CHENNAI MATHEMATICAL INSTITUTE
M.Sc. Data Science Entrance Examination 2023

IMPORTANT INSTRUCTIONS

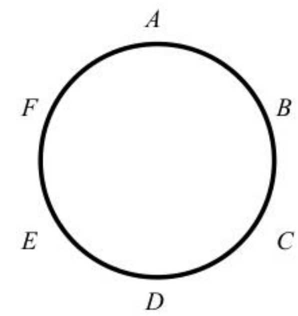
- Part (A) consists of multiple-choice questions. There may be multiple correct choices. You have to select all the correct options and no incorrect option to get full marks. **There is no partial credit.**
- For questions in part (A), you have to provide the answers on the computer. You only have to 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. Six children — Abhay, Bhavna, Charanjit, Divya, Enakshi and Farid — are sitting, in that order, around a circular table at a birthday party, as shown on the right. Each of them may be seated or standing.

   They play a game, as follows. The birthday girl's father calls out the initials of a pair of children seated next to each other. The children whose initials are called out then change their position from sitting to standing or vice versa.

   Initially, all children are seated. Which of the following announcements will result in an arrangement where Abhay and Enakshi are standing and the other four children are seated?

   (a) FA, DE, BC, CD, AB, FA
   (b) AB, BC, DE, EF, EF, DE
   (c) AB, CD, BC, EF, DE, EF
   (d) AB, CD, FA, BC, AB, DE

2. Consider the following single variable, real valued functions:

$$f(x) = \frac{x^4+1}{x^3-1}; \quad g(x) = \frac{x^3-x}{x^3+x}.$$

   Which of the following is/are true?

   (a) The function f has a removable discontinuity at 1 and g has it at 0.
   (b) Only g is continuous everywhere.
   (c) As x tends to ∞ $f(x)$ also tends to ∞.
   (d) The range of g is bounded.

3. Let n be a positive integer such that

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{7} + \frac{1}{n}$$

   is an integer. Which of the following is/are true?

   (a) n has to be even.
   (b) 3 does not divide n
   (c) 7 divides n
   (d) 5 does not divide n

4. Let $a_1 < a_2 < a_3 < a_4 < \ldots < \ldots$ be a sequence of infinitely many positive integers which are in arithmetic progression. Furthermore, suppose a_2, a_4, a_8 are in geometric progression. Which of the following is/are true?

   (a) For every positive integer n, there is a triple of numbers $a_i < a_j < a_k$ which are in geometric progression with ratio n, i.e. $a_j = a_i n, a_k = a_j n$.
   (b) If $i < j$ then a_i divides a_j.
   (c) There are only finitely many n for which there are infinitely many triples $i < j < k$ with a_i, a_j, a_k in arithmetic progression with common difference n.
   (d) Integers a_{10}, a_{30}, a_{90} are in geometric progression.

5. Let A be a 20×11 matrix with real entries. After performing some row operations on A, we get a matrix B which has 12 nonzero rows. Which of the following is/ are true?

(a) Rank of A is 12.
(b) Rank of A and B are not related.
(c) If v is a vector such that $Av = 0$ then Bv is also 0.
(d) Rank of B is at most 11.

6. Let f be a single variable, real valued function such that all its derivatives exist. We say that f changes sign at some point x_0 in the domain if the product $f(x_0 + \epsilon) \cdot f(x_0 - \epsilon) < 0$ for small enough, positive ϵ.

   - A point x_0 in the domain is called a **turning point** of f, if the first derivative f' changes sign at x_0.
   - A point x_0 in the domain is called **flex point** of f, if the second derivative f'' changes sign at x_0.

   Which of the following statements is/are true?

   (a) If x_0 is a flex point of f then $f''(x_0) = 0$.
   (b) If $f'(x_0) = 0$ then x_0 is a turning point of f.
   (c) If $f''(x_0) = 0$ then x_0 is a flex point of f.
   (d) If x_0 is a flex point of f then $f'(x_0) \neq 0$.

7. Solitaire Tic-Tac-Toe is a new game on the market. Instead of adding X's and O's to an empty 3×3 grid, you start with a 3×3 grid in which every position already has an X or an O. In each move, you select a row, column or diagonal and reverse all the entries along the chosen line: all X's become O's and all O's become X's.

