CHENNAI MATHEMATICAL INSTITUTE

M.Sc. Data Science Entrance Examination 2022 – Solutions

Part A - Answers

A1.	(a),(c)	A11.	(c)
A2.	(d)	A12.	(a),(c),(d)
A3.	(b),(d)	A13.	(b)
A4.	(a),(c)	A14.	(c)
A5.	(c),(d)	A15.	(a),(c),(d)
A6.	(a),(c),(d)	A16.	(b)
A7.	(d)	A17.	(b),(d)
A8.	(b),(d)	A18.	(c)
A9.	(c)	A19.	(c)
A10.	(d)	A20.	(a),(d)

Part (A) – Multiple-choice questions

1. Let X be a Binomial(n, p) random variable with the probability mass function

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

Let T be a random variable defined as:

$$T = \begin{cases} 1 & \text{if } X = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbb{E}(T)$ and $\mathbb{V}(T)$ denote the expectation and the variance of a random variable T. Which of the following statements is/are true?

(a) $\mathbb{E}(T^2) = np(1-p)^{n-1}$ and $\mathbb{V}(T) = np(1-p)^{n-1}\{(1 - (np(1-p)^{n-1}))\}$

(b) $\mathbb{E}(T) = np$ and $\mathbb{V}(T) = np(1-p)$

(c) $\mathbb{E}(T) = np(1-p)^{n-1}$ and $\mathbb{E}(T^2) = np(1-p)^{n-1}$

(d) $\mathbb{E}(T) = p$ and $\mathbb{V}(T) = p(1-p)$

> **Solution:** Choices (a) and (c) are correct, and choices (b) and (d) are wrong. These follow from straightforward computation using standard formulas.

2. Consider the following code, in which `A` is an array indexed from `0` and `n` is the number of elements in `A`..

```
function foo(A,n) {
    L = 0;
    R = n - 1;

    while (L <= R) {
        i = ceil((L + R)/2);

        if (A[i] < i) {
            L = i + 1;
        } else {
            if (A[i] > i) {
                R = i - 1;
            } else {
                return(i);
            }
        }
    }

    return(-1);
}
```

Here, `ceil(x)` returns the smallest integer bigger than or equal to the number `x`.

If `A = [-5, -4, -3, -2, -1, 4, 6, 8, 10, 12]`, what will `foo(A, 10)` return?

(a) -5

(b) -1

(c) 4

(d) 6

Solution: Choice (d) is correct. Given a sorted array `A` with distinct integers and the length of `A` as arguments, this function finds an index `i` such that `A[i] = i`, or returns `-1` if there is no such index `i`. Since there is exactly one such index, namely 6, this is what we get back.

3. Which of the following statements is/are true?

(a) For any real number r with $|r| > 1$, $\lim_{n\to\infty} \sum_{i=1}^{n} 17r^n = \frac{17}{1-r}$.

(b) Let i be the positive square root of -1 and let x be any real number. Then

$$\tan x = \frac{(e^{ix} - e^{-ix})}{i(e^{ix} + e^{-ix})}.$$

(c) $\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots < 1.$

(d) The function $f(x)$ defined as

$$f(x) = \begin{cases} 3\cos 3x, & 0 < x < \pi/6 \\ 0 & \text{otherwise,} \end{cases}$$

is a probability density function.

Solution:

Choices (b) and (d) are correct, and choices (a) and (c) are wrong.

- (a) is wrong: This is a GP series. For $|r| > 1$ the limit is divergent.
- (b) is correct: Since $e^{-ix} = \cos x - i \sin x$, we have $\sin x = \frac{(e^{ix} - e^{-ix})}{2i}$, and $\cos x = \frac{(e^{ix} + e^{-ix})}{2}$.
- (c) is wrong:

$$\begin{aligned}
\frac{1}{1.2} + \frac{1}{2.3} + \frac{1}{3.4} + \cdots &= \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k(k+1)} \\
&= \lim_{n\to\infty} \left[\sum_{k=1}^{n} \frac{1}{k} - \frac{1}{k+1} \right] \\
&= \lim_{n\to\infty} \left[1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots + \frac{1}{n} - \frac{1}{n+1} \right] \\
&= \lim_{n\to\infty} \left[1 - \frac{1}{n+1} \right] = 1.
\end{aligned}$$

- (d) is correct: $3\cos 3x \geq 0 \ \forall x \in (0, \pi/6)$, and $\int_0^{\pi/6} 3\cos 3x \, dx = 3\left[\frac{1}{3}\sin 3x\right]_0^{\pi/6} = \sin\frac{3\pi}{6} = \sin\frac{\pi}{2} = 1.$

4. Let E and F be events such that $P(E \cap F^c) = 0.2$, $P(F \cap E^c) = 0.3$, $P((E \cap F)^c) = 0.7$, where, for an event E, the notation E^c denotes the complement of the event. Then we can conclude that

   (a) $P(E \cup F) = 0.8$.
   (b) $P(E^c \cap F^c) = 0.3$.
   (c) $P(F) = 0.6$.
   (d) $P(E) = 0.6$.

