

CHENNAI MATHEMATICAL INSTITUTE
M.Sc. Data Science Entrance Examination 2026

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 $\text{trace}(A)$ 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. An insurance company offers a special policy covering two independent risk events in a policyholder's life over one year: a **medical emergency** and a **vehicle accident**. Based on historical actuarial data:

$$P(\text{medical emergency}) = \frac{1}{10}, \quad P(\text{vehicle accident}) = \frac{1}{20}$$

A claim is considered *valid* if **at least one** of the two events occurs during the year. Each policyholder pays an annual premium of **Rs. 4,800**, and if a valid claim is filed, the company pays out **Rs. 6,000** (no payout otherwise). Which of the following statements are correct?

- (a) The probability of valid claim is $\frac{29}{200}$.
 (b) A single policyholder renews the same policy for four consecutive years. The probability that she files a valid claim in exactly one of the four years equals

$$\binom{4}{1} \left(\frac{29}{200}\right)^3 \left(\frac{171}{200}\right)^1 \approx 0.0095$$

- (c) A family of four members each independently hold this policy for a year. The probability that exactly one of the four members files a valid claim equals

$$\binom{4}{1} \left(\frac{29}{200}\right)^1 \left(\frac{171}{200}\right)^3 \approx 0.3257$$

- (d) If 800 policyholders are enrolled under this policy in a given year, the expected profit of the insurance company for the year equals **Rs. 19,44,000**.

2. Consider the single variable, real-valued function $f(x) = x^3 - 3x$. Which of the following statements about f are true?

- (a) f is increasing for all x .
 (b) f has exactly 2 critical points.
 (c) f has exactly one inflection point.
 (d) f is bounded below.

3. There are 10 switches labeled 1 to 10, all initially turned OFF. You perform the following operation: at step k , you toggle (an ON switch is turned OFF and vice versa) every switch whose label is divisible by k . After performing 10 such steps, which switches are in ON position?

- (a) 1, 4, 9
 (b) 2, 4, 6, 8
 (c) 1, 4, 9, 10
 (d) 1, 3, 6, 9

4. Which of the following inequalities involving real valued functions are true?

- (a) $|\sin x| \leq |x|, \quad \forall x \in \mathbb{R}$
 (b) $e^{\sqrt{x}} \leq \sqrt{x}, \quad \forall x \geq 0$
 (c) $\ln(1+x) \leq x, \quad \forall x > -1$
 (d) $\tan x \geq x, \quad \text{for all } x \text{ where defined}$

5. Let S be the set of all functions from $\{1, 2, 3, 4\} \rightarrow \{1, 2, 3\}$. Which of the following are true?

- (a) The number of functions in S whose image has exactly 2 elements is $3 \cdot (2^4 - 2)$
 (b) The number of functions in S whose image has exactly 3 elements is $3^4 - 3 \cdot 2^4 + 3$
 (c) The number of injective functions in S is nonzero
 (d) The number of surjective functions in S with $f(1) = 1$ and $f(2) = 2$ is 5.

6. A box contains 3 red balls and 2 blue balls. Two balls are drawn one after another without replacement. Let A be the event that both balls are red, and B the event that the first ball is red. Which of the following are true?
- (a) $P(A) = \frac{3}{10}$
 - (b) $P(B) = \frac{3}{5}$
 - (c) $P(B^c | A^c) = \frac{2}{3}$ (here X^c is the event that X does not take place).
 - (d) $P(A | B) = \frac{3}{5}$
7. Five people, A, B, C, D, E sit in a row. Two particular people, A and B , refuse to sit next to each other. Which of the following are true?
- (a) The number of valid arrangements is $5! - 4! \cdot 2$
 - (b) The probability that A and B are not adjacent is $\frac{3}{5}$
 - (c) The number of arrangements (where A and B are not adjacent) in which A and D sit next to each other is 36
 - (d) The number of arrangements (where A and B are not adjacent) in which D and E sit next to each other is 48.
8. Let A be an $n \times n$ real matrix. Which of the following statements are always true?
- (a) If A is invertible, then the equation $Ax = b$ has a unique solution for every $b \in \mathbb{R}^n$.
 - (b) If $Ax = 0$ has only the trivial solution, then A is invertible.
 - (c) If $A^2 = 0$, then $A = 0$.
 - (d) If $AB = 0$ for some $n \times n$ matrix $B \neq 0$, then A is not invertible.
9. How many 3-element subsets of $\{1, 2, \dots, 8\}$ contain exactly one pair of consecutive integers?
- (a) 20
 - (b) 36
 - (c) 30
 - (d) 56
10. Let a and b be the outcomes of two independent rolls of a fair die. Define $X = \max\{a, b\}$. Which of the following are true?
- (a) $P(X = 6) = \frac{11}{36}$
 - (b) $P(X = 1) = \frac{1}{36}$
 - (c) $E[X] > 4$
 - (d) $P(X \leq 3) = \frac{8}{36}$
11. Starting with the number $n = 1$, we generate a sequence of numbers. In the second step we replace 1 by either $2 = 2 \times 1$ or $3 = 2 \times 1 + 1$. In general, replace n by either $2n$ or $2n + 1$ to get a new value for n . Which of the following numbers can be obtained as values of n in this fashion?
- (a) 7
 - (b) 10
 - (c) 15
 - (d) 2026