   For instance, here is a sequence of possible moves.

$$\begin{array}{c|c|c} X & O & X \\ \hline O & X & O \\ \hline X & O & X \end{array} \xrightarrow[\text{row}]{\text{Bottom}} \begin{array}{c|c|c} X & O & X \\ \hline O & X & O \\ \hline O & X & O \end{array} \xrightarrow[\text{diagonal}]{\text{SW-to-NE}} \begin{array}{c|c|c} X & O & O \\ \hline O & O & O \\ \hline X & X & O \end{array} \xrightarrow[\text{column}]{\text{Middle}} \begin{array}{c|c|c} X & X & O \\ \hline O & X & O \\ \hline X & O & O \end{array}$$

   Given an arrangement S, we want to explore all arrangements of the grid that we can generate starting with S. What is the smallest number m such that each such arrangement can be reached using at most m moves, starting with S?

   (a) 8
   (b) 16
   (c) 64
   (d) 256

8. A bug starts walking from the origin of a two dimensional plane. It walks 1 step up, 1/2 step to the left then 1/4 step down, then 1/8 step to the right, then 1/16 step up and so on, forever. In (x, y)-coordinates, its final position is given by

   (a) $(-\frac{2}{3}, \frac{4}{5})$
   (b) $(-\frac{2}{5}, \frac{4}{3})$
   (c) $(-\frac{2}{5}, \frac{4}{5})$
   (d) $(-\frac{2}{3}, \frac{4}{3})$

9. Let A, B, C be events such that $P(A) = P(B) = P(C) = 0.5$, $P(A \cap B) = 0.3$, $P(A \cap C) = 0$. Which of the following is/are true?

   (a) $P(A \cup B) = 0.75$
   (b) $P(A \cup C) = 1$

(c) $P(B \cap C) = 0.2$

(d) $P(B \cup C) = 0.9$

10. The mean and variance of a data set of size $n \geq 100$ are 1.1 and 4.3 respectively. It was then discovered that one data point was wrongly recorded as -0.9 and should be ignored. Then we can conclude that the mean μ and variance σ^2 of the $n-1$ data points satisfy

    (a) $\sigma^2 < 4.3$

    (b) $\sigma^2 > 4.3$

    (c) $\mu > 1.1$

    (d) $\mu < 1.1$

11. An $n \times n$ matrix $A = (a_{ij})$ is said to be antisymmetric if $a_{ij} = -a_{ji}$ for all $1 \leq i, j \leq n$. Which of the following is/are true?

    (a) If A is antisymmetric then all its principal diagonal entries are zero.

    (b) If A, B are antisymmetric matrices then the product AB need not be antisymmetric.

    (c) If A is antisymmetric then A^2 is symmetric.

    (d) A nonzero 3×3 antisymmetric matrix is invertible.

12. Consider the set $S = \{0, 1, 2, \ldots, n-1\}$ such that $n \geq 2$. Let x be an element sampled uniformly at random from S. Let y denote $(x + 2^n) \mod n$. Which of the following statements is/are true?

    (a) $\Pr\left[y = \lfloor\sqrt{n-1}\rfloor\right] = \frac{\lfloor\sqrt{n-1}\rfloor}{n}$

    (b) $\Pr[x = a, y = b] = \Pr[x = a] \cdot \Pr[y = b]$ for all $a, b \in S$

    (c) y is always an element of S

    (d) $\Pr\left[y = \lfloor\log_2(n-1)\rfloor\right] = 1/n$

13. 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 $[1, 6, 2, 7, 3, 8, 4, 9, 5, 10]$.

    We start with the deck $[8, 1, 4, 5, 3, 6, 2, 7]$ and keep shuffling. Which card(s) will never appear next to 5?

    (a) 1

    (b) 2

    (c) 7

    (d) 8

14. Consider the following pseudocodes of two functions, where `u % 2` denotes the remainder when `u` is divided by 2. The function `abs(v)` returns the absolute value of an integer `v`.

```
function foo1(u):
    if u % 2 == 0 AND abs(u) > 0:
        u = u + foo2(u-1)
    return u

function foo2(v):
    if v % 2 == 1 AND abs(v) > 1:
        v = v + foo1(v-1)
    return v
```

Which of the following is/are true?