Solution: Choices (a) and (c) are correct, and choices (b) and (d) are wrong. These follow from straightforward computation, e.g., by using a Venn diagram.

5. Let A, B be $n \times n$ invertible matrices of real numbers. Let $C = I + AA^T$, $D = I + BAA^TB^T$. We can conclude that

   (a) $(AB)^{-1} = (I + B^{-1}A^{-1})$
   (b) $(AB)^{-1} = A^{-1}B^{-1}$
   (c) $C = C^T$
   (d) $D = D^T$

Solution: Choices (c) and (d) are correct, and choices (a) and (b) are wrong. These follow by straightforward computation using standard properties of matrices.

6. A matrix C is said to be symmetric if $C^T = C$. Which of the following is/are true? Let A, B be $n \times n$ matrices.

   (a) If A is symmetric and invertible, then A^{-1} is also symmetric and invertible.
   (b) If A and B are symmetric, then $C = AB$ is also symmetric.
   (c) If A and B are invertible, then $C = AB$ is also invertible.
   (d) If A and B are symmetric, then $D = A + B$ is also symmetric.

Solution: Choices (a), (c) and (d) are correct, and choice (b) is wrong. These follow from standard properties of matrices.

7. Which of the following statements are true?

   (a) Let A be a $n \times m$ matrix with $n < m$. Then there is a nonzero solution y with $Ay = 0$ only if A has full row rank.
   (b) Let A be a $n \times m$ matrix with $n > m$. There is a nonzero solution y with $Ay = 0$.

(c) The row rank of an $n \times m$ matrix is equal to its column rank only when $n = m$.

(d) Let A be an $n \times n$ matrix. Suppose $A = BC$, where B has size $n \times r$ and C has size $r \times n$. The rank of A is less than or equal to r.

Solution: Choice (d) is correct, and choices (a), (b) and (c) are wrong:

- (a) is wrong: If A is all zeroes, then $Ay = 0$ for any vector y, and A is clearly not full rank.
- (b) is wrong: Let the first m rows of A be identical to the $m \times m$ identity matrix, and let the remaining rows be all zero. Then there is no such nonzero solution y.
- (c) is wrong: Consider
$$\begin{bmatrix} I_5 \\ 0 \end{bmatrix}.$$
This matrix has $n = 6, m = 5$, and row and column ranks both equal to 5.
- (d) is correct: The rank of A is the least such r.

8. Let $n \geq 5$ be a natural number, let $X = \{x_1, x_2, ..., x_n\}$, and let $Y = \{y_1, y_2\}$. Let F be the set of functions from X to Y and G be the set of bijective functions from X to Y. Then

(a) The number of functions in F equals n^2.

(b) The number of functions in F equals 2^n.

(c) The number of functions in G equals $n^2 - 2$.

(d) The number of functions in G equals 0.

Solution: Choices (b) and (d) are correct, and choices (a) and (c) are wrong:

- (a) is wrong: The number of functions in F is 2^n.
- (b) is correct: The same reason as (a).
- (c) is wrong: Since X, Y have different cardinalities, there are no bijections from X to Y.
- (d) is correct: The same reason as (c).

9. In order to select a debating team to represent a school, 7 students from class XII and 13 students from class XI were shortlisted and were undergoing trials. The coach had to select a team of 5 students, out of which at least two students should be from each class. One out of the 5 students was to be named as team leader, who was to be from class XII. Two teams with the same members but different leaders are considered to be two different teams. The number of different teams the coach can select is

(a) $2 \times \left(\binom{7}{2} \times \binom{13}{3} + \binom{7}{3} \times \binom{13}{2} \right)$

(b) $\binom{7}{2} \times \binom{13}{3} + \binom{7}{3} \times \binom{13}{2}$

(c) $2 \times \binom{7}{2} \times \binom{13}{3} + 3 \times \binom{7}{3} \times \binom{13}{2}$

(d) $3 \times \binom{7}{2} \times \binom{13}{3} + 2 \times \binom{7}{3} \times \binom{13}{2}$

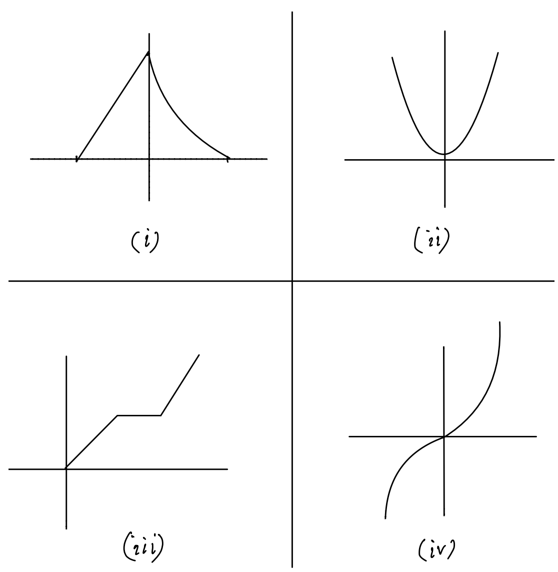
Solution: Choice (c) is correct, and choices (a), (b) and (d) are wrong.

