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

1<sup>st</sup> August 2021

Enter your Admit Card Number: D	] – 🗌		] – [				
---------------------------------	-------	--	-------	--	--	--	--

## 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.:  $\binom{n}{r}$ ,  ${}^{n}P_{r}$ , n!.

### 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

For questions in part (A), you have to provide the answers on the computer. Choose only the appropriate answer(s) from the choices provided. If the answer is parts (a) and (c), choose only (a) and (c). There is no partial credit.

1. There are two longest subsequences, not necessarily contiguous, common to the strings "ARTIFICIAL" and "INTELLIGENCE". They are "IIC" and "TIC" which are of length three.

Consider two strings S1 = "CORONAVIRUS" and S2 = "SARSCOVID". Let x be the length of a longest common subsequence between S1 and S2 and let y be the number of such longest common subsequences of length x between S1 and S2. What is x + 5y?

- (a) 13
- (b) 15
- (c) 14
- (d) 16

2. The roots of the polynomial  $p(x) = x^4 - 2x^3 - 2x^2 + 8x - 8$  are:

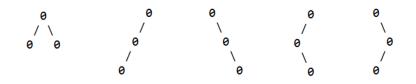
- (a) 1, -1, 2, 2+3i(b) 1+i, 1-i, 2, -2 (c) 1, -1+i, 2, 2+3i(d) 1+i, -1+i, 2, -2
- 3. Consider the following code, in which A is an array indexed from 0.

```
function foo(A,n) {
    m = A[0];
    x = 0;
    for i = 0 to n-1 {
        x = x + A[i];
        if (m < x) {
            m = x;
        }
        if (x < 0) {
            x = 0;
        }
        return(m);
}
If A = [-12,-3,5,10,8,-16,-23,12,-5,7], what will foo(A,10) return?</pre>
```

```
(a) 23
```

- (b) 17
- (c) -17
- (d) 35

- 4. A student has an average score of 80 from her first four Mathematics tests, and 88 from her first five Physics tests. How much must she score in her upcoming tests to raise her average score in Mathematics and Physics to 82 and 89, respectively?
  - (a) 92 in Mathematics and 100 in Physics.
  - (b) 86 in Mathematics and 90 in Physics.
  - (c) 88 in Mathematics and 92 in Physics.
  - (d) 90 in Mathematics and 94 in Physics.
- 5. A binary tree starts with a single root node at the top of the tree. Each node can have either a left child or a right child, or both, or neither. The children of a node are drawn below it, connected by edges. Here are the five possible binary trees with three nodes.



Note that the directions left and right of the children matter. In the second tree, the root has a left child that has a left child, while, in the fourth tree, the root has a left child that has a right child, and so on.

How many different binary trees can be constructed with four nodes?

- (a) 13
- (b) 15
- (c) 14
- (d) 30

6. Consider the following code, in which A is an array indexed from 0.

```
function foo(A,n) {
    x = 0;
    for i = 0 to n-1 {
        x = A[i]^x;
    }
    return(x);
}
```

Here, a<sup>b</sup> represents the bitwise *Exclusive OR* function over variables a and b.

Given integers a and b, we write them in binary, padded by zeros to the left to make them of equal length. We then apply Exclusive OR to these binary representations bitwise. The operation  $a^b$  denotes the integer value obtained by performing bitwise Exclusive OR on the binary values of a and b. For example,  $3^4 = 011^100 = 111 = 7$ , and  $9^5 = 1001^0101 = 1100 = 12$ . The truth table for the Exclusive OR function is provided below.

a	b	a^b
0	0	0
0	1	1
1	0	1
1	1	0

If A = [2,3,7,2,3], what will foo(A,5) return?

