

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

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**.

Solution: (a), (c)

- (a) Finding $P(\text{valid claim})$

A claim is valid if **at least one** event occurs. Using the complement rule:

$$P(\text{no claim}) = P(\text{no medical emergency}) \times P(\text{no vehicle accident}) = \left(1 - \frac{1}{10}\right) \left(1 - \frac{1}{20}\right) = \frac{9}{10} \times \frac{19}{20} = \frac{171}{200}$$

$$\therefore P(\text{valid claim}) = 1 - \frac{171}{200} = \frac{29}{200}$$

Simply adding $\frac{1}{10} + \frac{1}{20} = \frac{30}{200}$ overcounts the case where both events occur simultaneously. The correct approach via inclusion-exclusion gives the same result:

$$P(A \cup B) = \frac{1}{10} + \frac{1}{20} - \frac{1}{10} \times \frac{1}{20} = \frac{20}{200} + \frac{10}{200} - \frac{1}{200} = \frac{29}{200} \quad \square$$

- (b) The correct expression for $P(\text{exactly one claim in 4 years})$ is:

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

instead the option is given as

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

- (c) $P(\text{exactly one of the four members files a valid claim})$ is correct.

- (d) Per policy the revenue is:

$$E = (+4800) \times \frac{171}{200} + (-1200) \times \frac{29}{200} = \text{Rs. } 3,930$$

So Expected revenue for 800 policies is $800 \times 3930 = 31,44,000$. So (d) is wrong

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.

Solution: (b), (c)

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

Solution: (a)

Only switches numbered by perfect squares will be ON: 1, 4, 9. Note that a switch will be ON after an odd number of steps. So we are looking for numbers that have an odd number of divisors including 1 and itself, and are less than 10.

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) $\tan x \geq x, \quad \text{for all } x \text{ where defined}$

Solution: (a), (c)

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.

Solution: (a), (b), (d)

(a) is true: choose 2-element images and count the onto functions $- 2^4 - 2$. (b) is true by inclusion-exclusion. (c) is false since domain size $>$ codomain size. (d) is true: Since $f(1), f(2)$ are already determined, valid pairs for $(f(3), f(4))$ are $(3, 3), (3, 1), (3, 2), (1, 3), (2, 3)$.

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) $P(A | B) = \frac{3}{5}$

Solution: (a), (b)

$P(A) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}$, so (a) is true. $P(B) = \frac{3}{5}$, so (b) is true.

For (c), note that $P(A^c) = 1 - \frac{3}{10} = \frac{7}{10}$. The event $B^c \cap A^c$ is the event that the first ball is blue (since if the first ball is blue, both balls certainly cannot be red), so $P(B^c \cap A^c) = P(B^c) = \frac{2}{5}$. Hence

$$P(B^c | A^c) = \frac{P(B^c \cap A^c)}{P(A^c)} = \frac{2/5}{7/10} = \frac{4}{7} \neq \frac{2}{3},$$

so (c) is false.

For (d), $P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{3/10}{3/5} = \frac{1}{2} \neq \frac{3}{5}$, so (d) is false.

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.

Solution: (a), (b), (c)

The total number of arrangements is $5! = 120$. Treating A and B as a block gives $4! \cdot 2 = 48$ arrangements together. So the arrangements where they are not together is $120 - 48 = 72$; thus (a) is true. The probability of this happening is $\frac{72}{120} = \frac{3}{5}$, so (b) is also true. There are 48 arrangements where A, D can sit besides each other. Now, out of these arrangements that contain DA, B or B, AD are not allowed, these are 12 in number hence (c) is also correct. Option (d) is false, because we have to subtract arrangements where A, B are sitting next to each other.

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.

Solution: (a), (b), (d)

For option (c) there are straightforward counterexamples.

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

Solution: (c)

There are exactly 7 pairs of the form $(i, i + 1)$. We split the counting in 2 cases. First we count the subsets that contain either $\{1, 2\}$ or $\{7, 8\}$. In order to extend these subsets to a 3-element subset there are exactly 5 choices each (since for the first pair 3 is not allowed for the second 6 is not allowed). Then we count subsets containing one of the remaining 5 pairs. In this case, the third element can neither be $i - 1$ nor $i + 2$. So we are left with 4 choices. The total count is $(2 \times 5) + (5 \times 4) = 30$.

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) $P(X \leq 3) = \frac{8}{36}$

Solution: (a), (b), (c)

$P(X = 6) = \frac{11}{36}$ and $P(X = 1) = \frac{1}{36}$ are standard counts, so (a), (b) are true. One can compute $E[X] \approx 4.47 > 4$, so (c) is true. For $X \leq 3$, both rolls must be ≤ 3 , giving $9/36$, so (d) is false.

