CHENNAI MATHEMATICAL INSTITUTE

M.Sc. Data Science Entrance Examination 2024

IMPORTANT INSTRUCTIONS

- Part (A) consists of multiple-choice questions. There may be multiple correct choices. You have to select all the correct options and no incorrect option to get full marks. **There is no partial credit.**
- For questions in part (A), you have to provide the answers on the computer. You only have to choose the appropriate answer(s) from the choices provided. If the answer is parts (a) and (c), choose only (a) and (c).
- For questions in part (B), you have to write your answer with a short explanation in the space provided for the question.
- **Part(A) will be used for screening.** Part (B) will be graded only if you score a certain minimum in part (A). However your scores in both parts will be used while making the final decision.
- For numerical answers, the following forms are acceptable: fractions, decimals, symbolic e.g.: $\binom{n}{r}$, ${}^nP_r$, $n!$ etc.

Notation and terminology

- A function f from a set A to a set B is said to be **injective (or one-to-one)** if $f(x) = f(y)$ implies $x = y$ for all $x, y \in A$;
- f is said to be **surjective (or onto)** for every $y \in B$ there exists $x \in A$ such that $f(x) = y$;
- f is said to be **bijective** if it is both injective and surjective;
- f is said to be **invertible** if there exists a function g from B to A such that $f(g(y)) = y$ for all $y \in B$ and $g(f(x)) = x$ for all $x \in A$.
- For a matrix A, A^T denotes the transpose of A. For a square matrix A, $|A|$ denotes the determinant of A and trace(A) denotes the *trace* of A — namely the sum of the diagonal elements of A.
- A *diagonal matrix* is a square matrix D with all off-diagonal entries equal to zero i.e. $d_{ij} = 0$ for all $i \neq j$.
- An *upper triangular matrix* is a square matrix A for which all entries below the diagonal are zero, i.e. $a_{ij} = 0$ for $i > j$.
- A *symmetric matrix* is a square matrix S for which $s_{ij} = s_{ji}$ for all $i \neq j$.

Part (A) - Multiple-choice questions

1. Consider the following positive integers:

$$A = 1234679580, B = 1234789560.$$

Which of the following statements are true?

(a) Both A and B are divisible by 36.
(b) Both A and B are divisible by 72.
(c) Only B is divisible by 120.
(d) The number A is divisible by 80.

2. Which of the following statements are true?

(a) Let $f(x) = x^3 + x - 1$ for $x \in \mathbb{R}$. Then the equation $f(x) = 0$ has at least one root in $[-1, 0)$.
(b) Let $f(x) = x^3 + x - 1$ for $x \in \mathbb{R}$. Then the equation $f(x) = 0$ has at least one root in $[0, 1)$.
(c) Let $f(x) = x^3 + x + 1$ for $x \in \mathbb{R}$. Then the equation $f(x) = 0$ has at least one root in $[0, 1)$.
(d) Let $f(x) = x^3 + x + 1$ for $x \in \mathbb{R}$. Then the equation $f(x) = 0$ has at least one root in $[-1, 0)$.

3. 26 children participated in a chess tournament. A child got two points for winning, zero for losing and one point for a draw. Each child played against every other child. After the tournament was over no child had an odd score. There were no draws in the entire tournament and no two players had the same score. For convenience assume that the children are named $A, B, C, \ldots, Z$, by the 26 letters of the English alphabet in increasing order of scores, so A has lowest score, Z has highest score. Thus we have $A < B < C < \cdots < X < Y < Z$. Pick all statements which are true.

(a) K lost to Q
(b) K lost to B
(c) M lost to N
(d) If L lost to M then N lost to M.

4. 14 teams participate in a volleyball tournament. Each team plays the other exactly once. There are no draws. Assume the teams are labeled by j, $1 \leq j \leq 14$. Let x_j denote the number of games team j wins and y_j the number of games that team j lost. Pick the correct alternative(s):

(a) $\sum_j x_j^2 = \sum_j y_j^2$.
(b) $\sum_j x_j^2 > \sum_j y_j^2$.
(c) $\sum_j x_j = \sum_j y_j$.
(d) $\sum_j |x_j| = \sum_j |y_j|$.