- (a) 0
- (b) 2
- (c) 3
- (d) 7
- 7. Fifteen telephones are received at a service center. Of these, 5 are mobile, 6 are cordless, and 4 are wired. These 15 phones are randomly numbered from 1 to 15 to establish the order in which they are serviced. Which of the following statement(s) is/are correct?
  - (a) The probability that among the first 3 serviced, the first and third are mobile and the second is not, is  $\frac{5 \times 10 \times 4}{15 \times 14 \times 13}$ .
  - (b) The probability that the first four serviced are all the wired phones, is  $\frac{1}{(^{15})}$ .
  - (c) The probability that after servicing ten of these phones, only one of the three types remain to be serviced, is  $\frac{\binom{6}{5}}{\binom{15}{5}}$ .
  - (d) The probability that two phones of each type are among the first six serviced, is  $\frac{\binom{5}{2} + \binom{6}{2} + \binom{4}{2}}{\binom{15}{6}}$ .

- 8. Which of the following statement(s) is/are true?
  - (a) For k = 69597 and  $x_1, x_2, \dots, x_k \in (0, \infty)$ ,  $\frac{x_1 + x_2 + \dots + x_{2^k}}{2^k} \le (x_1 x_2 \cdots x_{2^k})^{\frac{1}{2^k}}$ .
  - (b) For any three real numbers  $x, y, z, |x z| \le |x y| + |y z|$ .
  - (c) For |r| < 1,  $\sum_{n=0}^{\infty} 37 \cdot r^n = \frac{37}{1-r}$ .
  - (d)  $(n^3 n)$  is divisible by 3 for all positive integers n.
- 9. Which of the following statement(s) is/are true for an arbitrary  $n \times n$  matrix A?
  - (a) Exchanging two rows of A does not change its determinant.
  - (b) Exchanging two rows of A does not change its trace.
  - (c) Replacing each diagonal element of A with a 1 does not change its determinant.
  - (d) Exchanging two columns of A negates its determinant.
- 10. Let A, B be  $n \times n$  matrices. Which of the following properties of A and B is/are preserved under matrix multiplication?
  - (a) Being upper triangular.
  - (b) All diagonal elements being zero.
  - (c) Being diagonal.
  - (d) Being symmetric.

11. The proportion of visitors to an e-commerce website in a week who would purchase a product follows the beta distribution with the probability density function

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha - 1} (1 - x)^{\beta - 1} \quad \text{for } 0 < x < 1,$$

where  $\Gamma$  is the gamma function,  $\alpha > 0$ ,  $\beta > 0$  and  $(\alpha + \beta) > 2$ . The mean and mode of the beta distribution are  $\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta}$  and  $\mathbb{M}ode(X) = \frac{\alpha - 1}{\alpha + \beta - 2}$ . From historical data the mean and mode of the proportion of buyers are estimated as  $\frac{1}{4}$  and  $\frac{1}{5}$  respectively. Which of the following statement(s) is/are correct?

- (a)  $\alpha = 3$  and  $\beta = 9$ .
- (b) The probability that in a given week the proportion of visitors who buy the product is between 10% and 20% is given by

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left[ \int_0^{0.2} x^{\alpha-1} (1-x)^{\beta-1} \, dx - \int_0^{0.1} x^{\alpha-1} (1-x)^{\beta-1} \, dx \right].$$

(c) The mean proportion of buyers can be calculated as follows

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx.$$

(d) The probability that in a given week the proportion of visitors who buy the product is less than 20% is given by

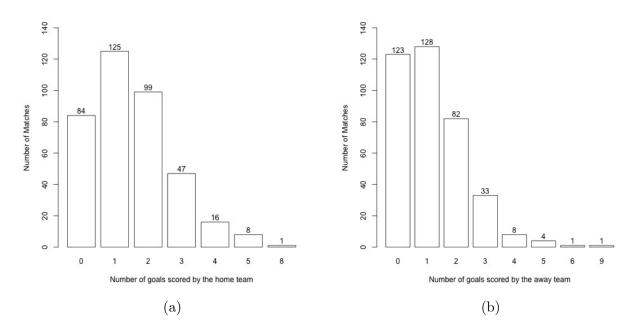
$$1 - \int_{0.2}^{1} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha} (1 - x)^{\beta} dx$$