Any choice of team either has exactly 2 students from class XII and exactly 3 students from class XI, **OR** it has exactly 3 students from class XII and exactly 2 students from class XI. The two terms in the sum in choice (c) correspond to these two cases.

For each choice of $k \in \{2, 3\}$ students from class XII, there are k ways of choosing the leader. This is the leading term in each product in choice (c).

Here is an alternate derivation: There are 7 ways of selecting a leader. For each leader selected, we can now pick either a single student from class XII or two students from class XII and the rest need to be from class XI. To form a team by selecting a single non-leader student from class XII, there are $7 \cdot \binom{6}{1} \cdot \binom{13}{3}$ ways. To form a team by selecting exactly two non-leader class XII students, there are exactly $7 \cdot \binom{6}{2} \cdot \binom{13}{2}$ ways. The answer is the sum of these.

10. Which of the following plots correspond to a function whose derivative is continuous?



(a) (i) and (ii)

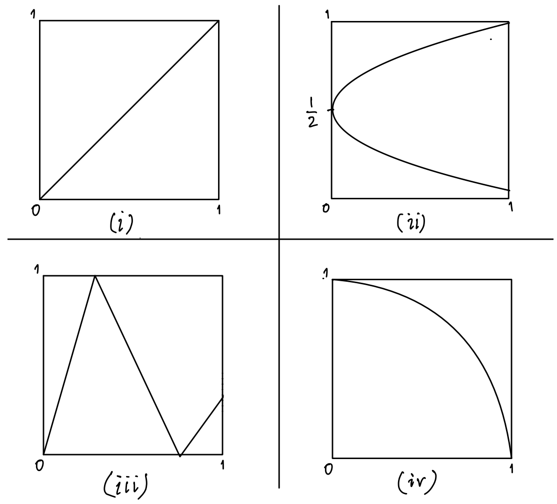
(b) (ii) and (iii)

(c) (iii) and (iv)

(d) (ii) and (iv)

Solution: The only correct choice is (d), by inspection.

11. Which of the following plots correspond to a bijective function?



(a) (i) and (ii)
(b) (ii) and (iii)
(c) (i) and (iv)
(d) (iii) and (iv)

Solution: The only correct choice is (c), by inspection.

12. A relation R on the set $A = \{a, b, c, d\}$ is defined by reading the columns of the following table from top to bottom. If a column in the table reads $(x, y, 1)$ it means x is related to y in R. If a column in the table reads $(x, y, 0)$ it means x is not related to y.

a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
a	a	a	a	b	b	b	b	c	c	c	c	d	d	d	d
0	0	1	0	1	0	0	0	1	1	0	1	0	0	1	0

For instance, from the fifth column we have, $(a, b) \in R$, and from the second we have $(b, a) \notin R$.

Another relation S on the set A is defined as: for any $x, y \in A$, the pair (x, y) is in S if and only if there exists $z \in A$ such that *both* $(x, z) \in R$ and $(z, y) \in R$ hold.

Which of the following pairs are in S?

(a) (a, a)
(b) (b, b)
(c) (c, c)
(d) (d, d)

Solution:

Choices (a), (c), and (d) are correct, and choice (b) is wrong:

- (a) is correct: $(a, a) \in S$ since $(a, c) \in R$ and $(c, a) \in R$.
- (b) is wrong: $(b, b) \notin S$ since c is the only image of b, and $(c, b) \notin R$.
- (c) is correct: $(c, c) \in S$ since $(c, a) \in R$ and $(a, c) \in R$.
- (d) is correct: $(d, d) \in S$ since $(d, c) \in R$ and $(c, d) \in R$.

13. We say that a subset S of a finite set U is *large* if $|S| > |U \setminus S|$. Here $U \setminus S$ denotes the elements of U which are not in S and the notation $|T|$ denotes the number of elements in a set T. Let x be the number of large subsets of the set $X = \{1, 2, \ldots, 10\}$, and let y be the number of large subsets of $Y = \{1, 2, \ldots, 9\}$. Which of the following is true?

    (a) $x = 260, y = 256$
    (b) $x = 386, y = 256$
    (c) $x = 386, y = 130$
    (d) $x = 512, y = 256$.

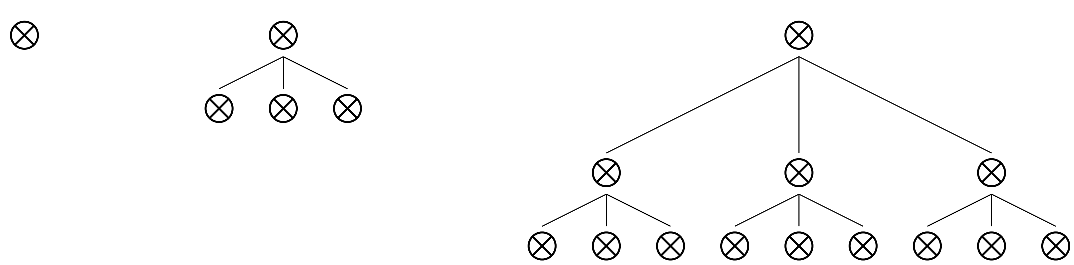
Solution: Choice (b) is correct; the other choices are wrong.

