INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: (2018 - 2019)

B. Stat 1st Year

Introduction to Programming and Data Structures

Date: 04.09.2018 Maximum Marks: 40 Duration: 2 hours

Answer all questions in brief.

1. Write a program to rearrange a list of strings alphabetically. [7]

2. What is the difference between (explain with suitable example) [3 + 3 + 3 = 9]

   a) int *ptr[10] and int (*ptr)[10]
   b) const int *ptr and const int *const ptr
   c) structure and union

3. Answer the following: [2 + 3 + 3 + 2 + 2 + 2 = 14]

   a) Explain the error (compile-time or run-time), if any, in the following C code:

```
#include<stdio.h>
struct emp
{
   char name[20];
   int age;
};
int main( )
{
   struct emp e1 = {"Ganguly",45};
   struct emp e2 = e1;
   if(e1= = e2)
         printf("Two structures are equal");
   else
         printf("Two structures are unequal");
   return 0;
}
```

   b) Explain the difference between int **data and int *(*data)( ).
   c) Illustrate the main purpose of passing variables by pointers, with the help of a function that swaps two integers.
   d) How can one dynamically allocate a two-dimensional array?

e) Explain the output of the following program:

```
#include<stdio.h>
int main( )
{
    struct node
    {
            int value;
            struct node *ptr;
    };
    struct node *p,q;
    printf("\n%d\t%d\n",(int)sizeof(p),(int)sizeof(q));
    return 0;
}
```

f) Explain the error (compile-time or run-time), if any, in the following C code:

```
#include<stdio.h>
int main( )
{
    int x = 10;
    int *ptr;
    float y;
    *ptr = 15;
    y = (float)x/*ptr;
    printf("The value of y = %f\n",y);
}
```

4. Distinguish between recursion and iteration. Write a program to find maximum of a set of integers using recursion. [3 + 7 = 10]

INDIAN STATISTICAL INSTITUTE, KOLKATA

MIDTERM EXAMINATION: FIRST SEMESTER 2018 -'19

B. STAT I YEAR

Subject: **Analysis I** Date: **September 05, 2018**

Duration: **3 hours** Time: **10:30 AM to 12:30 PM**

Maximum score (after scaling): **30**

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 (X, d) be any metric space and $\Gamma := \{V : V \text{ is an open subset of } X\}$. Let "$\leq$" denote the partial order on Γ given by

$$V_1 \leq V_2 \Leftrightarrow V_1 \subseteq V_2 .$$

(i) When will a subset of Γ be bounded above or bounded below?
(ii) Show that $(\Gamma, \leq)$ has the least upper bound property.
(iii) Given any subset Λ of Γ, find glb(Λ).

[2+3+4=9 marks]

(2) Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ satisfying the following property

$$|a_{n+1} - a_n| \leq \frac{2^{n+1}}{5^n} \text{ for all } n \in \mathbb{N}.$$

Show that the above sequence is convergent.

[6 marks]

(3) Let (X, d) be a metric space such that **any** family of closed sets having FIP must have the grand intersection non-empty, that is,
for **every** family $\{E_i\}_{i\in I}$ consisting of closed sets satisfying $\bigcap_{i\in F} E_i \neq \emptyset$ for all **finite** subset F of I, we have $\bigcap_{i\in I} E_i \neq \emptyset$.

Let $\{V_\gamma\}_{\gamma\in\Gamma}$ be an open cover of X. Show that it has a finite subcover.

[8 marks]

(4) Show that $\lim_{n\to\infty} \left(\sqrt{n^2 + 2n} - n\right)$ exists using the definition of "limit of a sequence".

[6 marks]

(5) Consider the metric space $\mathbb{Q}$ with the usual metric $d(x, y) = |x - y|$. Let $A := \left\{r \in \mathbb{Q} : -\sqrt{2} < r < \sqrt{2}\right\}$.
(i) Show that A is closed in $\mathbb{Q}$.
(ii) Is A compact? Provide a proof if 'yes' or justify if 'no'.

[2+5 marks]

Indian Statistical Institute
Probability Theory I
B-I, Midsem

Date: Sep 06, 2018 Duration: 2hrs.

Attempt all questions. The maximum you can score is 50. Justify all your steps.

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. We have n letters and n corresponding envelopes. The letters are placed randomly in the envelopes (all possible conugurtions being equally likely). Use inclusion-exclusion principle to find the chance p_n that no letter is in its correct envelope. What is $\lim_{n\to\infty} p_n$?

   [8 + 2 marks]

2. Let $F : \mathbb{R} \to \mathbb{R}$ be the CDF of a real valued random variable X. Let $a \in \mathbb{R}$. Show that $P(X < a) = F(a-)$. [You may assume existence of $F(a-)$.]

   [10 marks]

3. Prove or disprove the following *infinite* version of the theorem of total probability:

   Let $\{A_n : n \in \mathbb{N}\}$ be an infinite collection of mutually exclusive and exhaustive events each with positive probability. Let B be any event. Then the series

   $$\sum_{n=1}^{\infty} P(A_n)P(B|A_n)$$

   must converge to $P(B)$.

   [10 marks]

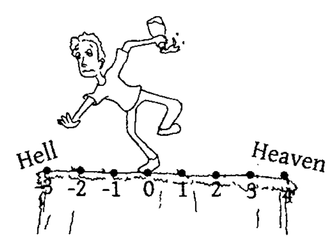
4. Let X be a discrete random variable with expectation $\mu \in \mathbb{R}$. Let ϵ be any positive number. Show that

   $$P(|X-\mu| \geq \epsilon) \leq \frac{E(|X-\mu|)}{\epsilon}.$$

   [Hint: In the series defining $E(|X-\mu|)$ consider those values of X that lie at a distance $\geq \epsilon$ from μ.] [10 marks]

[More trouble on the other side!]

5. A drunkard was standing on a cliff top as shown. Every second he moved by one unit in either direction with equal probability (unless he fell off by trying to step beyond -3 or 4). His steps were all independent. Given that he has fallen off the cliff within the first 10 seconds, what is the probability that he went to heaven?



[15 marks]

INDIAN STATISTICAL INSTITUTE

Mid Semester Examination, 1^{st} Semester, 2018-19

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

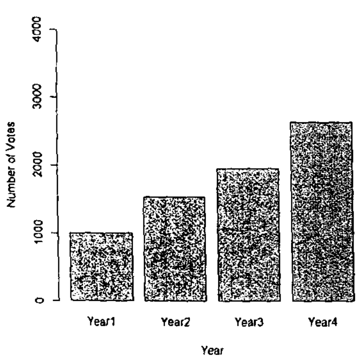
Date: September 7, 2018 Time:2 hours

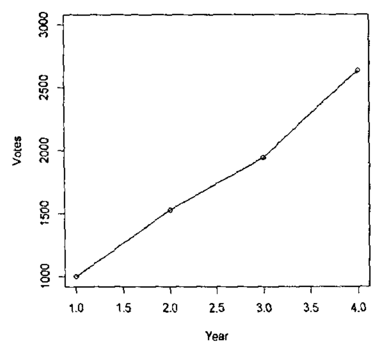
This paper carries 25 marks including 5 marks for data analyses using R

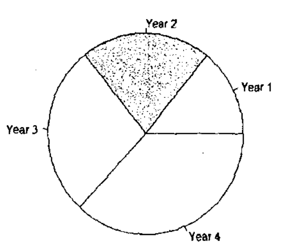
1. Suppose $x_1, x_2, \ldots, x_n$ be a set of observations. Obtain a measure of center θ that minimizes the quantity $\sum_{i=1}^{n} |e^{-x_i} - e^{\theta}|$ w.r.t. θ. [2]

2. In a study on a large number of individuals who have pacemakers and have traveled on direct flights between India and USA, it was found that 93% of the individuals had suffered from breathlessness during the flight. Is this a prospective, a retrospective or an epidemiological study design? Is this an observational or a controlled study? Do you think that the study demonstrates that having pacemakers cause discomfort to passengers during long flights? [3]

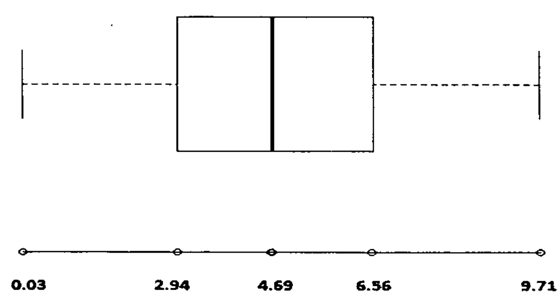
3. Suppose we wish to present the relative popularity of a political party in four consecutive elections. Which of the following diagrams is appropriate for this purpose? [2]



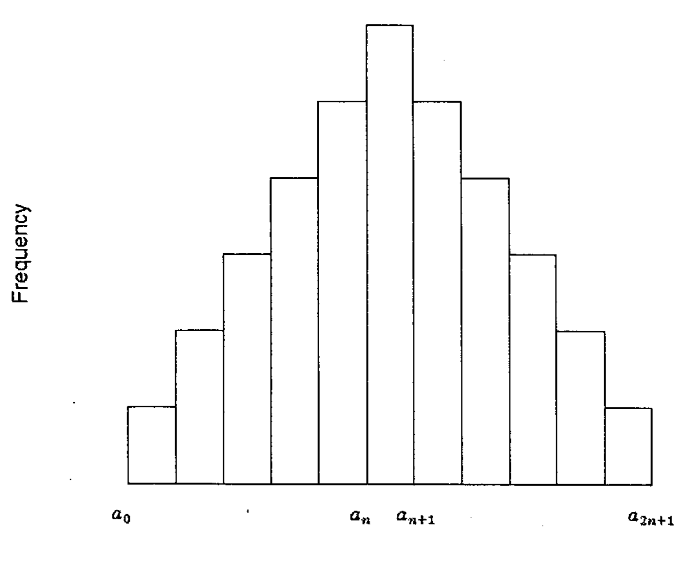




4. Consider the following box and whisker chart for a set of 20 observations. What can be said about the mean of this set? [3]



5. The following frequency distribution comprises $(2n+1)$ classes with equal class widths and the frequency of the i^{th} class equal to f_i, $i = 1, 2, \ldots, 2n+1$. If $f_i = f_{2n+2-i}$, $i = 1, 2, \ldots, n$, show that the mean, median and mode of this frequency distribution are identical. [5]



6. In a study on the difference between the mean weights of individuals having Type 2 Diabetes and having normal fasting glucose levels, it was found that the mean weight of the first group was 10 kg more than that of the second group while the standard deviations of the weights within the two groups were 6 kg and 8 kg, respectively. If the two groups are pooled, what can be concluded about the standard deviation of the weights in the pooled group? [5]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: Semester I (2018-19)

B.STAT. 1^{ST} YEAR

Course: REMEDIAL ENGLISH

Date: September 11, 2018 **Total Marks 30** **Time: 90 minutes**

A. Read the passage below and answer the questions: [8×1=8]

The next few decades will see great changes in the way energy is supplied and used. In some major oil producing nations, 'peak oil' has already been reached, and there are increasing fears of global warming. Consequently, many countries are focusing on the switch to a low carbon economy. This transition will lead to major changes in the supply and use of electricity. **[A]** Firstly, there will be an increase in overall demand, as consumers switch from oil and gas to electricity to power their homes and vehicles. **[B]**Secondly, there will be an increase in power generation, not only in terms of how much is generated, but also how it is generated, as there is growing electricity generation from renewable sources. **[C]** To meet these challenges, countries are investing in Smart Grid technology. **[D]** This system aims to provide the electricity industry with a better understanding of power generation and demand, and to use this information to create a more efficient power network.

Smart Grid technology basically involves the application of a computer system to the electricity network. The computer system can be used to collect information about supply and demand and improve engineer's ability to manage the system. With better information about electricity demand, the network will be able to increase the amount of electricity delivered per unit generated, leading to potential reductions in fuel needs and carbon emissions. Moreover, the computer system will assist in reducing operational and maintenance costs.

Smart Grid technology offers benefits to the consumer too. They will be able to collect real-time information on their energy use for each appliance. Varying tariffs throughout the day will give customers the incentive to use appliances at times when supply greatly exceeds demand, leading to great reductions in bills. For example, they may use their washing machines at night. Smart meters can also be connected to the internet or telephone system, allowing customers to switch appliances on or off remotely. Furthermore, if houses are fitted with the apparatus to generate their own power, appliances can be set to run directly from the on-site power source, and any excess can be sold to the grid.

With these changes comes a range of challenges. The first involves managing the supply and demand. Sources of renewable energy, such as wind, wave and solar, are notoriously unpredictable, and nuclear power, which is also set to increase as nations switch to alternative energy sources, is inflexible. With oil and gas, it is relatively simple to increase the supply of energy to match the increasing demand during peak times of the day or year. With alternative sources, this is far more difficult, and may lead to blackouts or system collapse. Potential solutions include investigating new and efficient ways to store energy and encouraging consumers to use electricity at off-peak times.