5. Which of the following statements is/are true for real numbers x, y?

(a) If $x^2 = y^2$ then $x = y$.
(b) If $x^3 = y^3$ then $x = y$.
(c) If $x < y$ then $x^2 < y^2$.
(d) If $x < y$ then $x^3 < y^3$.

6. Suppose that A is an $n \times n$ matrix with $n \geq 6$. For an $n \times 1$ vector b consider the equations

$$Ax = b, \text{ for an } n \times 1 \text{ vector } x \qquad (1)$$
$$A^T x = b \text{ for an } n \times 1 \text{ vector } x \qquad (2)$$

where A^T is the transpose of the matrix A. Which of the following statements are correct?

(a) If equation (**??**) admits a solution for all b, then A^{-1} exists.
(b) If equation (**??**) admits a solution for all b, then equation (**??**) also admits a solution for all b.
(c) If equation (**??**) admits a solution for some b, then A^{-1} exists.
(d) If equation (**??**) admits a solution for some b, then equation (**??**) also admits a solution for that b.

7. Which of the following inequalities are correct?

(a) $1 + x \leq e^x$ for all $x \in \mathbb{R}$
(b) $2^n \leq 2n!$ for all positive integers n
(c) $(1 + x^2)^n \leq (1 + x)^{2n}$ for all $x \in (0, \infty)$
(d) $\sin(x) \leq \tan(x)$ for all $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$

8. Which of the following statements are valid for all $n \times n$ matrices A, B:

(a) $(A^T A)^T = AA^T$.
(b) If A, B are invertible, then inverse of AB is $A^{-1}B^{-1}$.
(c) $(A + B)^T = A^T + B^T$
(d) $Ax = Bx$ for some $n \times 1$ vector x implies that $Ay = By$ for all $n \times 1$ vectors y.

9. Let x be a variable that takes real values, and let $f(x), g(x), h(x)$ be three polynomials with real-valued coefficients. Further, let $f(x)$ be a polynomial of degree 1, $g(x)$ a polynomial of degree 2, and $h(x)$ a polynomial of degree 3. Which of the following statements is/are **always true** for *any three* polynomials of these types?

(a) The graphs of $f(x)$ and $g(x)$ intersect at one or more points.
(b) The graphs of $f(x)$ and $h(x)$ intersect at one or more points.
(c) The graphs of $g(x)$ and $h(x)$ intersect at one or more points.

10. A spherical ball of ice of radius $20m$ is dropped in a vat of hot water. The ice melts in such a way that (i) the shape of the ball remains spherical, and (ii) the radius of the ball decreases at a constant rate of $0.5ms^{-1}$. At what rate does the *volume* of the ice ball decrease, when the radius of the ball is $15m$?

(a) $100\pi m^3 s^{-1}$
(b) $225\pi m^3 s^{-1}$
(c) $450\pi m^3 s^{-1}$
(d) $600\pi m^3 s^{-1}$

11. Let x be a real-valued variable. What is the range of the function $f(x) = \sin^2 x - \sin x + 2$?

(a) $[0, 2]$
(b) $[1, 2]$
(c) $[1, 4]$
(d) $[\frac{7}{4}, 4]$

12. In the following code, A is an array indexed from 0 whose elements are all positive integers, and n is the number of elements in A. It is given that n is at least 2. The operator * denotes multiplication.

```
function foo(A,n) {
  if A[0] > A[1] {
    first = A[0];
    second = A[1];
  } else {
    first = A[1];
    second = A[0];
  }

  for i from 2 to (n-1) {
    if A[i] > second {
      if A[i] > first {
           second = first;
           first = A[i];
      } else {
           second = A[i];
      }
    }
  }
  return(first * second);
}
```

If A = [15, 7, 16, 12, 17, 14, 16, 4, 13, 12], what will foo(A, 10) return?