(a) `foo1`, when called on a positive integer u, returns $u(u+1)/2$

(b) `foo2`, when called on an odd positive integer v, returns $v(v+1)/2$

(c) `foo2(-13)` goes into infinite recursion

(d) `foo1(14)` returns `105`

15. Which of the following statement(s) is/ are true?

(a) $(2023!)^2 < (2023)^{2023}$.

(b) $2^{\sqrt{7}} > 2^{\frac{21}{8}}$.

(c) $e^{\frac{1}{e}} < \sqrt{2}$.

(d) $\log_{10} 7 < \frac{1+\log_{10} 5}{2}$.

16. Let X, Y, Z be finite sets. Which of the following statement(s) is/are true?

(a) Let $f : X \to Y$, $g : Y \to Z$ be functions. such that the composite function $g \circ f : X \to Z$ is onto. Then both g, f have to be onto.

(b) Assume the number of elements of X, denoted as $|X|$, is strictly less than $|Y|$ and also assume that $|X| \leq |Z|$. There exists a one-one function $f : X \to Y$ and a $g : Y \to Z$ which is not one-one such that $g \circ f : X \to Z$ is one-one.

(c) Let id_X be the identity function on X, i.e. $id_X(x) = x$ for all $x \in X$. Let $f : X \to Y$ be a function. Then there exists a function $g : Y \to X$ such that $g \circ f = id_X$ exists.

(d) Denote by id_Y the identity function on Y, $id_Y(y) = y$ for all $y \in Y$. Then there exists $g : Y \to X$ such that $f \circ g = id_Y$ if and only if f is onto.

17. Consider the following real valued functions defined on appropriate domains:

$$f(x) = \frac{x}{x + \sin x}, \quad g(x) = x \ln(x).$$

Which of the following is/are true?

(a) Limit as $x \to 0$ of $f(x)$ is $\frac{1}{2}$.

(b) Limit as $x \to 0^+$ of $g(x)$ does not exist.

(c) Limit as $x \to \infty$ of $f(x)$ is finite.

(d) Limit as $x \to \infty$ of $g(x)$ is finite.

18. Consider the pseudocode below, where `n % 6` denotes the remainder when `n` is divided by `6`. The notation `n//2` stands for integer division, i.e., `15//2 = 7`, `36//2 = 18`, `49//2 = 24`.

```
function sixer(n):
    count = 0
    while n > 0:
        if n % 6 == 0:
            n = n - 1
        else:
            n = n//2
        count = count + 1
    return count
```

Which of the following is true when `sixer`(n) is invoked with sufficiently large n, say $n \geq 10^6$?

(a) The value of `count` is approximately $n/6$

(b) The value of `count` is approximately $n/2$

(c) The value of `count` is approximately $\log_2(n)$

(d) The value of `count` is approximately $(\log_2(n))^2$

19. In a football league, a collection of teams play against each other a number of times. A team gets 3 points for each match that it wins, 1 point for each match that it draws and 0 points for each match that it loses.

The following table gives the point standings after four teams have played a total of 7 matches. At this stage, each team has played every other team at least one.

Team	Matches	Points
Kolkata	4	8
Delhi	4	4
Chennai	3	4
Mumbai	3	1

How many matches have been drawn so far?

(a) 3

(b) 4

(c) 5

(d) 6

20. A new game show on TV has 100 boxes numbered 1, 2, ..., 100, each containing a mystery prize. The prizes are of different types, a, b, c, ..., in decreasing order of value, and are distributed randomly among the boxes. The most expensive item is of type a, a diamond ring, and there is exactly 1 of these. You are told that the number of items at least doubles as you move to the next category — there are at least twice as many items of type b as of type a, at least twice as many items of type c as of type b and so on.

You ask for the type of item in box 45. Instead of being given a direct answer, you are told that there are 31 items of the same type as box 45 in boxes 1 to 44 and 43 items of the same type as box 45 in boxes 46 to 100.

Which of the following statements can be true?

(a) There are exactly 3 types of prizes.

(b) There are exactly 4 types of prizes.

(c) There are exactly 5 types of prizes.

(d) There are exactly 6 types of prizes.

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. There are 5 friends A, B, C, D, E. Because of a heated argument not all of them are on speaking terms anymore. The array below describes who is talking to whom. A value of 1 indicates that the pair is on speaking terms, while a 0 indicates that they are not.