When we choose a subset we can pair it with its complement. So the number of pairs of subsets is 2^9, since 2^{10} is the number of subsets. From a pair we can always choose the larger one unless both have 5 elements. So we get $x = 2^9 - (\frac{1}{2})\binom{10}{5} = 386$, $y = 2^8 = 256$.

14. A ternary tree starts with a single root node at the top of the tree. Each node in the tree can have up to three nodes as its children. No node in the tree is the child of two different nodes. A node which has no children is called a *leaf* node.

    The children of a node are drawn below it, connected by edges. The *level* of a node v in the ternary tree is the number of edges in the (unique) path from the root node to v. Thus, for instance, the root node is at level 0, and each child of the root node is at level 1.

    A *complete* ternary tree is a ternary tree in which (i) each non-leaf node has *exactly* three children, and (ii) all leaf nodes are at the same level. This latter level is called the *height* of the complete ternary tree. The complete ternary trees of heights $0, 1$, and 2, respectively are shown in the figure below, where we use the symbol $\otimes$ to denote a node:



What is the total number of nodes in a complete ternary tree of height 9?

(a) $2^{11}-1$
(b) $\frac{2^{10}+2}{3}$
(c) $\frac{3^{10}-1}{2}$
(d) $\frac{3^{10}+1}{2}$

Solution: Choice (c) is correct: the number of nodes is the value of the sum $1+3^1+3^2+\ldots+3^9$. Choices (a), (b), (d) are wrong.

15. Which of the following expressions take integral values for all integers $n > 100$?

(a) $\frac{3^n-1}{2}$
(b) $\frac{4^n-1}{2}$
(c) $\frac{4^n-1}{3}$
(d) $\frac{5^n-1}{2}$

Solution:

Choices (a), (c) and (d) are correct, and choice (b) is wrong:

- (a) is correct: the expression is of the form $\frac{a^n-1}{a-1}$ which is equal to the sum $1+a+a^2+\cdots+a^{n-1}$.
- (b) is wrong: 4^n-1 is an odd number for every positive integer n, since it is one less than a power of 4. So it is not evenly divisible by 2.
- (c) is correct: the expression is of the form $\frac{a^n-1}{a-1}$ which is equal to the sum $1+a+a^2+\cdots+a^{n-1}$.
- (d) is correct: 5^n-1 is an even number for every positive integer n, since it is one less than a power of 5 which is an odd number.

16. At a conference attended by 1235 people, some attendees shake hands with other attendees. As a part of the local COVID-19 tracing protocol the organizers ask each attendee to note down the *number* of other attendees with whom they shook hands. Let $\mathcal{N}$ be the *sum* of all the numbers noted down by the attendees. That is, $\mathcal{N}$ is the sum, taken over all attendees, of the number of other attendees with whom they shook hands. Which of the following is guaranteed to be true about $\mathcal{N}$?

(a) $\mathcal{N}$ is always a multiple of 1235
(b) $\mathcal{N}$ is always a multiple of 2
(c) $\mathcal{N}$ is always a multiple of 3
(d) $\mathcal{N}$ is always a multiple of 5

Solution: The only correct choice is (b): Each handshake adds 2 to $\mathcal{N}$, so $\mathcal{N}$ must be an even number.

17. Let n be a positive integer and let k be a positive integer which is greater than 1. Then n can be represented as

$$n = n_t k^t + n_{t-1} k^{t-1} + \cdots + n_1 k + n_0$$

for a unique set of positive integers $n_0, \ldots n_t$. This expression is called the representation of n in base k and is denoted by $(n_t\ n_{t-1}\ \ldots\ n_0)_k$. For example the integer 264 expressed in base 3 is represented as $(1\ 0\ 0\ 2\ 1\ 0)_3$ because $264 = 1 \times 3^5 + 2 \times 3^2 + 1 \times 3$.

Which of the following expressions represent the integer $(1\ 1\ 0\ 1\ 1\ 1\ 1\ 0)_2$?

(a) $(1\ 1\ 0\ 0)_8$

(b) $(1\ 3\ 4\ 2)_5$

(c) $(4\ 3\ 2)_7$

(d) $(2\ 2\ 0\ 2\ 0)_3$

Solution: Choices (b) and (d) are correct, and choices (a) and (c) are wrong. The given integer, in decimal notation, is 222.

18. There are two villages X and Y in a faraway land. It is known that each person from village X always tells the truth, and that each person from village Y always lies.

You meet three people A, B, C who are from these villages. You are told that A and B are from the same village. A says, "If B is from village X, then I am from village Y". Now C says, "If I am from village X, then $2 + 2 = 4$".

What can you infer about the villages to which A, B, and C belong?

(a) A, B are from village X and C is from village Y.

(b) All three of them are from village Y.

(c) A, B are from village Y and C is from village X.