- 12. There are  $10^{10}$  functions from  $\{1, 2, ..., 10\}$  to  $\{1, 2, ..., 10\}$ . Pick one such function f at random. Let  $p_1$  be the probability that  $f^{-1}(3)$  has exactly four elements, and  $p_2$  the probability that  $f^{-1}(2)$  has exactly six elements. Which of the following statement(s) is/are true?
  - (a)  $p_1 > p_2$
  - (b)  $p_1 < p_2$
  - (c)  $p_1 = p_2$
  - (d)  $p_1 = 3^4 p_2$

13. For a non-zero real number *a*, the inverse of  $J = \begin{pmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{pmatrix}$  is

(a) 
$$\begin{pmatrix} a^{-1} & a^{-2} & a^{-3} \\ 0 & a^{-1} & a^{-2} \\ 0 & 0 & a^{-1} \end{pmatrix}$$
  
(b) 
$$\begin{pmatrix} a^{-1} & -a^{-2} & a^{-3} \\ 0 & a^{-1} & -a^{-2} \\ 0 & 0 & a^{-1} \end{pmatrix}$$
  
(c) 
$$\begin{pmatrix} a^{-1} & 1 & 0 \\ 0 & a^{-1} & 1 \\ 0 & 0 & a^{-1} \end{pmatrix}$$
  
(d) 
$$\begin{pmatrix} a^{-1} & a^{-2} & 0 \\ 0 & a^{-1} & a^{-2} \\ 0 & 0 & a^{-1} \end{pmatrix}$$

**Description for following two questions**: In the 2019-2020 season of the English Premier League (EPL), 380 matches were played in a home and away format. The figure below describes the number of goals scored by the home team and the away team against the number of matches played. For example, the home team scored one goal in 125 matches.



14. Which of the following statement(s) is/are correct?

- (a) The average number of goals scored by the home team over 380 matches is less than 1.5, and for away teams are less than 1.
- (b) In the whole season of the EPL, a total of 1034 goals were scored.
- (c) The probability that away team failed to score any goal is 22.11%.
- (d) The probability that home team scored more than 3 goals is 6.6%.
- 15. Which of the following statement(s) is/are correct?
  - (a) In more than 50% matches, the home team scored at most one goal.
  - (b) In more than 10% matches, the away team scored more than two goals.
  - (c) In more than 15% matches, the home team scored three or more goals.
  - (d) In more than 90% matches, the away team scored less than three goals.
- 16. You are given a sequence of N digits. From this you can generate N-digit numbers as follows: pick up each digit in the given order, starting from the first one. Place it either to the left or to the right of the number you have generated so far. For instance, if you start with the sequence [4, 3, 2, 5, 2], you can generate 53422 through the steps  $4 \rightarrow 34 \rightarrow 342 \rightarrow 5342 \rightarrow 53422$ , and 22345 through the steps  $4 \rightarrow 34 \rightarrow 234 \rightarrow 2345 \rightarrow 22345$ .

Which of the following sequences could have generated 51324?

- (a) [2, 1, 3, 4, 5]
- (b) [2, 3, 1, 5, 4]
- (c) [2, 3, 4, 1, 5]
- (d) [2, 4, 1, 3, 5]

17. Which of the following is/are true?

- (a)  $2^{\frac{1}{2}} < 3^{\frac{1}{3}}$  and  $3^{\frac{1}{3}} > 4^{\frac{1}{4}}$ .
- (b)  $2^{r-1}r! \ge (2r-2)!$  for all natural numbers r.
- (c)  $\sin(2x) \ge \cos(x)$  for all values of x in  $\left[\frac{\pi}{6}, \frac{\pi}{2}\right]$ .
- (d)  $|x-1| 1 \le ||x-1| 1|$  for all real numbers x.