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

Solution: (a), (b), (c), (d)

Every positive integer can be obtained by reversing the process: if even divide by 2, if odd subtract 1, eventually reaching 1. Hence all listed numbers are reachable.

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

Solution: (b)

Use inclusion–exclusion:

Total number of arrangements is $5! = 120$.

Let A_i be the set of arrangements where i is well-placed. Then

$$|A_i| = 4!, \quad |A_i \cap A_j| = 3!, \quad |A_i \cap A_j \cap A_k| = 2!, \quad \text{etc.}$$

Hence the number is

$$\begin{aligned} 5! - \binom{5}{1}4! + \binom{5}{2}3! - \binom{5}{3}2! + \binom{5}{4}1! - \binom{5}{5}0! \\ = 120 - 5 \cdot 24 + 10 \cdot 6 - 10 \cdot 2 + 5 \cdot 1 - 1 \\ = 120 - 120 + 60 - 20 + 5 - 1 = 44. \end{aligned}$$

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.

Solution: (d)

If S1 is true, consistency fails. If S1 is false, then S2 and S3 must both be true, which is consistent. Thus 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) $\frac{49}{64}$

Solution: (d)

Total subsets = $2^6 = 64$. Subsets with only even elements: $2^3 = 8$. Only odd: $2^3 = 8$. Empty counted twice, subtract once: $8 + 8 - 1 = 15$. Good subsets = $64 - 15 = 49$. Probability = $49/64$.

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.

Solution: (a), (d)

The code correctly computes the average. The line that updates `avg` can safely be moved out of the loop.

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

Solution: (b)

The code computes $0 - 1 + 2 - 3 + 4 - 5 + 6 - 7 + 8 - 9 + 10 = 5$.

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

Solution: (a)

The function sorts the array in ascending order and returns the first (smallest) element.

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.

Solution: (a), (b), (c)

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.

Solution: (a), (c)

Options (a) and (c) are true, since this function is similar to the bubble sort and the loops are fixed. Option (b) is false because the number of comparisons is fixed. Similarly, (d) is also false, all the comparisons will be performed.

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

Solution: (b)

Trace: $6 \rightarrow 3 \rightarrow 10 \rightarrow 5 \rightarrow 16 \rightarrow 8 \rightarrow 4 \rightarrow 2 \rightarrow 1$, which is 8 steps.

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 $\text{trace}(A)$.

Solution: The matrix satisfies the characteristic polynomial.

$$A^2 - tA + dI = 0,$$

where t is the trace and d is the determinant. Substitute $A^2 = I$ and $d = \pm 1$ in the above equation to get the three possible values of t .

Alternatively, the given matrix equation implies that each eigen value satisfies $\lambda^2 = 1$. So the eigenvalues come from the set $\{+1, -1\}$. Hence the trace could be $-2, 0$ or $+2$.

2. How many binary strings of length 8 contain exactly three 1s such that no two 1s are adjacent.

Solution: In particular, the string should contain 5 zeros, which create 6 gaps (before, between and after). Choose 3 positions from 6 gaps: $\binom{6}{3} = 20$. We have to choose 3 gaps to insert a single 1 each.

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).

Solution: First, assume $q = 1$. Then $f(p) = f(1 + \dots + 1) = f(1) \times \dots \times f(1) = (f(1))^p$. Let q be any integer. Then $f(qr) = f(p) = (f(1))^p$. Now, qr is also r added q times, hence $f(qr) = (f(r))^q$. But, $(f(r))^q = (f(1))^p$ implies the answer.

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.

Solution: Let $f(1) = a$. Assumption implies that

$$f(x) = a^x \quad x \in \mathbb{R}.$$

We want the function to be strictly increasing, hence $x < y$ should imply $a^x < a^y$, this is possible only when $a > 1$.

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.

Solution: Let S_1 be the first sum and S_2 be the second sum. We want S_2/S_1 to be an integer:

$$\frac{S_2}{S_1} = \frac{n(n+1)(2n+1)/6}{n(n+1)/2} = \frac{2n+1}{3}.$$

Now, $2n+1$ is a multiple of 3 if and only if n is of the form $3k+1$. First few integers are 1, 4, 7, 10 etc.

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.

Solution: Exactly one of the first i entries must be maximum, so the probability is $\frac{1}{i}$.

7. What is the expected number of new maxima in a list?

Let X_i be a random variable which takes the value 1 when the i th position is a new maximum, and the value 0 otherwise. Then the random variable $X = \sum_{i=1}^n X_i$ denotes the total number of new maxima in a list. The expected number of new maxima is thus calculated as follows:

$$\begin{aligned} E[X] &= \sum E[X_i] \\ &= \sum P(X_i = 1) \\ &= \sum \frac{1}{i}. \end{aligned}$$

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)`.