A second problem is the fact that many renewable power generation sources are located in remote areas, such as windy uplands and coastal regions, where there is currently a lack of electrical infrastructure. New infrastructures therefore must be built. Thankfully, with improved smart technology, this can be done more efficiently by reducing the reinforcement or construction costs.

Although Smart Technology is still in its infancy, pilot schemes to promote and test it are already underway. Consumers are currently testing the new smart meters which can be used in their homes to manage electricity use. There are also a number of demonstrations being planned to show how the smart technology could practically work, and trials are in place to test the new electrical infrastructure. It is likely that technology will be added in 'layers', starting with 'quick win' methods which will provide initial carbon savings, to be followed by more advanced systems at a later date. Cities are prime candidates for investment into smart energy, due to the high population density and high energy use. It is here where Smart Technology is likely to be promoted first, utilising a range of sustainable power sources, transport solutions and an infrastructure for charging electrically powered vehicles. The infrastructure is already changing fast. By the year 2050, changes in the energy supply will have transformed our homes, our roads and our behaviour.

1. According to paragraph 1, what has happened in some oil producing countries?

- A They are unwilling to sell their oil any more.
- B They are not producing as much oil as they used to.
- C The supply of oil is unpredictable.
- D Global warming is more sever here than in other countries.

2. Where in paragraph 1 can the following sentence be placed?

There is also likely more electricity generation centres, as households and communities take up the opportunity to install photovoltaic cells and small scale wind turbines.

- A
- B
- C
- D

3. Which of the following is NOT a benefit of Smart Grid technology to consumers?

- A It can reduce their electricity bills.
- B It can tell them how much energy each appliance is using.
- C It can allow them to turn appliances on and off when they are not at home.
- D It can reduce the amount of energy needed to power appliances.

4. According to paragraph 4, what is the problem with using renewable sources of power?

- A They do not provide much energy.
- B They often cause system failure and blackouts.
- C They do not supply a continuous flow of energy.
- D They can't be used at off-peak times.

5. In paragraph 5, what can be inferred about cities in the future?

- A More people will be living in cities in the future than nowadays.
- B People in cities will be using cars and buses powered by electricity.
- C All buildings will generate their own electricity.
- D Smart Grid technology will only be available in cities.

6. The word 'underway' in paragraph 6 is closest in meaning to:

- A permanent
- B complete
- C beneficial
- D in progress

7. What is the main idea of the final paragraph? (paragraph 6).

- A To describe who will benefit from Smart Grid technology first.
- B To outline the advantages of Smart Grid technology.
- C To summarise the main ideas in the previous paragraphs.
- D To describe how, where and when Smart Technology will be introduced.

8. In paragraph 6, what can be inferred about the introduction of Smart Grid Technology?

- A The technologies which produce most benefits will be introduced first.
- B The cheapest technologies will be introduced first.
- C The technologies which are most difficult to put into place will be introduced first.
- D Technologically advanced systems will be introduced first.

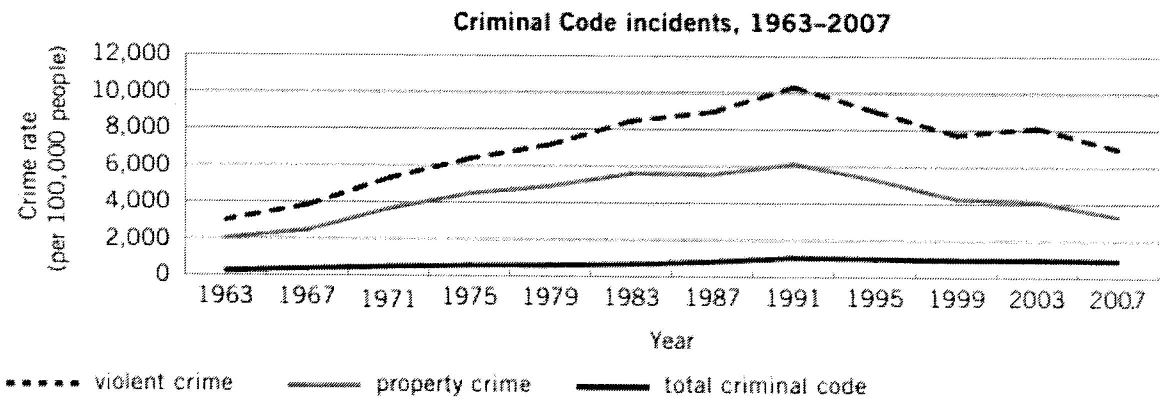
B. Fill in the gaps with phrasal verbs. Sometimes, you must use a pronoun between the main verb and the particle. If this is the case, the required pronoun can be found in brackets after the space. [12×1=12]

1. I'm tired and stressed and the kids have been ....all day. They don't seem to know how to behave properly.
2. I met Harold in the supermarket. He was .... you and I said you were fine and really happy about this new job you've got.
3. I'll .... (your boss) for dinner tonight if he isn't busy. It will make a good impression, don't you think?
4. If your boss accuses you of missing yesterday's meeting on purpose, I will ....
5. I can hardly read this document. Go and...... .... (it) so it is at least twice the size.
6. These statistics are too confusing. We need to...... .... (them) down into the four different seasons of the year for a clearer picture.
7. If someone phones me while I am in the meeting, I will have to......... ....for a few minutes as I am expecting a really important call.
8. After the First World War, the Austro-Hungarian empire was........... ......into several pieces.
9. Sales are up 10% on this period last year. We need to......... .....that if we are to save this company.
10. When I told her about her mother's illness being incurable, she........... ....tears.
11. He............ .... the police inspector by buying his wife a nice present and complimenting his car, hoping it would help his son who was being held in custody.
12. The trade union leader may be a problem with these new conditions but Mr. Dutt seems to think he can be.............. .....with a few thousand rupees.

C. Look at the graph below and write a paragraph. [10]

Write a paragraph to describe the graph below. Begin with the provided topic sentence and prompts, and indicate possible reasons for the highs and lows in your supporting details.

The crime rate has changed significantly over the past 45 years.



Data Source: Statistics Canada. Total criminal code refers to crimes addressed by criminal law and excludes traffic offences

INDIAN STATISTICAL INSTITUTE

Semestral Examination : First Semester 2018-19

Course : B Stat I year

Subject : Vectors and Matrices

Date : 12th November 2018 | Maximum Marks : 65 | Duration : 3 Hours

While using any result clearly state that.

1. Let V_1 be a 4-dimensional vector space with ordered basis $B_1 = \{x_1, x_2, x_3, x_4\}$ and V_2 be a 3-dimensional vector space with ordered basis $B_2 = \{y_1, y_2, y_3\}$. Let $T : V_1 \to V_2$, be a linear transformation. Suppose that with respect to the bases B_1, B_2 the linear transformation T has the matrix representation

$$T_{B_1,B_2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

What would be the matrix of T with respect to the bases $\{x_1, x_2, x_3, x_4\}, \{y_1 + y_2 + y_3, y_2 + y_3, y_3\}$ of V_1, V_2 respectively? **[7]**

2. Reduce the matrix

$$A = \begin{bmatrix} 0 & 0 & 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 0 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}$$

to row echelon form. Exhibit a basis for the kernel of the linear transformation given by A. Exhibit a basis for the annihilator of the kernel. **[7+3+3=13]**

3. Let $A \in M_n(\mathbb{R}), u, v \in M_{n\times 1}(\mathbb{R})$. Suppose A is invertible. Show that $A + uv^T$ is invertible iff $v^T A^{-1} u \neq -1$. **[10]**

4. Let W be an affine subspace of a vector space V. Then show that there is a subspace V_W of V and $b \in W$ such that

$$W = V_W + b = \{v + b : v \in V_W\}.$$

Recall that the dimension of W is defined as $\dim V_W$. **[10]**

5. Let $A \in M_{m\times n}(\mathbb{R})$ and $\mathscr{G}_A = \{G \in M_{n\times m}(\mathbb{R}) : AGA = A\} \subseteq M_{n\times m}(\mathbb{R})$ be the collection of g-inverses of A. Show that $\mathscr{G}_A$ is an affine subspace of $M_{n\times m}(\mathbb{R})$. If A has rank r what would be the dimension of $\mathscr{G}_A$? Prove your assertion. **[2+2+6=10]**

6. Let V be a finite dimensional $\mathbb{R}$-vector space and $J : V \to V$ an $\mathbb{R}$-linear map such that $J^2 = -1$. Show that V must be even dimensional. **[10]**

7. A square matrix $A \in M_n(\mathbb{R})$ is called an integral matrix if all entries of A are integers. In other words $A \in M_n(\mathbb{Z})$. Show that there exists $B \in M_n(\mathbb{Z})$ such that $AB = BA = I$ iff $\mathrm{Det}(A) \in \{\pm 1\}$. **[7+8=15]**

INDIAN STATISTICAL INSTITUTE

First-Semestral Examination: (2018 - 2019)

B. Stat 1st Year

Introduction to Programming and Data Structures

Date: 15.11.2018 **Maximum Marks:** 100 **Duration:** 3.5 hours

---

The question paper is divided into following two groups: **Group A:** 70 marks, 2 hours duration; **Group B:** 30 marks, 1.5 hours duration.

Group A

Marks: 70 **Duration:** 2 hours

Answer all questions in brief.

1. Answer the following: [4 + 2 + 2 + 2 = 10]

   a) Illustrate the main purpose of passing variables by pointers, with the help of a function that swaps two integers.
   b) Explain with example the difference between a string and an array of characters.
   c) What is the speciality of void pointer?
   d) Explain the error (compile-time or run-time), if any, in the following C code:

```
#include<stdio.h>
struct emp{
    char name[20];
    int age;
};
int main( ){
    struct emp e1 = {"Ganguly", 45};
    struct emp e2 = e1;
    if(e1= = e2)
        printf("Two structures are equal");
    else
        printf("Two structures are unequal");
    return 0;
}
```

2. Write the function ***strindex(char *S, char T)***, which returns the position of the rightmost occurrence of character ***T*** in string ***S***, or -1 if there is none. [6]

3. Write a function to multiply two matrices with appropriate orders and return the resultant matrix using the following function prototype: ***int **multiply(int **, int, int, int **, int, int)***. [10]

4. Explain with suitable examples the differences between [5 X 4 = 20]

   a) int *data( ) and int (*data)( )
   b) const int *ptr and int *const ptr
   c) void fn(int a [ ]) and void fn(int *a)

d) fscanf( ) and fread( )
e) stack and queue

5. Write the function ***rearrange(char **S, int N)***, which rearranges a set of ***N*** strings ***S*** alphabetically. [8]

6. Answer the following: [4 + 4 + 3 + 3 + 2 = 16]

a) Explain with example the differences between realloc( ) and calloc( ), in terms of the functions they perform.
b) Define a macro ***swap(t, x, y)*** that interchanges two variables ***x*** and ***y*** of data type ***t***.
c) Explain the output produced by the following C program:

```
#include<stdio.h>
void show1(void){
    static int x = 10; x + = 5; printf("%d\n",x);
}
void show2(void){
    int x = 10; x + = 5; printf("%d\n",x);
}
void show3(void){
    static int x; x = 10; x + = 5; printf("%d\n",x);
}
int main( ){
    show1( );show1( );
    show2( );show2( );
    show3( );show3( );
    return 0;
}
```

d) Explain the output produced by the following C program:

```
#include<stdio.h>
int main( ){
    char string1[ ] = "ISI";
    char string2[ ] = "ISI";
    if(string1= = string2) printf("Two strings are same");
    else printf("Two strings are different");
    return 0;
}
```

e) Explain the output of the following C program:

```
#include<stdio.h>
int main( ) {
    struct node{
            int value;
            struct node *ptr;
    };
    struct node *p,q;
    printf("\n%d\t%d\n",(int)sizeof(p),(int)sizeof(q));
    return 0;
}
```

INDIAN STATISTICAL INSTITUTE

First-Semestral Examination: (2018 - 2019)

B. Stat 1st Year

Introduction to Programming and Data Structures

Date: 15.11.2018 **Maximum Marks:** 100 **Duration:** 3.5 hours

---

The question paper is divided into following two groups: **Group A:** 70 marks, 2 hours duration; **Group B:** 30 marks, 1.5 hours duration.

Group B

Marks: 30 **Duration:** 1.5 hours

Write a C program to perform the following tasks: [6 + 8 + 12 + 4 = 30]

(1) *Create* two linked lists, namely, **L1** and **L2**, with ***m*** and ***n*** nodes, respectively.

[**L1 → 12 → 5 → 7 → 11 → 9 → 17 → 4 → NULL** and **L2 → 10 → 9 → 5 → 13 → NULL**]

(2) *Sort* **List1** and **List2** in ascending order.