(a) 28
(b) 144
(c) 180
(d) 272

13. In the following code, `A` is an array indexed from `0` whose elements are all positive integers, and `n` is the number of elements in `A`.

```
function foo(A,n) {
  max = 0;
  curr = 0;

  for i from 1 to (n-1) {
    if A[i] > A[i-1] {
      curr = curr + 1;
      if curr > max {
        max = curr;
      }
    } else {
      curr = 0;
    }
  }
  return(max+1);
}
```

If `A = [1, 3, 5, 2, 4, 7, 6, 8]`, what will `foo(A, 8)` return?

(a) 2
(b) 3
(c) 4
(d) 5

14. A prominent newspaper conducted a survey among a carefully chosen sample of 1,00,000 voters under the age of 25. A randomly chosen subset of 60% of this sample were asked the following question: *"Have you ever voted in a general election?"* The remaining 40% were asked the following, different question: *"Is it true that you have* never *voted in a general election?"*

The editor in charge of collating the data from the survey, accidentally deleted some of the contents of the survey database. The only information that survived this mishap was that a total of 54,000 respondents replied YES to the question that they got. Specifically, it is not known how many of these YES answers correspond to the first question, and how many to the second.

Assume that the sampling was done properly and that the respondents answered truthfully. What estimate can we derive from just the total number of YES answers, about the percentage of voters under the age of 25 that have voted in some general election?

(a) At most 30%.
(b) More than 30%, but less than 60%.
(c) At least 60%, but less than 75%.
(d) No such estimate can be drawn from just the total number of YES answers.

15. Let f, g be two real valued functions defined on the set of real numbers. Which of the following statements are true?

   (a) If $f(x)$ is continuous and $g(x)$ discontinuous then the function $f(x)+g(x)$ is necessarily discontinuous.

   (b) If $f(x)$ is continuous and $g(x)$ discontinuous then the function $f(x) \cdot g(x)$ is necessarily discontinuous.

   (c) If the functions $f(x), g(x)$ are discontinuous at a point c in the domain the function $f(x) + g(x)$ is also discontinuous at c.

   (d) If the functions $f(x), g(x)$ are discontinuous at a point c in the domain the function $f(x) \cdot g(x)$ is also discontinuous at c.

16. In the following code, `A` is an array indexed from `0` whose elements are all positive integers, `n` is the number of elements in `A`, and `x` is a positive integer. The call `abs(z)` returns the absolute value of integer `z`.

```
function foo(A,n,x) {
  max = 0;

  for i from 0 to (n-1) {
    for j from (i+1) to (n-1) {
      diff = abs(A[i]-A[j]);
      if (diff >= x) and (diff >= max) {
        max = diff;
      }
    }
  }

  return(max);
}
```

   If `A = [10, 8, 10, 4, 10, 7, 1, 2, 2, 9]`, what will `foo(A, 10, 5)` return?

   (a) 2

   (b) 4

   (c) 6

   (d) 9

17. One day, Captain Haddock receives a mysterious letter with a confusing paragraph. Captain Haddock and Tintin are investigating the matter. There are two possible suspects: Professor Calculus and Thomson & Thompson. Based on their past experience:

   - The probability that Professor Calculus sends a letter is 60%, while the probability that Thomson & Thompson send a letter is 40%.
   - When Professor Calculus sends letters, there is an 80% probability that the letter contains a confusing paragraph.
   - When Thomson & Thompson send letters, there is a 5% probability that the letter contains a confusing paragraph.

   What is the probability that the letter was sent by Professor Calculus?

   (a) 0.96

   (b) 0.80

   (c) 0.50

   (d) 0.48

18. Let f be a function on the positive real numbers such that $f(xy) = f(x) + f(y)$. If $f(2024) = 2$ then which of the following statement(s) is/ are true?

    (a) $f(\frac{1}{2024}) = 1$
    (b) $f(\frac{1}{2024}) = -1$
    (c) $f(\frac{1}{2024}) = -2$
    (d) $f(\frac{1}{2024}) = 2$

The following description is for questions 19 and 20.