Solution:

$$13 \rightarrow 12 \rightarrow 6 \rightarrow 3 \rightarrow 2 \rightarrow 1$$

This takes 5 steps. Hence `Steps(13) = 5`.

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?

Solution: Each step either divides by 2 (removes a binary digit) or subtracts 1 (reduces the number of 1's in the binary representation of n). The number of steps is

$$(\text{number of bits} - 1) + (\text{number of 1's} - 1).$$

This is maximized when all k bits are 1, i.e., $n = 2^k - 1$. In that case,

$$\text{Steps}(n) = (k - 1) + (k - 1) = 2k - 2.$$

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.

Solution: Here is a trace through the sequence step by step:

Step	1	2	3	4	5	6	7	8	9	10
a_i	3	1	4	1	5	9	2	6	5	9
M	3	3	4	4	5	9	9	9	9	9
c	1	1	1	1	1	1	1	1	1	2

After all elements have been processed, $M = 9$ and $c = 2$.

11. What do the variables M and c represent at the end of processing? Justify your answer in one or two sentences.

Solution: At the end of processing, M is the maximum value in the sequence, and c is the number of times this maximum value appears in the sequence. This is because M is updated to a_i whenever a new maximum is encountered (resetting c to 1), and c is incremented whenever the current maximum is seen again; all elements smaller than the current M leave both variables unchanged.

12. Find the sum of all integers from 1 to 100 which are neither multiples of 3 nor multiples of 7.

Solution: The sum of all integers from 1 to 100 is $50 * 101 = 5050$.

We subtract the sum $3 + 6 + \dots + 99 = 3(1 + 2 + \dots + 33) = 3 \times 17 \times 33 = 1683$.

We also subtract $7 + 14 + \dots + 98 = 7(1 + 2 + \dots + 14) = 7 \times 7 \times 15$.

We add back $21 + 42 + \dots + 84 = 21(1 + 2 + \dots + 4) = 21 \times 10$.

The answer is $5050 - (3 \times 17 \times 33) - (7 \times 7 \times 15) + 21 \times 10 = 2842$

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 .

Solution: We have

$$\begin{aligned} n^2 + n + 1 &\equiv 0 \pmod{p} \\ (n-1)(n^2 + n + 1) &\equiv 0(n-1) \pmod{p} \\ n^3 - 1 &\equiv 0 \pmod{p}. \end{aligned}$$

Thus, we have $n^3 \equiv 1 \pmod{p}$; the only possible remainder is 1.

14. Suppose $p > 3$. Show that the numbers $1, n, n^2$ leave distinct remainders upon division by p .

Solution: We know that n^3 leaves residue 1 after dividing by p . Hence, we first need to show that neither n nor n^2 leave remainder 1. This is straightforward since $p > 3$. For example, If n leaves remainder 1 then $n^2 + n + 1$ leaves remainder 3, which is divisible by p . If n^2 leaves remainder 1 then p divides $n^2 - 1 = (n-1)(n+1)$ which further implies p divides $n-1$ or it divides $n+1$. The first case is already covered and the second contradicts the given information. Now n and n^2 can't leave the same remainder, otherwise p will divide $n^2 - n$, equivalently p divides n or it divides $n^2 - 1$, neither of which is possible.

15. Find all two-digit positive integers that are equal to three times the sum of their digits.

Solution: Let the number be $10a + b$. Then $10a + b = 3(a + b)$. Which leads to $b = \frac{7a}{2}$. Hence a has to be even. Trying all the even digits we get that 27 is the only number satisfying the given condition.

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.

Solution: There are several such possibilities. For example, $(2, 2, 2, 1, 1)$ is realized by $A \leftrightarrow B, B \leftrightarrow C, C \leftrightarrow A, D \leftrightarrow E$. Or, $(3, 3, 2, 2, 2)$ is realized by $A \leftrightarrow B, B \leftrightarrow C, C \leftrightarrow A, A \leftrightarrow E, B \leftrightarrow D$. Any valid sequence, realized by a configuration is correct.

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.

Solution. Try out the possibility that each model is the faulty one, and apply the stated rules.

Case 1: M_1 is faulty. Then M_2 and M_3 are non-faulty. Since M_2 is non-faulty, its statement (M_3 is faulty) must be true — but M_3 is non-faulty. Contradiction.