[**L1 → 4 → 5 → 7 → 9 → 11 → 12 → 17 → NULL** and **L2 → 5 → 9 → 10 → 13 → NULL**]

(3) *Merge* two sorted lists, obtained in (2), in such a way that the merged linked list becomes a sorted list with distinct nodes.

[**L1 → 4 → 5 → 7 → 9 → 10 → 11 → 12 → 13 → 17 → NULL**]

(4) *Display* the linked list(s) for each of the above tasks.

You may use constant amount of memory for tasks (2) and (3). For task (3), you cannot apply sorting algorithm on merged linked list. While writing the program, you need to take care of the following:

1. The names of the variables and functions should reflect their functionality.
2. Dynamic memory allocation/deallocation should be done.
3. Proper documentation of the program should be provided.

Indian Statistical Institute, Kolkata

Final Semester Examination : Semester 1 (2018-2019)

B.STAT. 1st YEAR

Course : REMEDIAL ENGLISH

Date : November 16, 2018 Total Marks: 60 Time: 2 hours 30 minutes

A. Read the following paragraph to answer the next eight questions (Questions 1-8). 1x8

One of the modern world's intriguing sources of mystery has been aeroplanes vanishing in mid-flight. One of the more famous of these was the disappearance in 1937 of a pioneer woman aviator, Amelia Earhart. On the second last stage of an attempted round the world flight, she had radioed her position as she and her navigator searched desperately for their destination, a tiny island in the Pacific. The plane never arrived at Howland Island. Did it crash and sink after running out of fuel? It had been a long haul from New Guinea, a twenty hour flight covering some four thousand kilometres. Did Earhart have enough fuel to set down on some other island on her radioed course? Or did she end up somewhere else altogether? One fanciful theory had her being captured by the Japanese in the Marshall Islands and later executed as an American spy; another had her living out her days under an assumed name as a housewife in New Jersey. Seventy years after Earhart's disappearance, 'myth busters' continue to search for her. She was the best-known American woman pilot in the world. People were tracking her flight with great interest when, suddenly, she vanished into thin air. Aircraft had developed rapidly in sophistication after World War One, with the 1920s and 1930s marked by an aeronautical record-setting frenzy. Conquest of the air had become a global obsession. While Earhart was making headlines with her solo flights, other aviators like high-altitude pioneer Wiley Post and industrialist Howard Hughes were grabbing some glory of their own. But only Earhart, the reserved tomboy from Kansas who disappeared three weeks shy of her 40th birthday, still grips the public imagination. Her disappearance has been the subject of at least fifty books, countless magazine and newspaper articles, and TV documentaries. It is seen by journalists as the last great American mystery. There are currently two main theories about Amelia Earhart's fate. There were reports of distress calls from the Phoenix Islands made on Earhart's radio frequency for days after she vanished. Some say the plane could have broadcast only if it were on land, not in the water. The Coast Guard and later the Navy, believing the distress calls were real, adjusted their searches, and newspapers at the time reported Earhart and her navigator were marooned on an island. No-one was able to trace the calls at the time, so whether Earhart was on land in the Phoenix Islands or there was a hoaxer in the Phoenix Islands using her radio remains a mystery. Others dismiss the radio calls as bogus and insist Earhart and her navigator ditched in the water. An Earhart researcher, Elgen Long, claims that Earhart's airplane ran out of gas within fifty-two miles of the island and is sitting somewhere in a 6,000-square-mile area, at a depth of 17,000 feet. At that depth, the fuselage would still be in shiny, pristine condition if ever anyone were able to locate it. It would not even be covered in a layer of silt. Those who subscribe to this explanation claim that fuel calculations, radio calls and other considerations all show that the plane plunged into the sea somewhere off Howland Island. Whatever the explanation, the prospect of finding the remains is unsettling to many. To recover skeletal remains or personal effects would be a grisly experience and an intrusion. They want to know where Amelia Earhart is, but that's as far as they would like to go. As one investigator has put it, "I'm convinced that the mystery is part of what keeps us interested. In part, we remember her because she's our favourite missing person."

Q1. Amelia Earhart's nationality was:

A: English B: Australian C: Canadian D: American E: South African

Q.2 All the following are theories about Amelia's fate EXCEPT:
A: she crashed on a remote island somewhere near her destination.
B: her plane ran out of fuel and crashed into the sea.
C: she was captured by the Japanese and executed as a spy.
D: she escaped incognito and lived under an assumed name.
E: she crashed somewhere on Howland Island.

Q.3 The most convincing evidence that Amelia crashed somewhere on land was:
A: the finding of aircraft remains.
B: sightings by islanders.
C: radio contact with the coastguard from the Phoenix Islands.
D: distress signals from the Phoenix Islands on her particular radio frequency.
E: All of these.

Q.4 If the aircraft were ever recovered from its probable sea grave:
A: it would be hardly recognisable.
B: it would be in pristine condition and considered highly valuable.
C: it may reveal some grisly evidence.
D: A and C together.
E: B and C together.

Q.5. The fate of Amelia Earhart still fascinates investigators for all the following reasons EXCEPT:
A: she was a famous female aviator and adventurer.
B: there are such conflicting theories about her disappearance.
C: she was so close to the end of her journey.
D: she may have staged her own disappearance.
E: she presents one of the twentieth century's great unsolved mysteries.

Q.6. " You cannot be a hero without being a coward." What does this sentence suggest?
A: Heroes are transformed cowards.
B: To be truly heroic, you first have to know the meaning of fear.
C: Heroes are cowards in disguise.
D: You can never be one or the other; it is always a combination of both.
E: None of these.

Q.7. What is fuselage and where all would you find it?

Q.8. How do you think the Coast Guard and the Navy 'adjusted their search" ?

B. Fill in the spaces with the correct tense forms: ½ x 36=18
1. When we reach Land's End we ____________________ 1,500 km. (walk)
2. I just remembered that I ____________________ the rent yet. I'm surprised that the landlord ____________________ me up and reminded me. (not pay, not ring)
3. It's a beautiful drive. I am sure you ____________________ the scenery. (enjoy)

4. The car ____________________. If you get in Tom and I ____________________ you a push. (not start, give)
5. I put the five-pound notes into one of the books; but the next day it ____________________ me ages to find it because I ____________________ which book I ____________________ it in. (take, forget, put)
6. He ____________________ the bagpipes since six this morning. He ____________________. (play, just stop)
7. My son ____________________ work yet. He's still at High School. – How long ____________________ at school? - He ____________________ there for six years. Before that he ____________________ five years at primary school. (not start, he be, be, spend)
8. Mary: I wonder what he ____________________ now. Ann: Well, his girlfriend ____________________ from Japan too, so I suppose he ____________________ Japanese. (say, come, speak)
9. While we ____________________ someone ____________________ into the house and ____________________us this note. (fish, break, leave)
10. It ____________________ for the last two hours so the game ____________________ (rain, be postponed)
11. When I ____________________ him he ____________________ a picture of his wife. – ____________________ it? (see, paint, you like)
12. This shop ____________________ for good next Monday. (close)
13. At 3 a.m. Jane ____________________ up her husband and said that she ____________________that someone ____________________ to get into the house. (wake, think, try)
14. This bike ____________________ in our family for the last 14 years. My father ____________________ it for the first five years, then my brother ____________________ it and I ____________________ it since then. (be, use,ride,have)
15. You see, Doctor, she ____________________ ill two days ago and since then she ____________________ anything. I'm sure she ____________________ several kilos. (fall, not eat, lose)

C. Fill in the spaces with the correct phrasal verbs: go on / pick up / come back / come up with / go back / find out / come out / go out / point out / grow up / set up / turn out / get out / come in(to) / take on with the correct phrasal verbs: 1x14

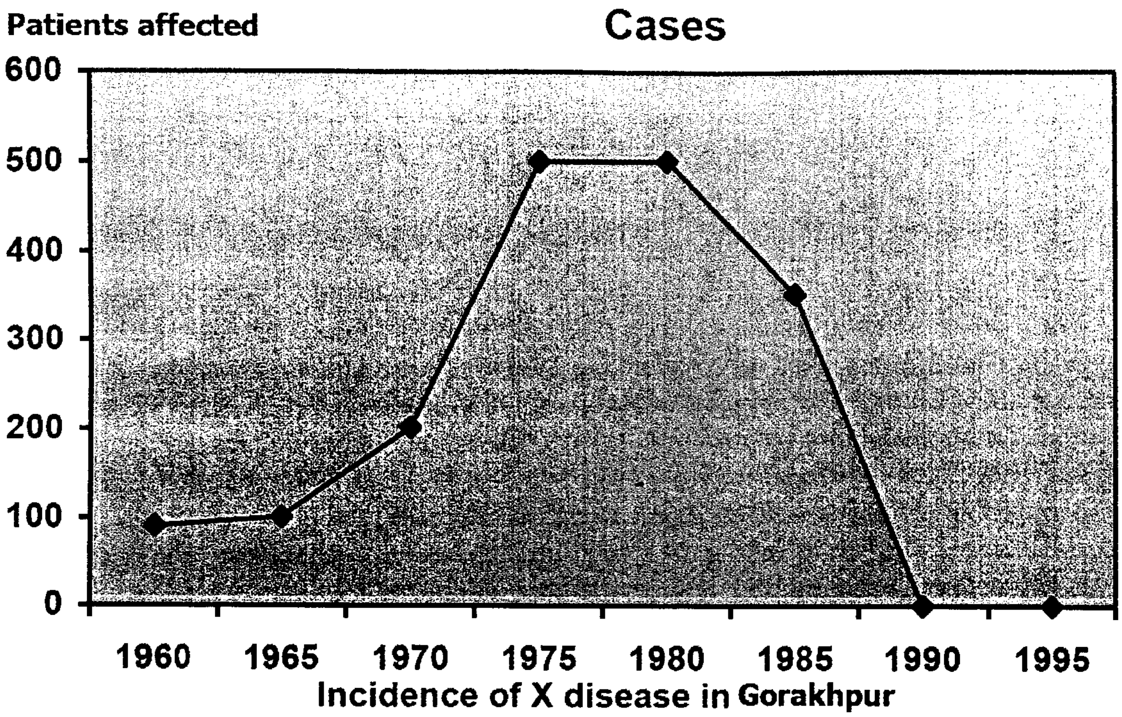
1. Can you ____ (think of an idea) a better idea?
2. She ____ (showed / mentioned) that the shops would already be closed.
3. I wish I hadn't ____ (become responsible for) so much work!
4. I ____ (went to an event) for dinner with my husband last night.
5. He ____ (entered a place where the speaker is) the kitchen and made some tea.
6. Where did you ____ (become an adult)?
7. I'd love to ____ (arrange / create) my own business.
8. I really want to ____ (leave a building) of this office and go for a walk.
9. As I arrived, he ____ (appeared from a place) of the door.

10. She ____________________________________ (got something from a place) some dinner on the way home.
11. Could you ____________________________________ (get information) what time we need to arrive?
12. I thought the conference was going to be boring but it ____________________________________ (in the end we discovered) to be quite useful.
13. What time did you ____________________________________ (return to a place where the speaker is) yesterday?
14. She ____________________________________ (appeared from a place) of the café and put on her gloves.

D. Summarize the following text in 100 words and give the summary a title: 10

In all learning, advances tend to come irregularly and in bursts, as you gain fresh insights into the subject. In order to obtain these insights you must thoroughly understand what you are studying. If you really understand a subject not only do you remember it easily, but you can apply your know ledge in new situations. The important thing is not what you know, but what you can do with what you know. The extra effort involved in getting a firm grounding in the essentials of a subject is repaid many times in later study. How are you to achieve understanding? Understanding involves (1) linking new knowledge to the old and (2) organizing it and remembering it in a systematic fashion. To retain and make sense of any new concept or fact it must be linked in as many ways as possible to your existing body of knowledge. All good introductory textbooks are constantly giving familiar examples, or using analogies, or appealing to common experience. In setting out the differences between daylight vision and twilight vision, for example, most writers point out that as twilight falls in the garden, blue flowers remain blue for some time after red blossoms appear black, illustrating, by appeal to common experience, that under dim illumination the colours of the blue end of the spectrum become relatively brighter than those of the red end. Or again, to illustrate that the movement of any particular electron during the passage of an electric current is only a few centimetres a second, although the velocity of the current is extremely great, the analogy is often used of a truck run into the end of a long line of trucks in a shunting yard, a corresponding truck being rapidly ejected from the far end. Linking new information to familiar experience in this fashion always helps understanding. In order to tie the new information to your stock of knowledge with as many links as possible, you must reflect on it, and try and relate it to what you already know. Thinking the matter over by yourself, writing out summaries of the main points, and talking to other students about it, are all valuable for fixing it more clearly in your mind.