A perfect shuffle of a deck of cards divides the deck into two equal parts and then interleaves the cards from each half, starting with the first card of the first half.

For instance, if we shuffle a deck of cards containing 10 cards arranged $[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$ we first create two equal decks with cards $[1, 2, 3, 4, 5]$ and $[6, 7, 8, 9, 10]$ and then interleave them to get a new deck $[1, 6, 2, 7, 3, 8, 4, 9, 5, 10]$.

19. We shuffle the deck $[3, 6, 11, 4, 7, 9, 2, 8, 5, 10, 12, 1]$. What are the neighbours of 4 after the shuffle?

    (a) 1
    (b) 5
    (c) 10
    (d) 12

20. After shuffling once we obtain a deck with cards $[15, 6, 4, 1, 13, 9, 8, 7, 12, 11, 5, 10, 3, 14, 2, 16]$. What were the neighbours of 2 before the deck was shuffled?

    (a) 3
    (b) 6
    (c) 7
    (d) 14

Part (B) - Short-answer questions

For questions in part (B), you have to write your answer with a short explanation in the space provided for the question in your answer sheet. If you need more space, you may continue on the pages provided for rough work. Any such overflows must be clearly labeled.

Questions 1 and 2 are based on the following description.

Each round of a TV game show consists of ten questions. Before each round the host takes ten boxes and places prizes in nine of them, leaving one empty. The host then shuffles these boxes and labels them from 1 to 10. When the guest answers a question correctly, the host opens the corresponding box. If the box has a prize, the guest earns the prize. If the box is empty, the round ends and the guest gets to keep their earnings so far.

1. In the Easy Round, all questions are easy, and each prize is ₹1000. Guest Chatur knows all the answers. What is the expected earnings for Chatur in this round? Explain your answer.
2. To proceed to the next round, a guest must earn at least ₹7000 in the Easy Round. What is the probability that guest Chatur, who knows all the answers of the Easy Round, progresses to the second round? Explain your answer.

---

3. Find values of x and y that satisfy both the following equations:

$$\sqrt{\frac{x}{y}} + \sqrt{\frac{y}{x}} = \frac{5}{2}$$
$$\frac{x}{\sqrt{y}} + \frac{y}{\sqrt{x}} = \frac{9}{2}.$$

Questions 4 and 5 are based on the following description.

The following question appeared in a quiz:

> "*Write the code for a function* `SecondBest(A,n)` *that takes an array* `A` *and a positive integer* `n` *as arguments. The elements of* `A` *are all integers, and* `n` *is the number of elements in* `A`*. The call* `SecondBest(A,n)` *should return the second largest element in* `A`*. If* `A` *has no second largest element, then the function should return the special value* `None`*.*"

A student submitted the code below as the answer to this question. In the code the array `A` is indexed from `0`.

```
function SecondBest(A, n) {
  if n == 1 {
    return(None);
  }

  first = A[0];
  second = A[1];

  for i from 2 to (n-1) {
    if (A[i] >= first) and (A[i] >= second) {
      second = first;
      first = A[i];
    } else {
      if A[i] >= second {
        second = A[i];
      }
    }
  }

  if first != second {
    return(second);
  } else {
    return(None);
  }
}
```

This answer turned out to be wrong; this function gives the correct answer for some valid inputs, and wrong answers for other valid inputs. Answer the next two questions about this function.

4. What do the following function calls return? Briefly explain your answer.

   (a) `SecondBest([2,2,3], 3)`

   (b) `SecondBest([3,2,2], 3)`

5. Give one example of an input array `A` with exactly 3 elements for which the call `SecondBest(A, 3)` returns a **wrong** answer. What is this wrong answer? What is the correct answer?

---

6. Let F be the set of points on the plane with coordinates (x, y) such that $||x| - |y|| + |x| + |y| = 4$. What is the number of points in F with integer coordinates?

7. We select three points at random on the circumference of a circle. What is the probability that ΔABC contains the centre O in the interior?