	A	B	C	D	E
A	-	1	1	1	1
B	1	-	1	0	1
C	1	1	-	1	0
D	1	0	1	-	1
E	1	1	0	1	1

A **gossip route** is an ordered sequence of all the five friends such that every person hears the gossip from exactly one friend (with whom they are on speaking terms) and shares it with exactly one other friend (with whom they are on speaking terms). The friend who starts the gossip doesn't hear it from anyone else and the fifth friend doesn't pass it on to anybody.

Answer the following questions:

(a) Suppose A starts spreading gossip along a gossip route and E hears it the last. Who is the third person to hear it?

(b) List all the gossip routes from B to E.

(c) There is another dispute between the friends. As a result, A is now not speaking with B and D. List all the gossip routes starting from A.

2. Two friends A and B are playing a coin flipping game with a fair coin, based on the following rules:

- When A flips the coin
  - If it is heads then A wins and the game ends.
  - If it is tails then B gets to flip the coin.
- When B flips the coin
  - If it is heads then B wins and the game ends.
  - If it is tails then A gets to flip the coin.
- The flipping continues till someone gets heads.

Suppose A is the first player to flip the coin. What is the probability that A will win the game.

3. Consider the following real valued function defined on the entire set of real numbers:

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

Show that $f'(x) = f(x)(1 - f(x))$.

4. Let f be the following function defined on integers:

$$f(n) = \frac{1}{3^n + \sqrt{3}}$$

Compute $\sqrt{3}\Big[f(-5) + f(-4) + f(-3) + \cdots + f(4) + f(5) + f(6)\Big]$

5. Find the determinant of the following 5×5 matrix:

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 2 & 5 & 2 & 0 & 0 \\ 0 & 2 & 5 & 2 & 0 \\ 0 & 0 & 2 & 5 & 2 \\ 0 & 0 & 0 & 2 & 5 \end{bmatrix}$$

6. Find the limit

$$\lim_{x\to\infty} \left(x - x\cos\frac{1}{\sqrt{x}} \right)$$

7. Let $N = 2^{150}$. How many bits are required to write N in binary (base 2)? How many digits are required to write N in decimal (base 10)? You may assume that $\log_{10} 2 = 0.30103$.

8. A domino is a sheet of paper with one number from $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ printed on each half. For example, a $4-4$ domino has 4 printed on both halves. A $0-4$ domino has 0 printed on one half and 4 on the other half. The order is not important, so a $0-4$ domino is the same as a $4-0$ domino.

   A complete set of dominos is a set of dominos containing exactly one copy of each different domino.

   (a) How many dominos are there in a complete set of dominos?

   (b) In how many ways can you select 2 dominos from a complete set so that at least one of the dominos has a 0 on it and at least one of the dominos has a 9 on it.

9. There are three classrooms, C_1, C_2, and C_3. In C_1 there are 5 boys and 5 girls. C_2 has 3 boys and 5 girls. C_3 has 5 boys and 3 girls. We select a classroom and then a student in the classroom. The probability of selecting classrooms C_1, C_2, and C_3 are $\frac{3}{10}$, $\frac{3}{10}$, and $\frac{4}{10}$, respectively. Once a room is selected, a student is picked uniformly at random from the class.

   (a) What is the probability that the selected classroom is C_3 and the chosen student is a girl?

   (b) What is the probability that the selected classroom is C_3 given that the chosen student is a girl?

10. How many integers in the set $\{1, 2, \ldots, 1000\}$ are relatively prime to 1000? Recall that two integers are relatively prime if the only common factor they have is 1.

11. Is the following inequality true, given that n is a positive integer?

$$\int_1^{n+1} \ln x \, dx \geq \ln n!$$

   Justify your answer.

12. Football teams T_1 and T_2 play two games against each other in the Premier League. It is assumed that the outcomes of the two games are independent of each other. The probabilities of T_1 winning, drawing and losing against T_2 are $\frac{1}{2}$, $\frac{1}{6}$ and $\frac{1}{3}$ respectively. Each team gets 3 points for a win, 1 point for a draw and 0 points for a loss in a game. Let X and Y denote the total points scored in these two games by team T_1 and T_2, respectively.

    (a) Find $P(X = Y)$.