E. Write a report for the HoD, Virology Department, Patna Medical College, after reading the data. Please do not mention causes or solutions, but just the facts. 10



Patients affected
Cases
600
500
400
300
200
100
0
1960
1965
1970
1975
1980
1985
1990
1995
Incidence of X disease in Gorakhpur

....

Indian Statistical Institute
Probability Theory I
B-I, Semestral 1

Date: Nov 19, 2018 Duration: 3hrs.

Attempt all questions. The maximum you can score is 50. Justify all your steps.

1. Prove or disprove each of the following statements.

   (a) If two real valued random variables X, Y are jointly continuous, then for any line L in $\mathbb{R}^2$ we must have $P((X,Y) \in L) = 0$.

   (b) If $X \sim \text{HyperGeom}(N_1, M_1, n_1)$ and $Y \sim \text{HyperGeom}(N_2, M_2, n_2)$ and they are independent, then $X+Y \sim \text{HyperGeom}(N_1+N_2, M_1+M_2, n_1+n_2)$.

   (c) Let X, Y be jointly distributed discrete real-valued random variables and let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function such that $E(f(X,Y)|X=4)$ and $E(f(4,Y))$ both exist finitely. Then we must have

$$E(f(X,Y)|X=4) = E(f(4,Y)).$$

[5+5+5]

2. Suppose that p is a fixed number in $[0,1]$. Suggest how you can use a fair coin to simulate an event with probability p. Justify your answer rigourously. [10]

3. 10 pieces of paper bearing the numbers 1, ..., 10 are arranged randomly in a line. We shall say that a *peak* occurs at position i if the i-th number is greater than its neighbour(s). Find the expected number of peaks. [10]

4. For each $n \in \mathbb{N}$ let X_n be a random variable taking values in $\mathbb{N}$. Also let X be a random variable taking values in $\mathbb{N}$. Let

$$\forall k \in \mathbb{N} \quad \lim_{n\to\infty} P(X_n = k) = P(X = k).$$

   Then is it true that $E(X_n) \to E(X)$? Assume finite existence of all the expectations involved. [10]

5. Prove that for jointly distrbuted random variables X, Y, Z with finite second moments, we have

$$cov(X,Y) = cov(E(X|Z), E(Y|Z)) + E(cov(X,Y|Z)).$$

[10]

INDIAN STATISTICAL INSTITUTE

Semester Examination, 1^{st} Semester, 2018-19

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

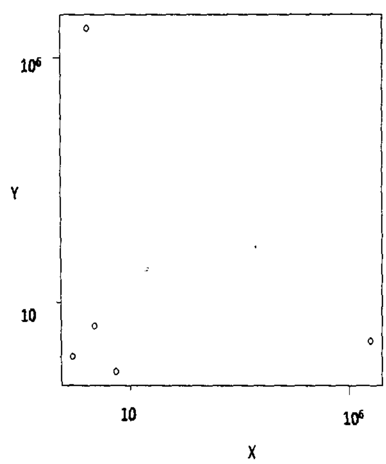
Date: November 22, 2018 Time:3 hours

This paper carries 50 marks. Answer all questions.

1. (a) Suppose $x_1, x_2, \ldots.., x_n$ is a set of observations with mean $\overline{x}$, standard deviation s and kurtosis β_2. Show that the proportion of observations that do not lie in the interval $(\overline{x} - 2s, \overline{x} + 2s)$ cannot exceed $\beta_2/16$.

   (b) Consider a set of distinct observations. If the minimum observation of the set is decreased and the maximum observation of the set is increased, would the mean deviation about the mean necessarily increase? Justify your answer. [3 + 7]

2. (a) Consider the following scatter plot for a set of 5 bivariate observations on (x, y). What is the approximate value of the correlation coefficient between x and y?



   (b) Suppose $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$ is a set of bivariate observations on (x, y) such that all data points are not collinear, $\overline{x}$ is the mean of the x_is and $\overline{y}$ is the mean of the y_is. When two additional observations (x_{n+1}, y_{n+1}) and (x_{n+2}, y_{n+2}) are included in the set, show that the least squares regression line of y on x and that of x on y both remain unchanged if and only if $x_{n+1} = x_{n+2} = \overline{x}$ and $y_{n+1} = y_{n+2} = \overline{y}$. [3 + 7]

3. (a) Suppose n individuals are ranked by two evaluators with no ties in the ranks assigned by either evaluator. Define R_i and S_i to be the ranks assigned to the i^{th} individual by the two evaluators such that $S_i = R_i + 1$ if $R_i < n$ and $S_i = 1$ if $R_i = n$, $i = 1, 2, \ldots, n$. If the value of Kendall's Tau for the two sets of ranks is 0, what is the value of the Spearman's rank correlation?

P.T.O

(b) In a study on possible association between obesity and oral cancer based on 1200 individuals, the data were tabulated as follows:

	Oral Cancer	Normal
Obese	720	360
Normal	480	840

However, when the individuals were divided into two groups: vegetarians and non-vegetarians, the corresponding tables were obtained as:

Table 1: Vegetarians

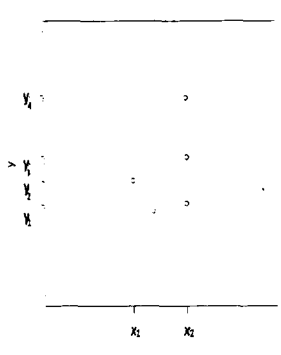
	Oral Cancer	Normal
Obese	80	200
Normal	320	800

Table 2: Non-vegetarians

	Oral Cancer	Normal
Obese	640	160
Normal	160	40

What is the nature of association between the two variables obtained from each of the above tables? Discuss any phenomenon underlying the contrasting evidence of association obtained from the different tables. [4 + 6]

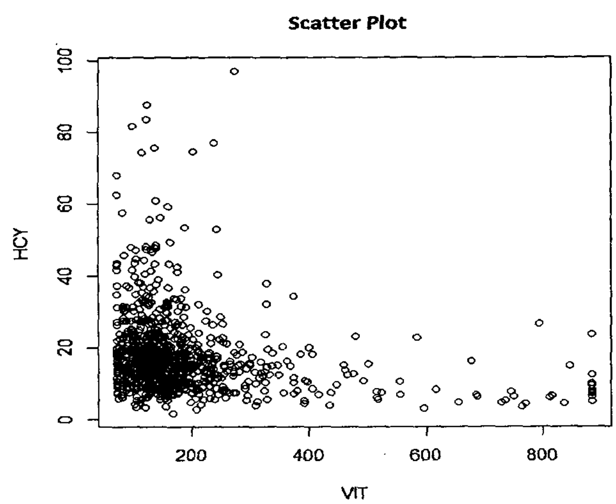
4. (a) Consider the following scatter plot for a set of 4 bivariate observations on $(x, y.)$ Obtain a necessary and sufficient condition for the least squares regression line of y on x to be identical to the Least Absolute Deviation line of y on x.



(b) Suppose $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$ is a set of bivariate observations on (x, y). Suppose a regression line of y on x is obtained by minimizing $\sum_{i=1}^{n} |y_i - \beta x_i|^{3/2}$ with respect to β. Describe a suitable algorithm to estimate β. [5 + 5]

5. In a study on Coronary Artery Disease (*CAD*), bivariate data on two quantitative precursor variables: homocysteine levels (*HCY*) and Vitamin B_{12} levels (*VIT*) were analyzed using *R*. Based on the following codes and corresponding outputs, write down your inferences with suitable justification. [10]

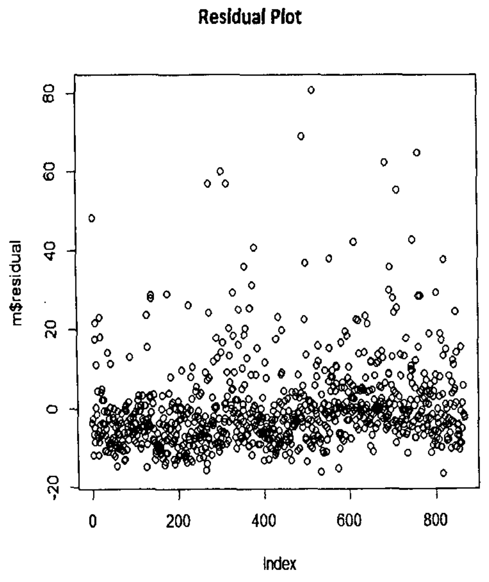
```
> Data=read.csv(file.choose(),header=TRUE)
> HCY=Data[,1]
> VIT=Data[,2]
> plot(VIT,HCY)
```



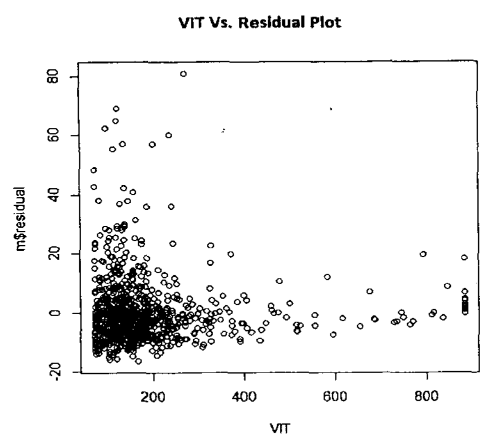

```
> m<-lm(HCY~VIT)
> m

Call:
lm(formula = HCY ~ VIT)

Coefficients:
(Intercept)        VIT
21.0124        -0.0177
```

P.T.O

```
> plot(m$residual)
```




```
> plot(VIT,m$residual)
```




```
> lm(VIT~HCY)

Call:
lm(formula = VIT ~ HCY)

Coefficients:
(Intercept)          HCY
    229.605       -2.535
```

INDIAN STATISTICAL INSTITUTE, KOLKATA
FINAL EXAMINATION: FIRST SEMESTER 2018 -'19
B. STAT I YEAR

Subject: **Analysis I** Date: **November 26, 2018**
Duration: **3 hours** Time: **10:30 AM to 1:30 PM**
Maximum score: **60**

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 $f : [0, \infty) \to \mathbb{R}$ be continuous on $[0, \infty)$ and differentiable on $(0, \infty)$. Further, $f(0) = 0$ and f' is increasing on $(0, \infty)$. Define $g : (0, \infty) \to \mathbb{R}$ by $g(x) := \dfrac{f(x)}{x}$. Show that g is increasing on its domain.

[12 marks]

(2) If $\{u_n\}_n$ is a sequence of non-negative real numbers converging to L, then show that $\dfrac{1}{n}\left(\sum_{i=1}^{n} u_i\right)$ converges to L.

[10 marks]

(3) Let the function $f : [0, \infty) \to \mathbb{R}$ be continuous. There exists $a \in (0, \infty)$ such that f is differentiable on (a, ∞) and f' is bounded. Show f is uniformly continuous.

[10 marks]

(4) Justify that there is no continuous function $f : \mathbb{R} \to \mathbb{R}$ which takes every value exactly twice.

[12 marks]

(5) Let $f : \mathbb{R} \to \mathbb{R}$ be a thrice differentiable function. Suppose $f(-1) = 0 = f(0) = f'(0)$ and $f(1) = 1$. Prove that there exists $s \in (0, 1)$ and $t \in (-1, 0)$ such that

$$f^{(3)}(s) + f^{(3)}(t) = 6.$$

[10 marks]

(6) Let $\{x_n\}_{n \geq 1}$ be a sequence defined recursively by

$$x_1 := \frac{1}{2} \quad \text{and} \quad x_{n+1} := \frac{x_n^2}{x_n^2 - x_n + 1}.$$

Prove that $\sum_{n=1}^{\infty} x_n$ is convergent. (*Hint: First, prove that (i)* $0 \leq x_n \leq \dfrac{1}{2}$ *for all* n *(using induction), and (ii)* $0 \leq \dfrac{a}{a^2 - a + 1} \leq \dfrac{2}{3}$ *for* $0 \leq a \leq \dfrac{1}{2}$.)

[12 marks]

INDIAN STATISTICAL INSTITUTE

Semestral Examination (Back Paper) : First Semester 2018-19

Course : B Stat I year

Subject : Vectors and Matrices

26.12.2018

Date : Yet to be decided | Maximum Marks :100 | Duration : 3 Hours

---

While using any result clearly state that.