18. Let  $A_{11}, A_{12}, A_{21}, A_{22}, B_{11}, B_{12}, B_{21}, B_{22}$  be  $n \times n$  matrices. Consider the  $2n \times 2n$  matrices  $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ 

- and  $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ . The product AB is given by: (a)  $\begin{bmatrix} A_{11}B_{11} & A_{12}B_{12} \\ A_{21}B_{21} & A_{22}B_{22} \end{bmatrix}$ (b)  $\begin{bmatrix} A_{11}B_{21} + A_{12}B_{11} & A_{11}B_{22} + A_{12}B_{21} \\ A_{21}B_{21} + A_{22}B_{11} & A_{21}B_{22} + A_{22}B_{21} \end{bmatrix}$ (c)  $\begin{bmatrix} A_{11}(B_{11} + B_{21}) & A_{12}(B_{12} + B_{22}) \\ A_{21}(B_{11} + B_{21}) & A_{22}(B_{12} + B_{22}) \end{bmatrix}$ (d)  $\begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}$
- 19. Consider right angled triangles with integer side lengths a, b, c where c is the hypotenuse. Call a right angled triangle *special* if a and b are odd integers. How many special right angled triangles exist?
  - (a) 23
  - (b) infinitely many
  - (c) 0
  - (d) 12
- 20. Given a sequence of numbers  $[x_1, x_2, \ldots, x_n]$ , we say that  $x_i$  is left-visible if each number in  $[x_1, x_2, \ldots, x_{i-1}]$  is strictly smaller than  $x_i$ . Similarly  $x_i$  is right-visible if each number in  $[x_{i+1}, x_{i+2}, \ldots, x_n]$  is strictly smaller than  $x_i$ . For instance, in the sequence [13, 37, 24, 55, 46], the numbers 13, 37, 55 are left-visible and 55, 46 are right-visible.

Suppose we construct a sequence from the set  $\{17, 22, 38, 49, 53\}$  such that 4 numbers are left-visible and 2 numbers are right-visible. In addition, we are told that 22 is at the second position from the left. What are the possible numbers at the third and fifth positions?

- (a)  $\{38, 53\}$
- (b)  $\{38, 49\}$
- (c)  $\{17, 49\}$
- (d)  $\{17, 38\}$

#### 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. Consider the following code.

```
function foo(n) {
    answer = 0:
    x = n;
    while (x > 0) {
        y = x \% 10;
        answer = (answer * 10) + y;
        x = x // 10;
    }
    x = answer;
    answer = (answer * 10) + 1;
    while(x > 0) {
        y = x \% 10;
        answer = (answer * 10) + y;
        x = x // 10;
    }
    return(answer);
```

} Here,

• a % b represents the remainder when a is divided by b. For example, 23 % 10 = 3.

• a // b represents integer division. For example, 23 // 10 = 2.

What will foo(2021) return?

2. Food delivery agents Aman, Boni, Chan, Dong and Eman were assessed on five parameters P1 to P5. They received an integer rating between 1 and 5 for each parameter. None of them received the same rating in four or more parameters. Everyone received a score of 1 in P2 or P5. All of them received the same rating in at least two parameters. The partial information related to the ratings are given in the table below.

Name	P1	P2	P3	P4	P5
Aman				3	
Boni	4				
Chan			5		
Dong	2				
Eman				1	

Considering the above information, answer the following questions:

- (a) What is the maximum average rating Boni could have achieved across all the five parameters?
- (b) What is the minimum average rating Chan could have received across all the five parameters?

3. Consider the following code where A is an array indexed from 0.

```
function foo(A, year, n) {
    l = 0, r = n - 1, c = 0;
    while (1 \le r) 
        c = c + 1;
        m = 1 + (r - 1) // 2;
        if (A[m] == year) {
            return(c * m);
        }
        if (A[m] < year) {
            1 = m + 1;
        } else {
            r = m - 1;
        }
    }
    return(-1);
}
function bar() {
    A = [2016, 2017, 2018, 2019, 2020, 2021, 2022];
    result = foo(A, 2021, 7);
    print(result);
}
```

Here, a // b represents integer division. For example, 23 // 10 = 2. What will be printed when bar() is executed?