12. Arrange the numbers 1, 2, 3, 4, 5 in a row. We say a number i , $1 \leq i \leq 5$ is *well-placed* in the given arrangement if it appears in position i , starting from the left. For example, in the arrangement 4, 1, 3, 2, 5, the numbers 3 and 5 are well-placed. We are interested in arrangements where no number is well-placed. How many such arrangements exist?

- (a) 20
- (b) 44
- (c) 60
- (d) 45

13. Three statements are made:

- S1: “Exactly one of S2 and S3 is true.”
- S2: “S1 is false.”
- S3: “S2 is true.”

Which of the following are possible?

- (a) Only S1 is true.
- (b) Only S2 is true.
- (c) Only S3 is true.
- (d) Exactly two statements are true.

14. A subset (possibly empty) of $\{1, 2, 3, 4, 5, 6\}$ is chosen uniformly at random. What is the probability that the chosen subset contains at least one even number and at least one odd number?

- (a) $\frac{1}{2}$
- (b) $\frac{3}{4}$
- (c) $\frac{7}{8}$
- (d) $\frac{49}{64}$

15. A student writes the following code to compute the average of numbers $A[1], \dots, A[n]$:

```
sum = 0
for i = 1 to n:
    sum = sum + A[i]
    avg = sum/n
return avg
```

Which of the following statements about this piece of code are correct?

- (a) The code always returns the correct average.
- (b) The output depends on the order of the input values
- (c) In general, the code doesn't calculate the correct average.
- (d) The variable `avg` is unnecessarily updated.

16. In the following code, the operator % denotes the remainder after integer division.

```
function mystery(n) {  
    result = 0;  
    for i from 1 to n {  
        if (i % 2 == 0) {  
            result = result + i;  
        } else {  
            result = result - i;  
        }  
    }  
    return(result);  
}
```

What does `mystery(10)` return?

- (a) -5
- (b) 5
- (c) 10
- (d) 0

Questions (17) to (19) are based on the following code.

The arguments to the function `enigma(A, n)` in the code below are: (i) an integer array `A` indexed from 0, and (ii) the number `n` of elements in `A`.

```
function enigma(A, n) {  
    for i from 0 to n - 2 {  
        for j from 0 to n - 2 - i {  
            if (A[j] > A[j+1]) {  
                temp = A[j];  
                A[j] = A[j+1];  
                A[j+1] = temp;  
            }  
        }  
    }  
    return(A[0]);  
}
```

Answer the next three questions about this function.

17. What does `enigma([3, 1, 4, 2, 5], 5)` return?

- (a) 1
- (b) 3
- (c) 4
- (d) 5

18. Which of the following statements about the function `enigma(A,n)` are true?

- (a) After execution, the array `A` is sorted in non-decreasing order.
- (b) The function returns the smallest element of the array.
- (c) The function modifies the input array `A`.
- (d) The function breaks down if the array contains repeated elements.

19. Which of the following statements about the number of operations performed in `enigma(A,n)` are true? Note that assigning a value to a variable is not considered as an operation.
- (a) The number of comparisons performed is proportional to n^2 .
 - (b) The total number of operations depends on the initial order of the array.
 - (c) The function always performs the same number of comparisons for a given n .
 - (d) The function may terminate early if the array is already sorted.
20. The argument to the function `steps(n)` is a positive integer n . The operator `%` denotes the remainder after integer division.

```
function steps(n) {  
    count = 0;  
    while (n != 1) {  
        if (n % 2 == 0) {  
            n = n / 2;  
        } else {  
            n = 3 * n + 1;  
        }  
        count = count + 1;  
    }  
    return(count);  
}
```

What does `steps(6)` return?

- (a) 6
- (b) 8
- (c) 3
- (d) 4

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.

1. Let A be a 2×2 real matrix such that $A^2 = I$. Determine all possible values of $\text{trace}(A)$.
2. How many binary strings of length 8 contain exactly three 1s such that no two 1s are adjacent.

Questions (3) and (4) are based on the following information

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function that satisfies

$$f(x+y) = f(x) \cdot f(y), \quad \forall x, y \in \mathbb{R},$$

and that $f(0) = 1$. The following two questions are based on this information.