    (b) Find $E(X)$.

13. Let `A`, `B` and `C` denote arrays of real numbers, where `B` has `n-1` entries and `A`, `C` have `n` entries each. Consider the following algorithm involving the three given arrays.

```
for k from 2 to n:
    t = B[k-1]
    B[k-1] = t/A[k-1]
    A[k] = A[k] - t * B[k-1]
end for
for k from 2 to n:
    C[k] = C[k] - B[k-1] * B[k-1]
end for
C[n] = C[n]/ A[n]
for k from n-1 to 1 with steps of -1:
    C[k] = (b[k]/A[k]) - B[k] * b[k+1]
end for
```

An arithmetic operation involves addition, subtraction, multiplication or division of real numbers. How many arithmetic operations, in total, are performed in the algorithm above?

Note: Integer subtraction in array indices, like `B[k-1]`, is not to be counted as an arithmetic operation.

14. In the following pseudocode segment, `A` denotes an array and `len(A)` denotes the number of elements in that array. The array is indexed from `1` to `len(A)`. Write down the array `A` in each iteration of the for loop when the input array is `<20, 3, 9, 12, 70, 6>`.

```
function weirdify(A):
    for i from 1 to len(A):
        if i % 2 == 0:
            A[i] = A[i/2]*A[i]
        else if i*i > len(A):
            A[i] = A[i-2] + 4
        else:
            A[i] = A[i] - 1
```

15. Consider a city with n East-West Streets (EWS) and n North-South Avenues (NSA). The EWS are the line segments $\{y = j \mid 1 \le x \le n\}$ for all $j \in \{1, 2, \ldots, n\}$ and the NSA are the line segments $\{x = i \mid 1 \le y \le n\}$ for all $i \in \{1, 2, \ldots, n\}$. A junction is a pair (i, j) where avenue i intersects street j. How many junctions are there in the city?

    For increased safety, the city council decides to place cameras at various junctions in the city. The cameras being super-powerful, can observe the entire street and avenue corresponding to the junction that they are placed at. For instance, a camera placed at the junction (i, j) can observe both avenue i and street j. Write down an expression for the minimum number of cameras needed to observe every street and avenue in the city.

    Justify your answers.

16. Find the largest value of xy subject to $x^2 + y^2 = 2$.

17. Each lawyer on a certain remote island is either honest or dishonest (but not both).

    - Honest lawyers always speak the truth.
    - Dishonest lawyers always lie.

    Is it possible for a lawyer on this island to claim that he/ she is dishonest? Explain. A judge asks lawyer A, "are you honest"? Before A could answer, lawyer B says "A will say yes. But then, he'll be lying". Which lawyer is honest and which one is dishonest? Explain.

18. A new restaurant chain *Burger Paradise* has recently opened in Chennai. Burger Paradise offers a variety of burgers, including vegetarian options, as well as sides and drinks. According to the restaurant's data,

20% of customers order vegetarian burgers. The restaurant's data also indicates that customers who order a burger are twice as likely to order a side of fries than a side of salad.

(a) If the restaurant serves 200 customers daily, what is the expected number of vegetarian burgers they will sell over a week?

(b) If 30% of the customers order a burger and 30% of those who order a burger order a side of salad, what is the probability that a customer orders a burger along with a side of fries?

19. Burger Paradise plans to expand to other cities and is considering opening a new branch in Coimbatore. They have identified three potential locations, A, B and C, and are evaluating the profitability of each location. Based on their research, they estimate that the fixed cost for opening a new location will be ₹$50,00,000$, ₹$70,00,000$ and ₹$65,00,000$ at A, B and C, respectively. They also estimate the variable cost per customer to be ₹85 and the average revenue per customer to be ₹150 at all three locations. They expect to serve $100,000$, $150,000$ and $130,000$ customers in the first year at locations A, B and C, respectively

    Which location in Coimbatore is expected to be the most profitable in the first year of operation, and what is the expected profit at this location?

20. Towns A, B and C are connected to each other by several roads, with at least one road connecting each pair. In going from one town to another, one can take the road that connects them directly or could travel via the third town. It is given that, in total, there are 33 roads from A to B, including those via C. Similarly, there are 23 roads from B to C, including those via A. How many routes are there from A to C, including those via B?