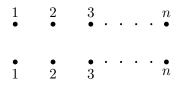
4. The CMI MSc DS 2021 batch of 60 students is holding an online event to celebrate their joining CMI. Student Aruni is in charge of organizing the musical section, and she sends out an online form where each student has to mark whether they agree or decline to performing two activities during the event: singing, and playing a musical instrument. Each student can mark one of four sets of choices: (i) agree to both singing and playing an instrument, (ii) agree to sing and decline to play an instrument, (iii) agree to play an instrument and decline to sing, or (iv) decline to do either activity.

All the students respond within the deadline, and Aruni sits down to tabulate the results so that she can plan the musical events. She finds that thirty five students agreed to sing or play an instrument, of whom twenty students agreed to do both.

How many students agreed to do exactly one activity, and how many declined to participate?

5. Show that every selection of 503 numbers from  $\{1, 2, 3, \ldots, 987\}$  has two numbers with g.c.d. 1. Recall that the g.c.d. of two positive integers x, y is the largest positive integer smaller than x, y which divides both x and y.

- 6. Let S be the sphere given by the equation  $x^2 + y^2 + z^2 6x + 4y 4z 272 = 0$ . What are the centre and radius of S? What is the minimum distance between S and the plane 2x + 2y + z = 214?
- 7. We draw two rows of n points with labels  $\{1, 2, ..., n\}$ , as shown in the figure below.



We connect each point on the top with a point on the bottom at random, making sure that no two points on the top are connected to the same point on the bottom. We say point j on the top is *special* if it gets connected to point j below.

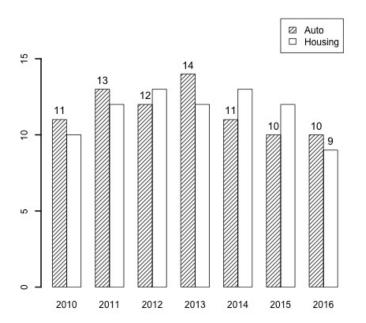
Let X be the number of special points. What is  $\mathbb{E}(X)$ , the expected value of X?

8. Let  $T_n$  denote the set of all sequences over  $\{0, 1, 2\}$  of length n. If we pick an element x in  $T_n$  uniformly at random what is the expected number of 1's in x?

and i are non-negative.

- 11. In the finale of the Indian Idol programme, Ladies Special, there are four women contestants Arunima, Priyamani, Razi and Shriya. The organisers send all participants a list of 7 songs, asking each one to pick one song for the finale. All the four girls pick the same song, *Pyar Hua Chupkese* from the film 1942 A Love Story. To resolve the tie, two days before the finale, the organisers prepare 7 sheets of paper and fold them. Exactly one of the 7 songs from the list is written on each folded sheet and the folded sheets are identical in all other aspects. First Arunima is asked to pick one sheet out of 7, then Priyamani is asked to pick one sheet out of the remaining 5 and finally Shriya is asked to pick one out of the remaining 4.
  - (a) What is the probability that the sheet chosen by Razi has the song *Pyar Hua Chupkese*?
  - (b) When Shriya is asked to choose a sheet, she argues that the three sheets chosen by the others should be opened, and if *Pyar Hua Chupkese* is not chosen by anyone, she should be allowed to pick that song. This is agreed to. What is the probability that Shriya gets to sing the song *Pyar Hua Chupkese*?

**Description for following two questions**: A Non-Banking Finance Corporation (NBFC) declares fixed annual rates of simple interest on their auto and housing loans each year. The rates of interest offered by the company differ from year to year depending on the variation in macro economic indicators like inflation, RBI's repo rate etc. The annual rates of interest offered by the company for the Auto and Housing sectors over the years are shown in the figure.