(d) The given information is insufficient to infer the villages to which A, B, C belong.

Solution: The only correct choice is (c).

A, B cannot both be truthful since, if that happens then A's statement would be a lie. So, A, B must both be liars (from village Y). C cannot be a liar, since if that is the case then his/her statement will be true. So C must be from village X.

19. A stick is fixed on the ground at an angle of 60 degrees as shown in Figure 1. The part of the stick which is above the ground—which is the part that is shown in Figure 1—has a length of 10 metres. At a certain time T_1 during the day, the sun is directly overhead at this location and its rays strike the ground at an angle of 90 degrees. This is shown in Figure 2. At a certain later time T_2 the sun's rays strike the ground at an angle of 60 degrees as shown in Figure 3. While Figures 2 and 3 show only the sun's rays for the sake of clarity, the stick is present at this location at times T_1 and T_2.

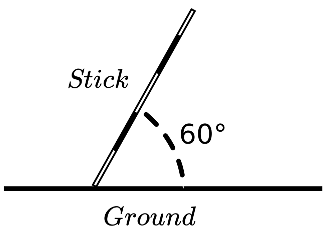


Figure 1

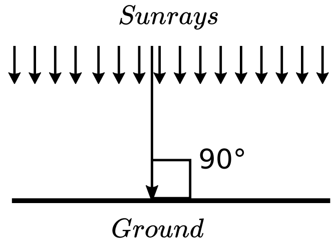


Figure 2

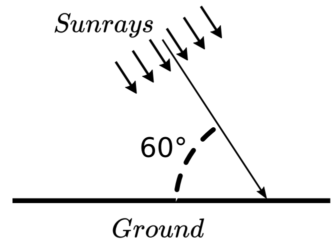


Figure 3

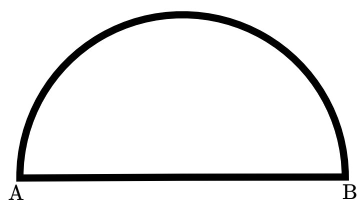
What are the lengths of the shadows of the stick at times T_1 and T_2, respectively?

(a) $5\sqrt{3}$ metres and $10\sqrt{3}$ metres
(b) 5 metres and $10\sqrt{3}$ metres
(c) 5 metres and 10 metres
(d) $5\sqrt{3}$ metres and 10 metres

Solution: The only correct choice is (c).

The length of the shadow at time T_1 is $10m \times \cos 60^\circ = 5m$. At time T_2 the stick, the ray of the sun from the tip of the stick to the ground, and the shadow of the stick make an equilateral triangle since all angles of this triangle are 60° each. So the shadow has the same length as the stick, which is $10m$ long.

20. A park in Chennai has walkways as depicted in the figure. The straight path from A to B is 100 metres long. The curved part is a semicircle whose centre coincides with the midpoint of line segment AB.



A woman and her dog take a walk in this park, both starting from point A at the same time. The woman takes the straight path from A to B, while the dog goes around the semicircular path. The woman and the dog both reach point B at the same time. If the woman takes 10 minutes to reach from point A to point B, which of the following statements about the woman's and dog's walks is/are true?

(a) The average speed of the woman's walk is between $0.15ms^{-1}$ and $0.17ms^{-1}$
(b) The average speed of the woman's walk is between $0.17ms^{-1}$ and $0.20ms^{-1}$
(c) The average speed of the dog's walk is between $0.25ms^{-1}$ and $0.27ms^{-1}$
(d) The average speed of the dog's walk is between $0.27ms^{-1}$ and $0.30ms^{-1}$

Solution: Choices (a) and (d) are correct, and choices (b) and (c) are wrong:

- The average speed of the woman's walk is $\frac{100m}{600s} \approx 0.1667ms^{-1}$.
- The length of the semicircular path is $\pi \times 50m \approx 157m$, with $\pi \approx 3.14$. The average speed of the dog's walk is thus $\approx \frac{157m}{600s} \approx 0.262ms^{-1}$.

Note that to get these answers right it is enough to compute that $0.15 < \frac{1}{6} < 0.20$, and $0.25 < \frac{160}{600} < 0.3$.

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 three different approaches that climbers can take to reach the summit of Mount Chernoff, namely, APPROACH-I, APPROACH-II, and APPROACH-III. Each attempt at the summit uses exactly one of these approaches. From climbers' logs it is known that APPROACH-I is used in 40% of the attempts, and APPROACH-II and APPROACH-III are each used in 30% of the attempts. From the same logs it is also known that in 10% of the attempts which used APPROACH-I the climber lost their way and had to be rescued. Similarly, 15% of the attempts that used APPROACH-II and 17% of the attempts which used APPROACH-III resulted in the climber losing their way.

   It is reported one day that climber Chebyshev has lost her way. What is the probability that she took APPROACH-II?

Solution: $\boxed{\frac{0.045}{0.136} \approx 0.3308824}$.