1. Let V_1, V_2 be finite dimensional $\mathbb{R}$-vector spaces and $T : V_1 \to V_2$ an $\mathbb{R}$-linear map. Show that
$$Ann_{V_1}(\ker(T)) := \{\phi \in V_1^* | \phi(v) = 0, \forall v \in \ker(T)\} = T^*(V_2^*).$$ **(15)**

2. Let W be an affine subspace of a vector space V. Then show that there is a subspace V_W of V and $b \in W$ such that
$$W = V_W + b = \{v + b : v \in V_W\}.$$
Recall that the dimension of W is defined as $\dim V_W$. **(15)**

3. Let $V_1, V_2 \subseteq V$ be two subspaces of a vector space V and $b_1, b_2 \in V$. Then the affine subspaces $W_1 = V_1 + b_1, W_2 = V_2 + b_2$ intersect iff $\pi(b_1) = \pi(b_2)$ where $\pi : V \to V/(V_1 + V_2)$ is the quotient map. If W_1, W_2 intersect what would be the dimension of their intersection? Prove your assertion. **(6+6+3+5=20)**

4. Let $G \in M_{n\times m}(\mathbb{R})$ be a g-inverse of $A \in M_{m\times n}(\mathbb{R})$. Show that $\{G + (I - GA)U + V(I - AG) : U, V \in M_{n\times m}(\mathbb{R})\text{arbitrary}\}$ is the class of all g-inverses of A. **(20)**

5. Let $n \geq 1$. Consider the $\mathbb{R}$-vector space $V = M_{2n}(\mathbb{R})$. Let $A \in M_{2n}(\mathbb{R})$ be of rank n. Consider the linear map $T_A : V \to V$ given by $T_A(B) = AB$. Compute $\mathrm{Det}(T_A)$. **(20)**

6. Let C be a binary code of length 7 with generator matrix
$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$
Write down a parity check matrix P for G. Show that every non null vector in kernel of P has Hamming distance at least three from the origin. **(4+6=10)**

INDIAN STATISTICAL INSTITUTE

First-Semestral Examination: (2018 – 2019)
(Back Paper)

B. Stat 1st Year

Introduction to Programming and Data Structures

Date: 27/12/18 **Maximum Marks:** 100 **Duration:** 3 hours

Answer all questions in brief.

1. Distinguish between recursion and iteration. Write a program to find the maximum of a set of integers using recursion.

   [5 + 10 = 15]

2. A concordance is an alphabetical list of all the words in a text along with the number of occurrence of each word. Write a program that makes a concordance from an input text file. While writing the code, you need to take care of the following:

   a) The names of the variables should reflect their functions.
   b) Dynamic memory allocation should be used.
   c) Proper documentation of the program should be provided.

   [25]

3. Answer the following: [3 + 4 + 3 = 10]

   a) How can one dynamically allocate memory for an one-dimensional array?
   b) Illustrate the main purpose of passing variables by pointers, with the help of a function that swaps two integers.
   c) Explain the error (compile-time or run-time), if any, in the following C code:

```
#include<stdio.h>
struct emp
{
    char name[20];
    int age;
};
int main( )
{
    struct emp e1 = {"Ganguly", 45};
    struct emp e2 = e1;
    if(e1= = e2)
        printf("Two structures are equal");
    else
        printf("Two structures are unequal");
    return 0;
}
```

4. Write a function to add two matrices with appropriate orders and return the resultant matrix using the following function prototype:

***int **addition(int **, int, int, int **, int, int)*.**

[15]

5. Explain with suitable examples the differences between (*any five*) [5 X 5 = 25]

   a) int *data[10] and int (*data)[10]
   b) malloc( ) and calloc( )
   c) fprintf( ) and fwrite( )
   d) stack and queue
   e) a string and an array of characters
   f) structure and union
   g) *struct x1{........}* and *typedef struct {......}x2*

6. Discuss the following with suitable examples: [5 X 2 = 10]

   a) Linked List;
   b) Void Pointer.

INDIAN STATISTICAL INSTITUTE
Backpaper Examination, 1^{st} Semester, 2018-19

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

Date: 31·12·2018 Time:3 hours

This paper carries 100 marks. Answer all questions.

1. Examine whether each of the following statements is true or false:

   (a) If both the minimum and the maximum observations of a set are changed, it is possible that neither the geometric mean nor the quartile deviation changes.

   (b) If the kurtosis of a set of observations is 3, the skewness of the set cannot be **-1.5**.

   (c) The proportion of variance in y that is not explained by a least squares linear regression on x cannot be less than the correlation coefficient between x and y.

   (d) Assuming no ties, the Spearman's rank correlation between two sets of ranks is **-1** if and only if Kendall's tau for the two sets is also **-1**. [5 x 4]

2. Show that for a set of observations, the absolute difference between the mean and the median of the set cannot exceed half the range of the set. You need to derive any result you may use in your proof. [20]

3. Suppose $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$ is a set of bivariate observations on (x, y).

   (a) If $\mathbf{x}$ denotes the vector $(x_1, x_2, \ldots, x_n)$ and $\mathbf{y}$ denotes the vector $(y_1, y_2, \ldots, y_n)$, show that the correlation coefficient between x and y can be expressed as $\frac{\mathbf{x}'H\mathbf{y}}{\sqrt{\mathbf{x}'H\mathbf{x}}\sqrt{\mathbf{y}'H\mathbf{y}}}$ where H is a suitable idempotent matrix of rank $(n-1)$.

   (b) Consider the usual regression model:

$$y_i = \alpha + \beta x_i + e_i;\ i = 1, 2, \ldots, n$$

   where, x_is are fixed and e_is are random errors with mean 0 and the same variance. If one performs a least squares linear regression of y on x, by assuming that $\beta{=}\beta_0$ (so that α is the only unknown parameter in the regression equation), show that the residual sum of squares for this regression is higher than that under the usual regression of y on x (where there is no restriction on β) by an amount $(\hat{\beta} - \beta_0)^2 \sum_{i=1}^{n} (x_i - \overline{x})^2$ where $\overline{x}$ is the mean of x_is and $\hat{\beta}$ is the least squares estimator of β. [10 + 10]

4. Consider a set of observations $x_1, x_2, \ldots, x_n$. Suppose that θ is a summary measure of location of the above set which is obtained by minimizing $\frac{1}{2}\sum_{i=1}^{n} |x_i - \theta|^{\frac{1}{2}} + \frac{1}{3}\sum_{i=1}^{n} |x_i - \theta| + \frac{1}{6}\sum_{i=1}^{n} |x_i - \theta|^2$ w.r.t θ. Describe a suitable algorithm to estimate θ. It is desirable that such a summary measure should lie in between minimum and maximum observations of the set. Verify whether this property is satisfied by your algorithm. [10]

P.T.O

5. In an epidemiological study, the following bivariate data on two quantitative variables: body mass index (*BMI*) and folate levels (*fol*) are available.

id	*BMI*	*fol*
1	25.4	5.1
2	22.5	3.2
3	21.8	1.4
4	29.4	1.2
5	21.2	4.5

(a) Which one among *BMI* and *fol* has a more skewed distribution?

(b) Obtain the residuals of the least squares regression line of *fol* on *BMI*.

(c) Obtain a LAD regression line of *BMI* on *fol*. [10 + 10 + 10]

Indian Statistical Institute
Probability Theory I
B-I, Backpaper Examination 1

Date: 01.01.2019 Duration: 3hrs.

Attempt all questions. The maximum you can score is 45. Justify all your steps.

1. Prove or disprove each of the following statements.

   (a) If two real valued random variables X, Y are marginally discrete, then they must be jointly discrete.

   (b) A fair coin is repeatedly tossed until 3 consecutive heads occur. Let X be the total number of heads required. Then $E(X) = \infty$.

   (c) Let X, Y, Z be jointly distributed random variables. If X, Y are independent, then $E(XY|Z) = E(X|Z)E(Y|Z)$ assuming finite existence of all the expectations involved.

   (d) It is possible to have a random variable X such that $E(X^k)$ exist finitely for all even values of k but not for odd values of k.

   [5+5+5+5]

2. Suppose that p is a fixed number in $[0, 1]$. Suggest how you can use a fair coin to simulate an event with probability p. Justify your answer rigourously.

   [10]

3. 10 pieces of paper bearing the numbers $1, ..., 10$ are arranged randomly in a line. We shall say that a *peak* occurs at position i if the i-th number is greater than its neighbour(s). Find the variance of the number of peaks.

   [15]

4. Let $X_1, ..., X_n$ be i.i.d. $Poisson(\lambda)$ for some $\lambda > 0$. Find $P(X = 5)$, where $X = \max\{X_1, ..., X_n\}$. [10]

5. A box contains N balls, $[\alpha N]$ of which are red, the rest being green. Here $[x]$ is the largest integer $\leq x$. Also α is a fixed number in $(0, 1)$. An SRSWOR of size n is drawn. Let X be number of red balls in the sample. Show that, as $N \to \infty$, the probability distribution of X approaches a binomial distribution. What are the parameters of this binomial distribution? [10+2]

6. If $X \sim Geom(p)$, then show that

$$\forall m, n \in \mathbb{N} \quad P(X > m + n | X > n) = P(X > m).$$

[10]

Please turn over

7. If X is a nonnegative integer valued random variable, show that

$$E(X) = \sum_{1}^{\infty} P(X \geq n).$$

[10]

8. Let $X_1, ..., X_n$ be an SRSWR from $\{1, ..., N\}$. Let there be K *distinct* numbers in the sample. Find $E(K)$ and $V(K)$.

[5+8]

Probability Theory II
B. Stat. Ist Year Semester 2
Indian Statistical Institute

Mid-semestral Examination

Time: 2 Hours Full Marks: 30
Date: February 18, 2019

This is an OPEN NOTE examination. You may use any handwritten notes that you have brought with you, but you are not allowed to share them. Also no notes in printed, photocopied, cyclostyled or any other form (including electronic storage) is allowed.
Calculators and any other computing devices are strictly not allowed.

1. If X_1, X_2, X_3 and X_4 are i.i.d. Pareto(2) random variables, find the probability that exactly two of them exceed 3. [4]
2. Suppose that a particle is fired from the origin at an angle chosen uniformly at random on $(-\frac{\pi}{2}, \frac{\pi}{2})$ and that it hits the line $x = 1$ at the point $(1, Y)$. Find the distribution of Y. [6]
3. Let $\{X_n\}$ be an i.i.d. sequence of Cauchy random variables. Define $M_n = \max\{X_1, \ldots, X_n\}$. Show that $P[M_n/n \leq x]$ converges for every real number x. Call the limit $F(x)$. Show that it is a distribution function and identify it. [4+2=6]
4. If Z is a standard normal variable, then $X = \exp(Z)$ is called a lognormal random variable. Find $E[X^k]$, where k is a natural number. Show that, for any $\lambda > 0$, $E[\exp(\lambda X)]$ does not exist. [4+4=8]
5. Let $X_1, \ldots, X_n$ be an i.i.d. sequence of continuous random variables with density f and distribution function F, with $n \geq 2$. Let Y be the second smallest one among the random variables. Find the density of Y. If F is the uniform distribution on $(0, 1)$, identify the distribution of Y. [4+2=6]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2018-2019

B. Stat. (Hons.) 1st Year. 2nd Semester

Vectors and Matrices II

Date: February 19, 2019 | Maximum Marks: 40 | Duration: 2 hours

---

• This question paper carries 46 points. Answer as much as you can. However, the maximum you can score is 40.

• **You should present all your arguments while answering a question.**

• You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.

---

1. Let $\mathbf{I}_n$ be the identity matrix of order n. Also, let $\mathbf{u}$ and $\mathbf{v}$ be two vectors, each of dimension $n \times 1$. Assume $\mathbf{v}^\mathrm{T}\mathbf{u} + 1 \neq 0$. Find the inverse of $\mathbf{I}_n + \mathbf{u}\mathbf{v}^\mathrm{T}$. [9]

2. (a) Show that

$$\langle f, g \rangle = \int_0^1 f(t)\, g(t)\, dt$$

is an inner product on the vector space V of all real-valued continuous functions on $[0, 1]$.

(b) Find an orthogonal basis of the subspace spanned by $\{1, t, t^2\}$. [7+7=14]

3. Let $N(\cdot)$ be a norm on $\mathbb{R}^n$. For an $n \times n$ matrix $\mathbf{A}$, define

$$\|\mathbf{A}\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{N(\mathbf{A}\mathbf{x})}{N(\mathbf{x})}.$$

It is known that $\|\cdot\|$ is a norm on $\mathbb{R}^{n\times n}$. This is called *the matrix norm induced by the vector norm N*. Show that the matrix norm induced by the L_∞-norm on $\mathbb{R}^n$ is

$$\|\mathbf{A}\| = \max_i \textstyle\sum_{j=1}^n |a_{ij}|. \qquad [6]$$

[P.T.O.]