3. Show that for any rational number r , we have $f(r) = (f(1))^r$. (**Hint:** Consider the reduced expression $r = p/q$, where p, q are integers).
4. Assuming that $f(x) = (f(1))^x$ holds for all real x , determine all values of $f(1)$ for which f is strictly increasing.
5. Determine all natural numbers $n \geq 1$ such that

$$\sum_{k=1}^n k \text{ divides } \sum_{k=1}^n k^2.$$

You may express your answer as a sequence of integers, giving the first few integers n for which the statement is true.

Question (6) and (7) are based on the following information

A *list* $(a_1, \dots, a_n)$ is an arrangement of natural numbers $1, 2, \dots, n$ in a random order, with all orderings equally likely. We say that a position i is a *new maximum* if $a_i > a_j$ for all $j < i$. For example, in the list

$$(3, 1, 4, 2, 5)$$

positions 1, 3, 5 are new maxima. Note that position 1 is always a new maximum. Now answer the two questions below based on this information.

6. For a given i , determine the probability that position i is a new maximum.
7. What is the expected number of new maxima in a list?

Questions (8) and (9) are based on the following description.

In the pseudocode below, the function **Steps**(n) takes a positive integer n as input. For positive integers a and b , the expression $a \% b$ evaluates to the remainder obtained when a is divided by b .

```
function Steps(n) {
    count = 0;
    while (n > 1) {
        if (n % 2 == 0) {
            n = n / 2;
        } else {
            n = n - 1;
        }
        count = count + 1;
    }
    return(count);
}
```

8. Compute the value of **Steps**(13).
9. Let n be a positive integer with exactly k digits in its binary representation. What is the maximum possible value of **Steps**(n)? For which value(s) of n (expressed in terms of k) is this maximum attained?

Questions (10) and (11) are based on the following information.

A finite sequence of $n \geq 1$ integers $a_1, a_2, \dots, a_n$ is processed one by one. Two variables M and c are maintained as follows:

- Initially, $M = a_1$ and $c = 1$.
 - For each subsequent number a_i , if $a_i > M$, set $M = a_i$ and $c = 1$; if $a_i = M$, set $c = c + 1$; otherwise, do nothing.
10. For the input sequence 3, 1, 4, 1, 5, 9, 2, 6, 5, 9, write down the values of M and c after all elements have been processed.
11. What do the variables M and c represent at the end of processing? Justify your answer in one or two sentences.
12. Find the sum of all integers from 1 to 100 which are neither multiples of 3 nor multiples of 7.

Questions (13) and (14) are based on the following information.

Let p be a prime and n be a natural number such that

$$p \text{ divides } n^2 + n + 1.$$

13. Find all possible remainders of n^3 upon division by p .
14. Suppose $p > 3$. Show that the numbers $1, n, n^2$ leave distinct remainders upon division by p .
15. Find all two-digit positive integers that are equal to three times the sum of their digits.
16. At a gathering of 5 people (named A, B, C, D, E), each person records how many other people they shook hands with. Each number must be between 0 and 4. Find one possible set of 5 recorded numbers, with at least 2 repeated values, but not all values equal, and describe a possible configuration of handshakes.
17. An LLM has generated 3 deep learning models M_1, M_2, M_3 . Exactly one of these models is faulty. After running internal diagnostic tests each model produced the following output:
- (a) **M₁**: *Model M_2 is faulty.*
 - (b) **M₂**: *Model M_3 is faulty.*
 - (c) **M₃**: *Model M_1 is not faulty.*

You know that only faulty models produce false statements and only non-faulty models produce true statements. Which model is faulty? Explain.

18. Let A be an $n \times n$ real matrix such that

$$A^2 = A.$$

- (a) Show that if x is in the image of A , then $Ax = x$.
- (b) Show that every vector $v \in \mathbb{R}^n$ can be uniquely written as

$$v = u + w,$$

where $Au = u$ and $Aw = 0$.

19. Let X_1, X_2, X_3 be independent and identically distributed $\text{Uniform}(0, \theta)$ random variables with probability density function (pdf),

$$f(x) = \frac{1}{\theta}, \quad 0 < x < \theta,$$

where unknown $\theta > 0$. Let $T = \max(X_1, X_2, X_3)$.

- (a) Derive the cdf and pdf of T
- (b) Find $E(T)$ and $\text{Var}(T)$.

20. **Chebyshev's Inequality states:** For any random variable X with mean μ and variance σ^2 , and for any $k > 0$:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{equivalently,} \quad P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Let $X \sim \text{Exponential}(\lambda)$ with mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$. A student claims:

$$P\left(\left|X - \frac{1}{\lambda}\right| < \frac{2}{\lambda}\right) \geq \frac{3}{4}$$

Is this claim true? Verify by computing the exact probability and comment on what your answer reveals about Chebyshev's inequality.