The following information is given in the description:

- $\mathsf{P}(\text{APPROACH-I}) = 0.4$ • $\mathsf{P}(\text{APPROACH-II}) = 0.3$ • $\mathsf{P}(\text{APPROACH-III}) = 0.3$
- $\mathsf{P}(\text{lost} \mid \text{APPROACH-I}) = 0.1$ • $\mathsf{P}(\text{lost} \mid \text{APPROACH-II}) = 0.15$ • $\mathsf{P}(\text{lost} \mid \text{APPROACH-III}) = 0.17$

We want to find $\mathsf{P}(\text{APPROACH-II} \mid \text{lost})$. This can be computed as follows.

$$\begin{aligned}\mathsf{P}(\text{APPROACH-II} \mid \text{lost}) &= \frac{\mathsf{P}(\text{APPROACH-II and lost})}{\mathsf{P}(\text{lost})} \\ &= \frac{\mathsf{P}(\text{lost} \mid \text{APPROACH-II})\mathsf{P}(\text{APPROACH-II})}{\mathsf{P}(\text{lost})}\end{aligned}$$

Now,

$$\begin{aligned}\mathsf{P}(\text{lost}) &= \mathsf{P}(\text{lost} \mid \text{APPROACH-I})\mathsf{P}(\text{APPROACH-I}) \\ &\quad + \mathsf{P}(\text{lost} \mid \text{APPROACH-II})\mathsf{P}(\text{APPROACH-II}) \\ &\quad + \mathsf{P}(\text{lost} \mid \text{APPROACH-III})\mathsf{P}(\text{APPROACH-III}) \\ &= 0.1 * 0.4 + 0.15 * 0.3 + 0.17 * 0.3 \\ &= 0.136.\end{aligned}$$

Hence

$$\mathsf{P}(\text{APPROACH-II} \mid \text{lost}) = \frac{0.15 * 0.3}{0.136} \approx 0.3308824$$

Description for the next four questions:

Dasholytics Inc runs an online analytics dashboard. The company employs three machines—Server 1, Server 2, and Server 3—to serve the data for the dashboard. At any point in time one of these three machines has the job of serving the data, and the other two are kept in standby mode. Dasholytics uses a server scheduler (software) that decides which of the three machines should serve the data at any given point of time. The scheduler switches between data servers without disrupting the data feed to the dashboard.

The machine which is serving the data sometimes fails to do so; in this case an alert is sent out to the Dasholytics team and they investigate and fix the problem. The time for which a machine fails to serve data is accounted as service outage caused by that machine. For technical reasons the server scheduler is deactivated when there is such an outage; the outage gets over only when the issue with the server is fixed and it is put back online.

The figures below describe the server usage and outage statistics as compiled over the last one year (365 days). Please use this information to answer the questions that follow.

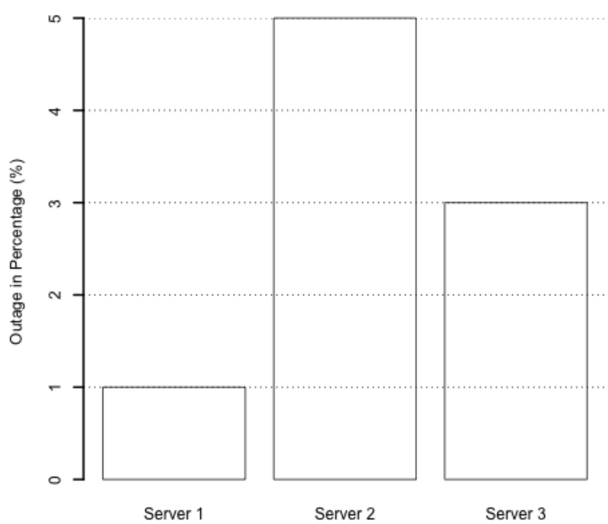


Fig. (a)

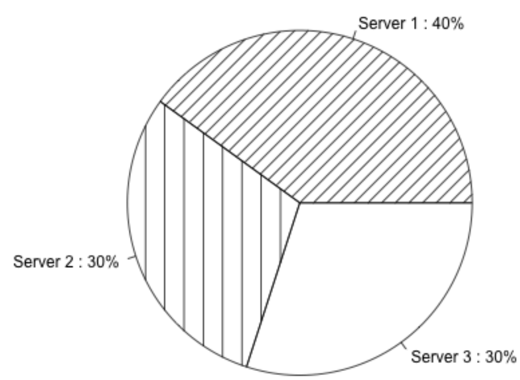


Fig. (b)

Figure 1: Figure (a) shows the total service outage caused by each server, as a percentage of the total time that it served data. Figure (b) shows the total time that each server served data, as a percentage of the total duration over which this information was collected.

2. Express the total service outage caused by all the three servers, as a percentage of the total duration over which the information was collected.

Solution: $\boxed{2.8\%}$.