- 12. In 2013, a customer took a housing loan and a car loan. The total loan amount was ₹60 lakhs. The interest paid by the customer after one year was ₹7.36 lakhs. What was the housing loan amount?
- 13. In 2012, Customer A took a fixed interest auto loan of ₹5 lakh for 4 years, which meant he paid interest according to the 2012 rate each year. Also in 2012, Customer B took a variable interest auto loan for ₹5 lakh for 4 years which meant his interest each year was calculated based on the prevailing interest rate for that year. Which of them paid more interest over 4 years? How much more?
- 14. Let Y be a continuous random variable that takes values in [0, 1].
  - (a) g(y) is a probability density function that can be used to model Y,

$$g(y) = c.y^2(1-y), \quad 0 \le y \le 1$$

Determine c.

(b) Using g(y), find the probability that Y is greater than 0.8.

15. The state government has decided to regulate the sale and distribution of solar power. There are multiple suppliers and consumers, as described in the tables below. The supplier table lists the amount of electricity generated by each supplier, in megawatts (MW), and the supplier's selling price per MW. The consumer table lists the amount of electricity required by each consumer, in MW, and the price the consumer will pay per MW.

Supplier ID	Quantity	Price		Consur
1	400	150	] [	1
2	180	225		2
3	250	170		3
4	120	300		4
5	240	200		

Consumer ID	Quantity	Price
1	400	180
2	300	210
3	560	200
4	240	160

The government wishes to fix a uniform market price per MW for solar power. If this price is p, any seller whose selling price is at most p will be willing to sell some or all of the electricity they generate at price p. Similarly, any consumer who is willing to pay at least p can buy as much electricity as they need at price p, upto their required quantity.

Find the value of p at which the maximum amount of solar power can be bought and sold on the market.

- 16. (a) If a polynomial f(x) has roots  $r_1, r_2, \ldots r_n$ , what are the roots of the polynomial f(x+1)?
  - (b) Let f be a polynomial of degree three. Set g(x) = f(x-1). Assume that -1 is not a root of f. What are the roots of the polynomial  $x^3g(1/x)$ ?
- 17. Is the following statement true?

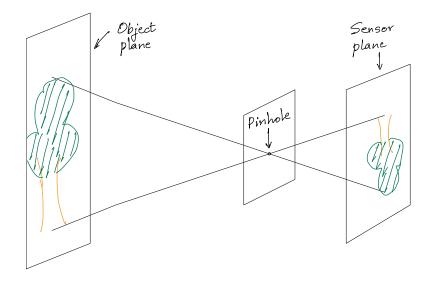
If 
$$f(x) \ge 0$$
 for all  $x$ , and  $\int_{-\infty}^{\infty} f(x) dx < \infty$  then  $\int_{-\infty}^{\infty} x^2 f(x) dx \ge \epsilon^2 \int_{\epsilon}^{\infty} f(x) dx$ , for all  $\epsilon > 0$ .

- 18. (a) Consider the function  $f(x) = \frac{\ln x}{x}$ . Find the critical point(s) of f and say whether f has a maximum or minimum at that point.
  - (b) Prove that for integers  $a > b \ge 3$ , we must have  $a^b < b^a$ .
- 19. Recall that if h is a function from X to Y and g is a function from Y to Z then,  $g \circ h$  is the function from X to Z such that  $(g \circ h)(x) = g(h(x))$ , for all  $x \in X$ .

Let S be the set of all functions f from  $\{1, 2, 3, 4, 5, 6\}$  to  $\{1, 2, 3, 4, 5, 6\}$  such that  $f \circ f = f$ .

- (a) Compute the number of functions  $f \in S$  whose range has three elements.
- (b) What is the cardinality of S?

20. A pinhole camera is placed between an object and a sensor. Recall that images from a pinhole camera are inverted (see figure below). Assume that the object lies on a plane. Also assume that the plane containing the object, the plane containing the pinhole and the sensor plane are all parallel to each other.



- (a) Show that straight lines are projected to straight lines. In particular, show that if A, B, C are three collinear points on an object, then their corresponding images a, b, c are also collinear.
- (b) Suppose two identical objects M and N having height 60cm are placed before the pinhole camera at distances 90cm and 120cm respectively. Suppose also that the distance between the pinhole and the sensor plane is 30cm. What are the heights of their images after projection through the pinhole camera?