4. Answer **any one** of the following: [10]

(a) Let $\mathbf{A}_n$ be the following $n \times n$ matrix

$$\begin{bmatrix} 2\cos\theta & 1 & 0 & \cdots & 0 & 0 \\ 1 & 2\cos\theta & 1 & \cdots & 0 & 0 \\ 0 & 1 & 2\cos\theta & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 2\cos\theta & 1 \\ 0 & 0 & 0 & \cdots & 1 & 2\cos\theta \end{bmatrix}, \text{ where } 0 < \theta < \pi.$$

Show that

$$|\mathbf{A}_n| = \frac{\sin{(n+1)\theta}}{\sin\theta}.$$

(b) Let $\mathbf{A}_n$ be the following $n \times n$ matrix

$$\begin{bmatrix} a_1^n & a_1^{n-2} & a_1^{n-3} & \cdots & a_1 & 1 \\ a_2^n & a_2^{n-2} & a_2^{n-3} & \cdots & a_2 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_n^n & a_n^{n-2} & a_n^{n-3} & \cdots & a_n & 1 \end{bmatrix}.$$

Show that

$$|\mathbf{A}_n| = (a_1 + \cdots + a_n) \prod_{1 \le i < j \le n} (a_i - a_j).$$

5. Let $\mathbf{A}$ be an $n \times n$ matrix with $\rho(\mathbf{A}) = n - 1$. Find $n \times 1$ vectors $\mathbf{u}$ and $\mathbf{v}$ such that $\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}}$ is nonsingular or show that such a pair of vectors cannot exist. [7]

INDIAN STATISTICAL INSTITUTE

Analysis 2 : B. Stat 1st year
Mid-Semester Examination: 2018-19
20.02.2019.

Maximum Marks: 40 Maximum Time: $2\frac{1}{2}$ hrs.

- *Answer all the questions. But maximum you can score is* 40.
- *In the following* $C[a,b]$ *stands for the set of real valued continuous functions on* $[a,b]$.

(1) Prove that every real valued monotonic function defined on $[a,b]$ is Riemann integrable. [4]

(2) Suppose $f : [a,b] \to [m,M]$ is Riemann integrable and $g \in C[m,M]$. Without using the characterization of integrability using measure zero sets, prove that the composite function $g \circ f$ is Riemann integrable on $[a,b]$. [8]

(3) (a) Suppose f and g are real valued continuous functions on $[a,b]$ with g non-negative. Prove that there exists $c \in [a,b]$ such that

$$\int_a^b f(x)g(x)dx = f(c)\int_a^b g(x)dx.$$

(b) If $a \in (-1,0)$ then evaluate the following limit

$$\lim_{n\to\infty}\int_a^0 \frac{x^n}{1+x}dx.$$

[4+6=10]

(4) Suppose $f : [0,1] \to (0,1)$ is a continuous function. Prove that the equation [P.T.O]

$$2x - \int_0^x f(t)dt = 1,$$

has exactly one solution in $(0, 1)$. [6]

(5) Suppose $f \in C[0, 1]$ is such that $\int_0^1 f(t)dt = 1$. Prove that there exists $c \in (0, 1)$ such that $f(c) = 3c^2$. [6]

(6) Find the precise range of p for which the improper integral

$$\int_0^\infty \frac{1 - \cos x}{x^p} dx,$$

converges. [10]

INDIAN STATISTICAL INSTITUTE

Midsemester Examination : (2018-2019)

B. Stat 1st Year

Statistical Methods -II

Date: 21. 2. 19 Maximum marks: 60 Duration: 2 hours.

Note: This paper carries 70 points. Maximum you can score is 60. You may use any results proved in the class by stating the results clearly. If you use other results not discussed in class, they need to be proved. You may use calculators.

1. Suppose Y is regressed on X_1, X_2 and X_3 with an intercept term, and the following matrices are computed.

$$\text{Y'Y}= 5000 \quad \text{Y'X}= (20, 30, 50, -40), \quad \text{X'X}= \begin{bmatrix} 20 & 0 & 0 & 0 \\ 0 & 19 & 3 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

   (a) Calculate the sample mean and sample variance of Y. [4]

   (b) Compute the regression equation. [6]

   (c) Compute the analysis of variance of Y by deriving Total, regression and residual sum of squares. Produce the corrected ANOVA table. [8]

   (d) Compute the estimate of error variance. Find the estimated variances of each regression coefficient. Also find the estimated covariances between them. [7]

2. Consider data on three variables X_1, X_2 and X_3. Given that $r_{12} = 0.863, r_{13} = 0.648$ and $r_{23} = 0.709$, find the multiple correlation coefficient between X_1 and the other two variables. Also find the partial correlation coefficients $r_{12.3}$ and $r_{13.2}$. Comment on your results. [8+4+3]

[P.T.O]

3. Let $r_{1(2.34...p)}$ be the correlation between X_1 and $x_{2.34...p}$, the residual of X_2 removing the effects of $X_3, \ldots, X_p$. Show that (following usual notations)

(a) $r^2_{1(2.34...p)} \leq r^2_{12.34...p}$

(b)

$$\sum_i (x_{1.23...p})_i{}^2 = \sum_i (x_{1.34...p})_i \left[x_{1i} - b_{12.34...p} \left(x_{2.34...p}\right)_i\right]$$

which is equal to

$$\sum_i (x_{1.34...p})_i{}^2 - b_{12.34...p} \sum_i x_{1i} \left(x_{2.34...p}\right)_i.$$

(c) Using part (b) or otherwise show that

$$r^2_{1.23...p} = r^2_{1.34...p} + r^2_{1(2.34...p)}$$

and hence

$$r^2_{1.23...p} = r^2_{1p} + r^2_{1(p-1.p)} + \ldots.. + r^2_{1(2.34...p)}$$

[5+10+5]

4. Assignments [10]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2018-19

Course Name: B. Stat. 1st year

Subject: Numerical Analysis

Date: 22. 02. 2019

Maximum Marks: 60 Duration: 2.5 hrs.

Answer as much as you can.

1. a) Determine the order of accuracy for the numerical integration rule shown below:

$$\int_a^b f(x)dx \approx \frac{b-a}{3}\left[2f(\frac{3a+b}{4}) - f\left(\frac{a+b}{2}\right) + 2f(\frac{a+3b}{4})\right]$$

b) Consider the following quadrature rule:

$$\int_0^1 f(x)dx \approx \frac{1}{2} f(x_1) + w_1 f(x_2).$$

Determine the weight w_2 and the nodes x_1 and x_2 so that this formula gives exact results for all polynomials up to the highest possible degree.

c) Consider the initial value problem:

$$\frac{dx}{dt} = \frac{1+t}{1+x}$$

Use a second order Range-Kutta method with time step $h = 0.5$ to approximate $x(2)$. Show the details of your calculations.

5+4+6 = 15

2. (a) Consider the equation:

$$x^3 - x - 1 = 0$$

(i) Prove that the given equation has a solution in the interval [1, 2].

(ii) Give a bound for the number of bisection iterations needed to achieve an approximation with absolute error less than 10^{-6}.

(iii) Explain, with reference to the order of convergence, which one of (a) Newton's method and (b) bisection method should achieve the same accuracy with lesser number of iterations.

(b) Consider the following fixed point iteration:

$$x_{n+1} = g(x_n), \quad g(x) = \frac{1}{2}\left(x + \frac{2}{x}\right) \text{ and initial point } x_0 \in [1,2]$$

Prove that independent of the value of x_0, the sequence $\{x_n\}$ converges to a unique fixed point x^*. Also find the value of this fixed point.

(2+3+5)+(3+2) = 15

3. a) The error function, which very often arises in Physics and Mathematics, is given by:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2}\, ds$$

A few values of erf (x) is tabulated below:

x	0	0.1	0.2	0.3
$\operatorname{erf}(x)$	0	0.1124	0.2227	0.3286

i) Use a simple linear interpolation to estimate erf (0.17).

ii) Derive a rigorous error bound for your answer in part (i).

b) Show that the formula:

$$D^{(2)}(x) = \frac{f(x) - 2f(x-h) + f(x-2h)}{h^2}$$

provides an approximation for the second derivative $f''(x)$. Find the order of accuracy of this approximation.

(4+6)+5 = 15

4. a) Use the Doolittle's algorithm to derive an LU factorization of the following matrix and show all your steps:

$$A = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & -2 \\ 3 & 2 & -4 \end{bmatrix}$$

Use this LU factorization to solve the linear system of equations $Ax = b$ with $b = (4, 4, 6)^T$

b) A clamped cubic spline s for a function f is defined on [1, 3] by:

$$s(x) = \begin{cases} s_0(x) = 3(x-1) + 2(x-1)^2 - (x-1)^3, & \text{if } 1 \le x \le 2, \\ s_1(x) = a + b(x-2) + c(x-2)^2 + d(x-2)^3, & \text{if } 2 \le x \le 3. \end{cases}$$

Given that $f'(1) = f'(3)$, find a, b, c, and d.

6+5+4 = 15

5. a) Approximate the solution of the system of linear equations:

$$\begin{cases} x_1 - 2x_2 + 2x_3 = 1 \\ -x_1 + x_2 - x_3 = 1 \\ -2x_1 - 2x_2 + x_3 = 1 \end{cases}$$

using 3 iterations (starting with $x_1 = 0, x_2 = 0$, and $x_3 = 0$) of the Gauss-Siedel iterative method. Will Jacobi iterations converge to the exact solution for this system? Justify.

b) Find the number of subintervals and the step-size h so that the error for the composite trapezoidal rule is less that 5×10^{-9} for approximating the integral $\int_2^7 \frac{dx}{x}$.

(5+4)+6 = 15

INDIAN STATISTICAL INSTITUTE

Analysis 2 (Math101B) : B. Stat 1st year
End Semester Examination: 2018-19
22.04.2019.

Maximum Marks: 80 Maximum Time: $3\frac{1}{2}$ hrs.

- *Answer all the questions. But maximum you can score is* 80.
- *In the following* $C[a,b]$ *and* $R[a,b]$ *stand for the set of continuous functions and Riemann integrable functions on* $[a,b]$ *respectively.*

(1) Investigate the existence of two iterated limits and the double limit of the double sequence

$$f(p,q) = \frac{p}{p+q},$$

where $p \in \mathbb{N}$, $q \in \mathbb{N}$. [6]

(2) Evaluate

$$\lim_{n\to\infty}\left[\left(1+\frac{1}{n^2}\right)\left(1+\frac{2^2}{n^2}\right)^2\cdots\left(1+\frac{n^2}{n^2}\right)^n\right]^{2/n^2}.$$

[6]

(3) Let $f \in R[-\pi,\pi]$ and $S \subset \mathbb{Z}$ be a finite set. Define

$$\begin{aligned} a_n &= \hat{f}(n), \quad n \in \mathbb{Z}\setminus S, \\ &= 0, \quad n \in S. \end{aligned}$$

Does there exist a $g \in R[-\pi,\pi]$ such that $\sum_{n=-\infty}^{\infty} a_n e^{inx}$ is the Fourier series of g? Justify your answer. [6]

(4) Suppose $\{P_m\}$ is a sequence of polynomials with real coefficients which converges uniformly on $\mathbb{R}$. Prove that the limit function is a polynomial. [6]

[P.T.O]

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

$$\int_0^1 x^n f(x)dx = 0,$$

for all odd natural numbers n. Prove that f is the zero function. [8]

(6) (a) Let $\{f_n\}$ be a sequence of continuous functions on $[0,1]$ such that the series $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on $[0,1)$. Prove that $\sum_{n=1}^{\infty} f_n(1)$ converges.

(b) Does the series $\sum_{n=1}^{\infty} e^{-nx}\cos(nx)$ converge uniformly on $(0,1]$? Justify your answer. [4+4=8]

(7) Prove that, $\zeta(x) = \sum_{n=1}^{\infty} \frac{1}{n^x}$, defines a continuous function in $(1,\infty)$. Is ζ differentiable on $(1,\infty)$? Justify your answer. [10]

(8) Let $f_n(x) = n^\alpha x(1-x^2)^n$ for $x \in [0,1]$, $n \geq 1$. Show that $\{f_n\}$ converges pointwise on $[0,1]$ for all $\alpha \in \mathbb{R}$. Find all α such that the convergence is uniform on $[0,1]$ and all α such that

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

[12]

(9) Find all values of $\alpha \in [1,\infty)$ for which the improper integral

$$\int_1^\infty x\sin(x^\alpha)dx,$$

converges. [12]

(10) For $r \in (0,1)$ consider the Poisson kernel P_r given by

$$P_r(\theta) = \sum_{n=-\infty}^{\infty} r^{|n|}e^{in\theta} = \frac{1-r^2}{1-2r\cos\theta + r^2}, \quad \theta \in [-\pi,\pi].$$

(a) If f is a continuous 2π-periodic function on $\mathbb{R}$ then prove that for all $\theta \in [-\pi,\pi]$, $\lim_{r\to 1-} f * P_r(\theta) = f(\theta)$.

(b) Hence or otherwise prove that for any continuous 2π-periodic function on $\mathbb{R}$

$$\lim_{r\to 1-} \sum_{n=-\infty}^{\infty} r^{|n|}\hat{f}(n)e^{in\theta} = f(\theta),$$

for all $x \in [-\pi,\pi]$. [8+6=14]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination : (2018-2019)

B. Stat 1st Year

Statistical Methods -II

Date: 24. 4. 19 Maximum marks: 100 Duration: 3 hours.