Case 2: M_2 is faulty. Then M_1 and M_3 are non-faulty. Since M_1 is non-faulty, its statement (M_2 is faulty) must be true. Since M_2 is faulty, its statement (M_3 is faulty) must be false, i.e. M_3 is not faulty. Since M_3 is non-faulty, its statement (M_1 is not faulty) must be true. All three statements are consistent.

Case 3: M_3 is faulty. Then M_1 and M_2 are non-faulty. Since M_1 is non-faulty, its statement (M_2 is faulty) must be true — but M_2 is non-faulty. Contradiction.

Therefore M_2 is the faulty model.

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 $Aw = 0$.

Solution: Suppose $x = Ay$ for some y . Then $Ax = A(Ay) = A^2y = Ay = x$. Given v , define $u := Av$ and $w := v - Av$. Then straightforward application shows $Au = u$ and $Aw = 0$. Uniqueness is also straightforward, just assume that there 2 different expressions $v = u_1 + w_1 = u_2 + w_2$ for the identity satisfied by A and then compare.

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 $\text{Var}(T)$.

Solution:

Part (a):

Since $X_i \sim \text{Uniform}(0, \theta)$, the CDF of each X_i is $F(x) = \frac{x}{\theta}$ for $0 \leq x \leq \theta$. The CDF of $T = \max(X_1, X_2, X_3)$ is:

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(\max(X_1, X_2, X_3) \leq t) \\ &= P(X_1 \leq t, X_2 \leq t, X_3 \leq t) \\ &= P(X_1 \leq t) P(X_2 \leq t) P(X_3 \leq t) \quad (\text{by independence}) \\ &= \left(\frac{t}{\theta}\right)^3, \quad 0 \leq t \leq \theta \end{aligned}$$

Differentiating, the CDF of T is:

$$f_T(t) = \frac{3t^2}{\theta^3}, \quad 0 \leq x \leq \theta,$$

we have the pdf of T .

Part (b):

The expected value is:

$$E(T) = \int_0^\theta t \cdot \frac{3t^2}{\theta^3} dt = \frac{3}{\theta^3} \int_0^\theta t^3 dt = \frac{3}{\theta^3} \cdot \frac{\theta^4}{4} = \boxed{\frac{3\theta}{4}}$$

We compute $E(T^2)$:

$$E(T^2) = \int_0^\theta t^2 \cdot \frac{3t^2}{\theta^3} dt = \frac{3}{\theta^3} \int_0^\theta t^4 dt = \frac{3}{\theta^3} \cdot \frac{\theta^5}{5} = \frac{3\theta^2}{5}$$

Therefore:

$$\text{Var}(T) = E(T^2) - [E(T)]^2 = \frac{3\theta^2}{5} - \left(\frac{3\theta}{4}\right)^2 = \frac{3\theta^2}{5} - \frac{9\theta^2}{16} = \theta^2 \left(\frac{48 - 45}{80}\right) = \boxed{\frac{3\theta^2}{80}}$$

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.

Solution:

Step 1: Verify the claim using Chebyshev's inequality.

For $X \sim \text{Exponential}(\lambda)$, we have $\mu = \frac{1}{\lambda}$ and $\sigma = \frac{1}{\lambda}$. The event $\left|X - \frac{1}{\lambda}\right| < \frac{2}{\lambda}$ is of the form $|X - \mu| < k\sigma$ with $k = 2$. By Chebyshev's inequality:

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

So the student's claim is **consistent with** Chebyshev's inequality. But this only guarantees the bound; it does not confirm whether $\frac{3}{4}$ is close to or far from the true probability.

Step 2: Compute the exact probability.

The event $\left|X - \frac{1}{\lambda}\right| < \frac{2}{\lambda}$ is equivalent to:

$$-\frac{1}{\lambda} < X < \frac{3}{\lambda}$$

Since $X \geq 0$, this reduces to $0 \leq X < \frac{3}{\lambda}$. Therefore:

$$P\left(\left|X - \frac{1}{\lambda}\right| < \frac{2}{\lambda}\right) = \int_0^{3/\lambda} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^{3/\lambda} = 1 - e^{-3} \approx 0.9502$$

Step 3: Compare and comment.

$$\underbrace{\frac{3}{4} = 0.75}_{\text{Chebyshev lower bound}} \leq \underbrace{1 - e^{-3} \approx 0.9502}_{\text{Exact probability}} \checkmark$$

The claim is **true**; the exact probability $1 - e^{-3} \approx 0.95$ comfortably exceeds the Chebyshev lower bound of $\frac{3}{4}$. The gap between 0.75 and 0.95 reveals two things: (i) Chebyshev's inequality is valid but conservative, since it must hold for *any* distribution with finite mean and variance, and (ii) when the specific distribution is known, the exact probability can be substantially tighter than the bound suggests.