The following information is given in the description:

1. $P(Outage \mid Server\ 1) = 0.01$ and $P(Server\ 1) = 0.4$
2. $P(Outage \mid Server\ 2) = 0.05$ and $P(Server\ 2) = 0.3$
3. $P(Outage \mid Server\ 3) = 0.03$ and $P(Server\ 3) = 0.3$

So we get

$$\begin{aligned} P(\text{Outage}) &= P(\text{Outage} \mid \text{Server 1}) \times P(\text{Server 1}) \\ &\quad + P(\text{Outage} \mid \text{Server 2}) \times P(\text{Server 2}) \\ &\quad + P(\text{Outage} \mid \text{Server 3}) \times P(\text{Server 3}) \\ &= 0.01 \times 0.4 + 0.05 \times 0.3 + 0.03 \times 0.3 \\ &= 0.028 = 2.8\%. \end{aligned}$$

3. What is the expected number of hours of service outage in an year (365 days)?

Solution: $\boxed{245.28 \approx 245 \text{ hours}}$.

From the previous question, we know that there is an outage 2.8% of the time. So the expected number of hours of outage over a year is $365 \times 24 \times 0.028 = 245.28 \approx 245$ hours.

4. As part of their due diligence, a potential customer of Dasholytics picks a random point in time during the year and checks the status of the dashboard. What is the probability that the customer finds that there was a service outage at this point in time, and that Server 1 was the server responsible for serving data at that time?

Solution: $\boxed{0.004}$.

The chance of service outage due to Server 1 is

$$P(\text{Outage and Server 1}) = P(\text{Outage} \mid \text{Server 1}) \times P(\text{Server 1}) = 0.01 \times 0.4 = 0.004,$$

5. An alert that a service outage has been caused by one of the three servers, was sent out to the Dasholyltics team. What is the probability that this outage was caused by Server 1?

Solution: $\boxed{\approx 0.1429, \text{ or } \approx 14.29\%}$

$$\begin{aligned} P(\text{Server 1} \mid \text{Outage}) &= \frac{P(\text{Outage} \mid \text{Server 1})P(\text{Server 1})}{P(\text{Outage})} \\ &= \frac{0.01 \times 0.4}{0.028} \\ &\approx 0.1429. \end{aligned}$$

6. In a question paper, there are 5 questions in part A, 4 in part B, and 3 in part C. In how many ways can a candidate make up her choice, if she has to select 3 from part A, 2 from part B, and 2 from part C?

Solution: $\boxed{180}$

The possible number of selections is

$$\binom{5}{3} \times \binom{4}{2} \times \binom{3}{2} = \frac{5 \times 4 \times 3}{3 \times 2} \times \frac{4 \times 3}{2 \times 1} \times \frac{3 \times 2}{2 \times 1} = 180.$$

Description for the next two questions

The probability density function of a normal distribution with mean μ and variance σ^2 is of the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5(\frac{x-\mu}{\sigma})^2}, \quad -\infty < x < \infty.$$

Let X be a random variable with mean $\mathbb{E}(X) = \mu$ and variance $\mathbb{V}ar(X) = \sigma^2$. Let $X_1, X_2, \ldots, X_n$ be n independently sampled values of X, and let $\bar{X} = \frac{1}{n}\sum_i X_i$ be the sample mean. The central limit theorem states that if the sample size n is large enough then $\bar{X}$ approximately follows the normal distribution with mean μ and variance σ^2/n. That is, $\bar{X} \approx N(\mu, \sigma^2/n)$. This in turn implies that

$$Y = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \approx N(0, 1).$$

Let Z be a random variable that follows the normal distribution with mean 0 and variance 1. For any real number a let $\Phi(a) = \mathbb{P}(Z < a)$ be the probability that Z takes values smaller than a. Then

$$\Phi(a) = \mathbb{P}(Z < a) = \int_{-\infty}^{a} \frac{e^{-u^2/2}}{\sqrt{2\pi}} du.$$

For solving the next two problems you may assume the following approximations: $\Phi(-2) = 0.02$, $\Phi(-1) = 0.16$, $\Phi(0) = 0.5$, $\Phi(1) = 0.84$, $\Phi(2) = 0.98$.

The weekly number of sales S at a certain car dealership is known to follow a probability distribution with mean $\mathbb{E}(X) = 10$ and variance $\mathbb{V}ar(X) = 144$. A performance audit picks a random sample of 36 weekly sales figures $S_1, S_2, \ldots, S_{36}$ from the last two years. They find that the sample mean is 10 and the sample variance is 144. Use this information to answer the next two questions. You may assume that $n = 36$ is a large enough sample size.

7. (a) What is the probability that the average number of sales in a week will be more than 8?

Solution: $\boxed{0.84}$

Note that $\bar{X} \approx N(10, 2)$, where $\mu = 10$, $\sigma = 12$, and $n = 36$. So we have

$$\begin{aligned} \mathbb{P}(\bar{X} > 8) &= \mathbb{P}\Big(\frac{\sqrt{36}(\bar{X} - 10)}{12} > \frac{\sqrt{36}(8 - 10)}{12}\Big) \\ &= \mathbb{P}\Big(Z > \frac{6 \times -2}{12}\Big) = \mathbb{P}(Z > -1) = 1 - \mathbb{P}(Z < -1) \\ &= 1 - \Phi(-1) = 1 - 0.16 = 0.84. \end{aligned}$$

(b) What is the total number of sales over the 36 weeks which were sampled?