Note: This paper carries 110 points. Answer as much as you can. Maximum you can score is 100. You can use any result proved in class, however, you have to state the result clearly and completely. You may use a calculator.

1. In the multiple regression model with an intercept, assume that there are p predictor variables which have mean 0.

   (a) Derive the expression for different sum of squares in the ANOVA decomposition and find their expected values. [15]

   (b) Assume further that the independent variables are uncorrelated. Consider p simple linear regressions with one independent variable taken at a time. Derive the relationship between the least-square estimates of the coefficients in the multiple regression model and the corresponding estimated coefficients in the p simple linear regression models. Also derive the relationship between the sum of squares due to regression for the multiple regression model and the sum of squares due to regression for the p simple linear regression models. [10]

2. Present a method to generate a value of the random variable X which has the following distribution:

$$P\{X=j\} = \left(\frac{1}{2}\right)^{j+1} + \frac{\left(\frac{1}{2}\right) 3^{j-1}}{4^j}, \quad j = 1, 2, \ldots$$

[15]

3. Suppose we want to generate a random number from the standard normal distribution.

   Give an algorithm based on the rejection method for this problem. Choose an appropriate value of the parameter for the distribution that you choose for this purpose so that the efficiency of this algorithm is maximized. [12 + 8]

[P.T.O.]

4. Find the MLE for θ based on n i.i.d. observations from $U(-\theta, \theta), \theta > 0$ distribution. [10]

5. A student was asked to roll a die 600 times. He reported the following frequency distribution:

Number	Frequency
1	98
2	103
3	96
4	102
5	104
6	97

Use a goodness of fit test to decide whether the die is fair or not. Also comment on the authenticity of the result with numerical justification. [10 + 5]

6. The distribution of the number of fatalities per regiment/year in the Prussian cavalry due to horse kicks is given below. See whether an appropriate poisson distribution can give a good fit.

Number of deaths	Frequency
0	109
1	65
2	22
3	3
4	1
≥ 5	0

(a) Check whether the method of moments and method of ML give same estimates of the parameter. [5]

(b) Calculate the chi-square statistic for goodness of fit. Comment on the results. [10]

7. Assignment. [10]

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2018–2019

B. Stat. (Hons.) 1st Year. 2nd Semester

Vectors and Matrices II

Date: April 26, 2019 | Maximum Marks: 60 | Duration: 3 hours

- Answer all the questions.
- **You should present all your arguments while answering a question.**
- You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.

---

1. (a) Show that the characteristic roots of the $n \times n$ matrix $\mathbf{P} := ((p_{ij}))$ with $p_{i,i+1} = 1$ for $i = 1, \ldots, n-1$, $p_{n1} = 1$, and $p_{ij} = 0$ otherwise, are the n-th roots of unity.

   (b) Consider the matrix

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ a_4 & a_1 & a_2 & a_3 \\ a_3 & a_4 & a_1 & a_2 \\ a_2 & a_3 & a_4 & a_1 \end{bmatrix}$$

   Show that the characteristic roots of $\mathbf{A}$ are $a_1 + a_2\omega_i + a_3\omega_i^2 + a_4\omega_i^3$, $i = 1, \ldots, 4$, where $\omega_1, \omega_2, \ldots, \omega_4$ are the 4-th roots of unity. [8+6=14]

2. Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ semi-simple matrices which satisfy $\mathbf{AB} = \mathbf{BA}$. Show that there exists a non-singular matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{AP}$ and $\mathbf{P}^{-1}\mathbf{BP}$ are diagonal. [12]

3. Suppose $\mathbf{A}$ is a 3×3 matrix all of whose leading principal minors are positive. Show that $\mathbf{A}$ is positive definite. [12]

4. If $\mathbf{A}$ and $\mathbf{B}$ are non-negative definite and if $\rho(\mathbf{A} - \mathbf{B}) = \rho(\mathbf{A}) - \rho(\mathbf{B})$, show that $\mathbf{A} - \mathbf{B}$ is non-negative definite. [12]

[P.T.O.]

5. Let $\mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{x}$ be a positive definite quadratic form, where $\mathbf{D}$ is $n \times n$. Also, let $\mathbf{A}\mathbf{x} = \mathbf{b}$ be a consistent system, where $\mathbf{A}$ is $m \times n$ with $\rho(\mathbf{A}) = m$. Show that

$$\min_{\mathbf{A}\mathbf{x}=\mathbf{b}} \mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{x} = \mathbf{b}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{b},$$

where $\mathbf{S} := \mathbf{A}\mathbf{D}^{-1}\mathbf{A}^{\mathrm{T}}$. [10]

PROBABILITY THEORY II

B. STAT. IST YEAR SEMESTER 2

INDIAN STATISTICAL INSTITUTE

Semestral Examination

Time: 3 Hours Full Marks: 70

Date: April 30, 2019

1. Let X be a Cauchy random variable and define $Z = 1/(1+X^2)$. Find density of Z. Identify it. Calculate its mean and variance. [[illegible]=10]
   5+1+2+2
2. Let X and Y be i.i.d. random variables with density function $\frac{1}{2}\exp(-|x|)$ on the entire real line. Find $\mathrm{P}[|\min(X,Y)| < 1]$. [10]
3. To come from home for the exam, you plan to catch the AC bus at Shyambazar, which is scheduled to arrive at 1:40 pm, but actually arrives according to a normal distribution with mean 1:40 pm and standard deviation 40 seconds. You arrive at the bus stop according to a normal distribution with mean 1:39 pm and standard deviation 30 seconds.
   (a) What is the probability that you will arrive before the bus actually does?
   (b) Given that you have actually arrived at 1:39 pm and the bus has not come till 1:42 pm, what is the probability that you have missed the bus?
   State your assumptions clearly. Your results can be in terms of Φ, standard normal distribution function. [5+5=10]
4. For $s \in \mathbb{N}$, show the Beta(r, s) distribution function is given by, for $0 < x < 1$,
$$\frac{x^r}{B(r,s)} \sum_{i=0}^{s-1} \binom{s-1}{i} (-1)^i \frac{x^i}{r+i}.$$
   [8]
5. Let (X, Y) be a point chosen uniformly at random from a square such that the origin and $(1,1)$ are two diametrically opposite vertices. If R is the distance of (X, Y) from the origin, find the density of R. [12]
6. Pick n points uniformly at random between 0 and 1. Let R be the distance between the points on the farthest right and the farthest left. Find $\mathrm{E}[R]$. [10]
7. Let X be distributed as Poisson random variable with parameter λ. Show that
$$\mathrm{P}[X \le \lambda/2] \le (2/e)^{\lambda/2}.$$
   [10]

INDIAN STATISTICAL INSTITUTE

End-Semestral Examination: 2018-19

Course Name: B. Stat. 1st year

Subject: Numerical Analysis

Date: 03. 05. 2019

Full Marks: 100 Duration: 3 hrs.

Answer as much as you can. Maximum marks you can get is 100.

1. a) Express -4.25 as a floating point number using the IEEE - 754 single precision format.

b) Let $f(x)$ be a real-valued function defined on the interval $[a,b]$ and k times differentiable in (a, b). If $x_0, x_1, \ldots, x_k$ are k distinct points in $[a, b]$, then prove or disprove the following statement:

"There exists $\xi \in (a,b)$ such that $f[x_0,\ldots,x_k] = \dfrac{f^{(k)}(\xi)}{k!}$",

where $f[x_0,\ldots,x_k]$ denotes the k-th order Newton's divided difference.

c) Consider the matrix:

$$A = \begin{bmatrix} 3 & 0 & 0 \\ -4 & 6 & 2 \\ 16 & -15 & -5 \end{bmatrix}$$

Estimate the dominant eigenvalue and the corresponding eigenvector of the matrix by using 4 iterations of the power method starting from the initial guess for the eigenvector as $x_0 = (1, 0.5, 0.25)^T$ Show all your steps and calculations.

d) Consider the following matrix:

$$A(\alpha) = \begin{bmatrix} 0.1\alpha & 0.1\alpha \\ 1.0 & 2.5 \end{bmatrix}.$$

If possible, find a value of α for which the condition number of the matrix $A(\alpha)$ is minimized when the maximum row norm is used.

2+5+8+5 = 20

2. a) Apply the Gerschgorin's theorem to the following matrix and determine the intervals in which the eigenvalues lie:

$$A = \begin{bmatrix} 0.5 & 0 & 0.2 \\ 0 & 3.15 & -1 \\ 0.57 & 0 & -7.43 \end{bmatrix}.$$

b) Estimate all eigenvalues of the following matrix by using the QR decomposition based method. Show all your steps:

$$A = \begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}.$$

6+14 = 20

3. a) Perform the Singular Value Decomposition (SVD) on the matrix and show the necessary steps:

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

b) Compute the Moore-Penrose pseudo-inverse of A by using the SVD obtained in Part (a).

c) Tri-diagonalize the following matrix by using Householder reflections:

$$A = \begin{pmatrix} 4 & 1 & -1 & 2 \\ 1 & 4 & 1 & -1 \\ -1 & 1 & 4 & 1 \\ 2 & -1 & 1 & 4 \end{pmatrix}:$$

9+4+7 = 20

4. a) Consider the problem of minimizing the following function:

$$f(x_1, x_2) = x_1^2 + 2x_2^2 - 2x_1x_2 - 2x_2 + 2x_1.$$

i) Compute the optimal solution vector $x^* = [x_1^*, x_2^*]^T$ to this problem. Is the optimal solution unique? Give reasons for your answer.

ii) Suppose the Newton's second order method is used to minimize the same function $f(x_1, x_2)$ in Q. 4(a) (i). Starting from the point $x^0 = [2,2]^T$, hand trace the iterations of the Newton's algorithm either till convergence or up to 3 iterations, whichever occurs earlier.

b) Consider the following system of non-linear equations:

$$\begin{cases} y = e^x \\ x^2 + y^2 = 1 \end{cases}.$$

i) Show graphically that this system has a solution near to $(x, y) = (0.7, 0.7)$.

ii) Use one iteration of the Newton's method (for system of non-linear equations) to find a better approximation of the solution starting with the initial point $(x, y) = (0.7, 0.7)$.

(2+1+3)+6+(3+5) = 20

PTO

5. a) Consider minimization of the function:

$$f(x_1, x_2) = -12x_2 + 4x_1^2 + 4x_2^2 + 4x_1x_2$$

by the method of gradient descent. Let us start with the initial solution $\boldsymbol{x}^0 = [0, 1]^T$. Hand trace the gradient descent algorithm with exact line search to determine the optimal learning rate parameter at every iteration. Continue up to the first 4 iterations and report the norm of the final solution gradient. Show all your work.

b) Use the Lagrange's multiplier method to find the coordinates of a point on the plane $2x + 3y - z = 5$ which is nearest to the origin $\mathbb{R}^3$.

c) Is the following function $f\colon \mathbb{R}^3 \to \mathbb{R}$ convex? Justify.

$$f(x_1, x_2, x_3) = 3x_1^2 - 4x_1x_2 + x_2^2 + x_3^2$$

12+5+3 = 20

6. a) Consider the quadrature rule of the form where w_1, w_2, and x_2 are such that the rule has the highest possible degree of accuracy:

$$\int_0^1 f(x)\,dx = w_1 f(0) + w_2 f'(x_2).$$

Approximate the following integral with a quadrature rule of the above form and derive the numerical value. What is the percentage relative error in this approximation?

$$\int_0^1 \frac{1}{x^2 + e^x}\,dx$$

b) Can you determine an approximation of the non-trivial solution of the following initial value problem by using the Euler's method? Jusify your answer.

$$y' = y^\alpha, \ \alpha < 1, \ y(0) = 0$$

6+4 = 10

INDIAN STATISTICAL INSTITUTE

Back Paper Examination: 2018-19

Course Name: B. Stat. 1st year

Subject: Numerical Analysis

Date: 08.07.2019

Full Marks: 100 Duration: 3 hrs.

Answer all the questions.

1. a) Suppose we wish to approximate values of $f(x) = \cos(x)$ by interpolating from the following table of values:

x	0	0.2	0.4	0.6
$\cos x$	1.000	0.980	0.921	0.825

(i) Construct the interpolating Lagrange polynomial $P(x)$ that agrees with all the tabulated values, and use it to estimate cos(0.45).

(ii) Give a rigorous error bound for your answer of part (i).

b) Compute the condition number of the matrix given below (relative to $\|\ \|_\infty$) for the matrix and show all the steps of your work:

$$\begin{bmatrix} 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \\ 1/4 & 1/5 & 1/6 \end{bmatrix}.$$

or

Consider the following tri-diagonal matrix:

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

Locate the second largest eignevalue of the matrix correct up to two decimal places by using the Sturm sequence property.