8. Let α be a fixed real number and let f be a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ defined as

$$f(x, y) = (x \cos\alpha - y \sin\alpha, x \sin\alpha + y \cos\alpha).$$

   Write down an expression for $f^{10}(x, y)$, where f^{10} denotes the function obtained by composing f with itself 10 times.

9. "*I've defeated more than* 10 *FIDE masters,*" a chess player boasted. "*Surely, it must be fewer than* 10," said the referee. "*Well, I suppose it was at least one,*" said the Grand Master. If exactly one of them spoke the truth, how many FIDE masters did the chess player actually defeated?

10. A supplier of art material has four reams of handmade paper, three boxes of acrylic colors and two printing blocks. The two artists in the shop want to buy one item each, but insist on having the same kind of art material. How many items does the supplier have to take out to be sure that the artists' demand is met?

11. In a farmer's stable three animals, a donkey, a cow and a horse had to share two types of feed bags, one containing hay and the other containing grain, as follows:

    - If the donkey ate grain then the horse ate what the cow ate.
    - If the horse ate grain then then donkey ate what the cow did not eat.
    - If the cow ate hay, then the donkey ate what the horse ate.

    Among all the assignments of feedbags that satisfy the above condition which animal always gets to eat from the same feedbag?

12. There are 7 elevators in a large shopping mall, each stopping at ground floor and at most six other floors. If at least 3 elevators stop at each floor and if it is possible to go from any floor to any other floor without changing elevators, what is the maximum number of floors in the mall?

13. Consider the following expression

$$\sqrt{x+2\sqrt{x-1}}+\sqrt{x-2\sqrt{x-1}}.$$

    Find two real numbers a, b such that the above expression is constant for $a \leq x \leq b$.

14. You are given a strange, analogue wall clock whose hour and minute hands are identical. Both the hands move continuously and there is no second hand. How many times are there from noon to midnight when it is not possible to tell what time is it by looking at the clock at that instant? For example, at 06 : 00 pm one can be sure that the upper hand is the minute hand, for otherwise at 12 : 30 pm the top hand is between 12 and 1. However, a bit after 1 : 15 pm and 3 : 06 pm the clock looks identical and you wouldn't be able to tell the exact time.

15. Does there exist a polynomial $q(x)$ with integer coefficients such that $q(1) = 2$ and $q(3) = 5$? Given an example if there is one. Justify, if there is not.

16. Find the number of all 4-digit natural numbers formed with exactly two distinct digits.

17. An electronic card shuffling machine always rearranges the cards in the same way relative to the order in which they are placed in it. One iteration of shuffling means that the cards are placed in the machine and they are taken out after the rearrangement. Two iterations means that you place the cards in the machine immediately after the first iteration. The ace through the king of hearts are arranged in order with the ace on top and the king at the bottom. After 2 iterations the order of the cards, from top to bottom, is

$$10, 9, Q, 8, K, 3, 4, A, 5, J, 6, 2, 7.$$

    What will be the order after 13 iterations?

Questions 18–20 are based on the following description.

In June 2017, a cyberattack named NotPetya spread all over the world. Big companies like Marex, Merck, and FedEx's TNT Express lost a lot of money because of this attack.

Company	Pre-attack Average Monthly Revenue (USD million)	Post-Attack Average Monthly Revenue (USD million)
Marex	1000	700
Merc	800	480
TNT Expanse	500	250

Table 1: Financial impact of NotPetya attack on three major companies

18. Which company experienced the smallest percentage decrease in total revenue over a one month period relative to its estimated total revenue loss over the same period?

19. Calculate the overall percentage decrease in revenue across all three companies in one month period following the NotPetya attack.

20. Assuming the monthly revenue of Marex follows a Gaussian distribution with a standard deviation of 75, (i) what is the probability that Marex's revenue will be less than \$850 million in the pre-attack scenario, and (ii) what is the probability that Marex's revenue will be more than \$850 million in the post-attack scenario?

z	$P(Z \leq z)$
-2	0.0227
-1	0.1586
0	0.5000
1	0.8413
2	0.9772

Table 2: Note that $Z \sim N(0, 1)$