(6+6)+8 = 20

PTO

2. Use the Householder construction to produce the QR factorization of the following matrix and show all your work:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 2 & -4 & 5 \end{bmatrix}$$

20

3. a) Approximate the solution of the system of linear equations:

$$\begin{cases} 3x_1 - x_2 + x_3 = 1, \\ 3x_1 + 6x_2 + 2x_3 = 0, \\ 3x_1 + 3x_2 + 7x_3 = 4, \end{cases}$$

using 3 iterations (starting with $x_1 = 0, x_2 = 0$ and $x_3 = 0$) of

(i) Jacobi iterative method

(ii) Gauss-Seidel iterative method

(b) Prove that, for the system of equations shown in part (a), Jacobi iterations converge to the exact solution.

(6+6)+8 = 20

4. Use the Rayleigh quotient iterations to find the dominant eigenvalue of the following matrix and show the details of your computation steps:

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

10

5. a) Determine the coefficients in the quadrature formula:

$$\int_0^{2h} x^{-\frac{1}{2}} f(x)dx = (2h)^{\frac{1}{2}}(w_0 f(0) + w_1 f(h) + w_2 f(2h)).$$

b) Give an initial guess x_0 for which the Newton-Raphson method fails to obtain the real root for the equation $\frac{1}{3}x^3 - x^2 + x + 1 = 0$. Justify why it fails.

PTO

c) Consider the following initial value problem: $\begin{cases} y' = y^2 \\ y(0) = 1 \end{cases}$.

Use two iterations of the Runge-Kutta 4th order method to approximate $y(0.2)$. 7+6+7 = 20

6. You are given a quadratic polynomial $f(x_1, x_2, x_3)$:

$$f(x_1,x_2,x_3) = 2x_1^2 - 2x_1x_2 - 4x_1x_3 + x_2^2 + 2x_2x_3 + 3x_3^2 - 2x_1 + 2x_3$$

i) Write the polynomial $f(x_1,x_2,x_3)$ in the form $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x}$, where $\boldsymbol{x} = [x_1, x_2, x_3]^T$, $\boldsymbol{A}$ is a real symmetric matrix, and $\boldsymbol{b}$ is some constant vector.

ii) Find the point $\boldsymbol{x}^* = [x_1^*, x_2^*, x_3^*]^T$ where $f(x_1,x_2,x_3)$ attains an extremum or stationary value.

iii) Is this point $\boldsymbol{x}^* = [x_1^*, x_2^*, x_3^*]^T$ a minimum, maximum, or saddle point of some kind? Justify your answer with suitable optimality tests.

3+3+4 = 10

or

Approximate the integral:

$$\int_0^{0.5} \frac{2}{x-4}\, dx$$

using the (non-composite) Simpson's rule. Give a rigorous error bound on this approximation.

4+6 = 10

INDIAN STATISTICAL INSTITUTE

Backpaper Examination: 2018–2019
B. Stat. (Hons.) 1st Year. 2nd Semester
Vectors and Matrices II

Date: July 09, 2019 Maximum Marks: 100 Duration: 3 hours

- Answer all the questions.
- **You should present all your arguments while answering a question.**
- You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.

---

1. If $\mathbf{A}$ and $\mathbf{D}$ are square and $\mathbf{A}$ non-singular, show that

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}| \cdot |\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}|. \qquad [12]$$

2. Obtain $\mathbf{P_A}$ where $\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 3 & -2 \\ -2 & 1 & -3 \end{bmatrix}$, where $\mathbf{P_A}$ is the orthogonal projector into $\mathscr{C}(\mathbf{A})$. [12]

3. Show that every $n \times n$ matrix $\mathbf{A}$ is similar to an upper triangular matrix over $\mathbb{C}$. [12]

4. Let $\mathbf{A}$ be semi-simple and let $\mathbf{A} = \sum_{i=1}^{k} \alpha_i \mathbf{E}_i$ be the spectral form of $\mathbf{A}$. Then prove that $\mathbf{B}$ commutes with $\mathbf{A}$ only if $\mathbf{B}$ commutes with $\mathbf{E}_i$ for $i = 1, \ldots, k$. [12]

5. Let $\mathbf{A}$ be a real symmetric matrix.

   (a) If $\mathbf{A}^k = \mathbf{I}$ for some positive integer k, show that $\mathbf{A}^2 = \mathbf{I}$.

   (b) If the eigenvalues of $\mathbf{A}$ are all positive and if $\mathbf{A}^k = \mathbf{I}$ for some positive integer k then show that $\mathbf{A} = \mathbf{I}$. [8+4=12]

[P.T.O.]

6. Find the rank and signature of the quadratic form: $\sum_{i,j=1}^{n}(x_i - x_j)^2$ [12]

7. Suppose **A** and **C** are non-negative definite matrices of the same order. Show that $\mathscr{C}(\mathbf{A}+\mathbf{C}) = \mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{C})$. [12]

8. Let **B** be positive definite and $\mathbf{u} \neq \mathbf{0}$. Show that

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{u}^{\mathrm{T}}\mathbf{x})^2}{\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}} = \mathbf{u}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{u}$$

and that the maximum is attained at $\mathbf{x}_0$ if and only if $\mathbf{x}_0$ is a scalar multiple of $\mathbf{B}^{-1}\mathbf{u}$. [8+8=16]

INDIAN STATISTICAL INSTITUTE

Analysis 2 (Math101B) : B. Stat 1st year
Back Paper Examination: 2018-19
,2019

Date: 10/07/2019

Full Marks: 100

Maximum Time: 3 hrs.

In the following $C[a,b]$ *stands for the set of real valued continuous functions on* $[a,b]$.

(1) Let $f \in C[0,1]$ and for each $n \in \mathbb{N}$, $c_i \in \left[\frac{i-1}{n}, \frac{i}{n}\right]$, $1 \leq i \leq n$ is given. Prove that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(c_i) = \int_0^1 f(x)dx.$$

[6]

(2) Let $g : [0,2] \to [0,2]$ be a decreasing function and consider the double sequence

$$f(p,q) = g\left(\frac{p+q}{pq}\right).$$

Does the double limit $\lim_{p,q\to\infty} f(p,q)$ exist. Justify your answer. [6]

(3) Let f and g be two continuous 2π-periodic real valued functions. Assume that there exists an open interval $I \subset (\pi,\pi)$ such that f and g agree on I. If for some $\theta \in I$

$$f(\theta) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{in\theta},$$

then prove that

$$g(\theta) = \sum_{n=-\infty}^{\infty} \hat{g}(n)e^{in\theta}.$$

[6]

[P.T.O.]

(4) Evaluate

$$\lim_{n\to\infty}\left[\left(1+\frac{1}{n}\right)\left(1+\frac{2}{n}\right)\dots\left(1+\frac{n}{n}\right)\right]^{1/n}.$$

[8]

(5) Prove that the sequence of functions

$$f_n(x) = nx^n(1-x),$$

converges uniformly on $[0,1]$. [8]

(6) For $x \in (0,\infty)$, define

$$f(x) = \int_x^{x+1} \sin(t^2)dt.$$

Prove that $|f(x)| \leq 1/x$, $x > 0$. [8]

(7) Determine whether the following improper integral converges

$$\int_0^\infty \frac{\sin(1/x)}{x}dx.$$

[10]

(8) Given $g \in C[0,1]$ consider the following sequence of functions

$$f_n(x) = x^n g(x), \quad x \in [0,1].$$

Prove that $\{f_n\}$ converges uniformly on $[0,1]$ if and only if $g(1) = 0$. [10]

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

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

for every sequence $\{x_n\} \subset I$ such that $x_n \to x \in I$. Show by an example that the conclusion may fail if the convergence is not uniform. [12]

(10) Prove that

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{1+x^n}, \quad x > 1,$$

is a continuous function on $(1,\infty)$. If f differentiable? Justify your answer. [12]

(11) (a) Compute the Fourier coefficients of the function $f(\theta) = |\theta|$, $\theta \in [-\pi,\pi]$.

(b) Prove that

$$\sum_{k=0}^{\infty} \frac{1}{(2k+1)^4} = \frac{\pi^4}{96}.$$

[8+6=14]

Probability Theory II
B. Stat. Ist Year Semester 2
Indian Statistical Institute

Backpaper Examination
Time: 3 Hours Full Marks: 100
Date: 11th July 2019

This is an OPEN NOTE examination. You may use any handwritten notes that you have brought with you, but you are not allowed to share them. Also no notes in printed, photocopied, cyclostyled or any other form (including electronic storage) is allowed. Calculators and any other computing devices are strictly not allowed.

1. Let T be a positive random variable with continuous density f, such that
$$G(t) := \mathrm{P}[T > t] > 0 \quad \text{for all } t > 0.$$
The hazard rate of such a random variable T is defined as
$$\lambda(t) := \lim_{h\to 0} \frac{1}{h}\mathrm{P}[T \in (t, t+h)|T > t], \quad \text{for } t > 0.$$
Show that $G(t) = \exp(-\int_0^t \lambda(u)du)$. Hence or otherwise show that only exponential distribution has constant hazard rate.
[Quote all results clearly and completely that you may use and check the conditions therein.] [7+5=12]

2. Pick n points uniformly at random between 0 and 1. Let R be the distance between the points on the farthest right and the farthest left. Find density of R. [14]

3. If Θ is uniformly distributed over $(0, 2\pi)$, show that $\cos 2\Theta$ and $\cos\Theta$ have same distribution. If X and Y are i.i.d. standard normal random variables, find the density of the random variable
$$\frac{X^2 - Y^2}{\sqrt{X^2 + Y^2}}.$$
If you use any property of standard normal random variables, state them clearly. [5+7=12]

4. A straight stick is broken into three pieces by snapping it at two points chosen at random. What is the probability that those three pieces will form a triangle? [14]

5. Let X and Y be two random variables satisfying $\mathrm{E}[(X-Y)^2] = 0$. <u>*Use Chebyshev inequality*</u> to show that $\mathrm{P}[X = Y] = 1$. Hence or otherwise show the following: Let X and Y have common mean and variance 0 and 1 respectively, while the correlation is -1. Then show that, with probability 1, $X + Y = 0$. [6+6=12]

6. Let (X, Y) be chosen uniformly at random from the triangle formed by the origin, $(0, 2)$ and $(2, 0)$. Find $\mathrm{P}[Y > 1|X = x]$ for different values of x. [12]

7. Let X and Y be i.i.d. Exponential(λ) random variables. Define $Z = \min(X,Y)/\max(X,Y)$. Without any explicit calculation of the distribution function, show that the distribution function of Z is free of λ. Identify the region in the first quadrant of the two dimensional plane which corresponds to the event $[Z \le z]$, for $0 < z < 1$. Hence, calculate the distribution function of Z. [3+4+5=12]

8. Let X be a nonnegative random variable. Prove that
$$\mathrm{E}[X^2] = 2\int_0^\infty x\,\mathrm{P}[X > x]dx.$$
[12]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination : Semester II (2018-19)

B. Stat 1st Year

Statistical Methods - II

Date: 12. 07. 2019 Maximum marks: 100 Duration: 3 hours.

You may use any result proved in class after clearly stating it. You may use a calculator.

1. In the multiple regression model with an intercept, assume that there are p independent variables which have mean 0.

   (a) Derive expressions for the means and variances of the (i) least-square estimates, (ii) the fitted values and (iii) the residuals. [10]

   (b) Show that the fitted values and the residuals are uncorrelated. [5]

   (c) Derive the expression for different sum of squares in the ANOVA decomposition in terms of quadratic forms and find their expected values. [10]

2. Express the $(n-1)$-th order error standard deviations in terms of the lower order error standard deviations. [6]

3. Derive the expression for the multiple correlation coefficient. [8]

   Show that this correlation coefficient

   (a) is necessarily between 0 and 1. [5]

   (b) is numerically not less than any partial correlation coefficient of any order involving the dependent variable. [6]

4. Suppose X has the following distribution function:

$$F(x) = 1 - exp(-2x^2), \quad 0 < x < \infty.$$

   (a) Give an algorithm based on the inverse transform method for simulating a random observation from this distribution. [5]

   (b) Give an algorithm based on the rejection method for the same problem. Choose an appropriate value of the parameter for the distribution that you choose for this purpose so that the efficiency of this algorithm is maximized. [15]

[P.T.O.]

5. In an ecological study of the feeding behaviour of birds, the number of hops between flights was counted for several birds. For the data given below, see whether an appropriate geometric distribution can give a good fit.

Number of hops	Frequency
1	48
2	31
3	20
4	9
5	6
6	5
7	4
8	2
9	1
10	1
11	2
12	1

(a) Check whether the method of moments and method of ML give same estimates of the parameter. [15]

(b) Calculate the chi-square statistic for goodness of fit. Comment on the results. [15]