Solution: $\boxed{360}$
The average sales per week is 10, i.e., $\bar{X} = 10$. Since the number of weeks $n = 36$, the total sales is $\sum X_i = n \times \bar{X} = 36 \times 10 = 360$.

8. Check if the following statement(s) are correct. Briefly explain your reasons.

   (a) The probability that the average number of sales in a week will be more than 8 but less than 14 is

$$\int_8^{14} \frac{1}{12\sqrt{2\pi}} e^{-0.5(\frac{u-10}{12})^2} du$$

Solution: $\boxed{\text{Wrong.}}$ The correct expression is

$$\int_8^{14} \frac{\sqrt{36}}{12\sqrt{2\pi}} e^{-0.5(\frac{\sqrt{36}(u-10)}{12})^2} du.$$

   (b) The probability that a salesperson would be able to sell ten or more products in a week is 50%.

Solution: $\boxed{\text{Correct.}}$
It is given that the weekly average sales follows a normal distribution with mean 10. The probability that the weekly average sales is at least the mean, is half.

9. Two particles move in opposite directions around a circular track. The first moves at a constant speed of 10 m/s. The speed of the second increases at a constant rate of 2 m/s every second. The particles are at the same position A at time 0, with the second particle being momentarily at rest at $t = 0$. We are told that the second meeting of these particles after time 0 takes place at the point A. We assume that the particles magically cross each other the first time they meet with no change in their instantaneous velocities. Given this information:

   (a) What is the circumference of the track?

Solution: $\boxed{\text{100 metres.}}$
Say they met for the first time at t_1 and at t_2 the second time. Then the circumference is equal to $10t_1 + 1/2 * 2 * t_1^2$. But this is also $10t_2$ and is also equal to $1/2 * 2 * t_2^2$. So we have $10t_2 = t_2^2$, so $t_2 = 10$. So the circumference is $10t_2 = 100$ metres.

   (b) At what time did the particles meet for the first time after time 0?

Solution: $\boxed{(\sqrt{125} - 5) = (5\sqrt{5} - 5) = 5 * (\sqrt{5} - 1) \approx 6.18 \text{ seconds.}}$
Solving $100 = 10t_1 + t_1^2$, we get $t_1 = (-10 + \sqrt{500})/2 = 5 * (\sqrt{5} - 1)$.

10. Words are formed using the characters 0 and 1. The length of a word is the number of characters in it. We say there is a path from word x to word y if starting from the word x you can get the word y by applying the following sequence of transformation rules finitely many times in any order.

   - Replacing an occurrence of the string 101 by a 0.

- Replacing an occurrence of the string 010 by a 1.
- Replacing a 0 by a 101.
- Replacing a 1 by by a 010.

If there is a path from word x to word y we say x and y are equivalent or that they are in the same equivalence class. For example the four letter word 1011 is equivalent to the two letter word 01, since the prefix 101 in 1011 can be replaced by a 0 to get 01.

We say x has a shorter description if there is a word of shorter length equivalent to it.

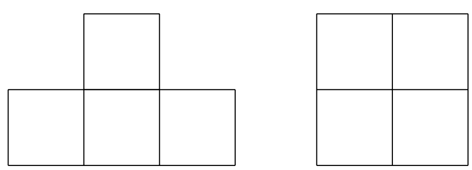
(a) State true or false: for any word x there is a unique shortest word in its equivalence class.

> **Solution:** $\boxed{\text{False.}}$
> E.g.: $1010 \equiv 11$, and $1010 \equiv 00$.

(b) How many three letter words are there in this language which have no shorter descriptions and which are all in different equivalance classes?

> **Solution:** $\boxed{\text{Two. One such choice is: 000 and 111.}}$
> 101 has the shorter description 0. 010 has the shorter description 1.
> Of the remaining six three-letter strings, none has a shorter description (by inspection). The following four map to 000 or 111:
>
> $$001 \rightarrow 10101 \rightarrow 111$$
> $$011 \rightarrow 01010 \rightarrow 000$$
> $$100 \rightarrow 10101 \rightarrow 111$$
> $$110 \rightarrow 01010 \rightarrow 000$$

11. You are given an 8×8 chessboard and a large supply of T-shaped tiles and square-shaped tiles of the kind shown below.



Note that the squares in these two shaped tiles are of the same size as the black/white squares on the chessboard.

We wish to cover the squares of the chess board using these tiles. Each square of the chessboard must be occupied by exactly one square from a tile; we are not allowed to break the tiles.

(a) Is it possible to cover the chessboard with 16 T shaped tiles? Why?

> **Solution:** $\boxed{\text{Yes.}}$
> Greedy tiling will work for a 4×8 grid, by trial and error. Doing this twice will cover the entire chess board.

(b) Is it possible to cover the chessboard with 15 T shaped tiles and one square tile? Why?

Solution: $\boxed{\text{No.}}$

Note that a square tile occupies two black and two white squares on the chessboard. So 15 square tiles occupy 30 black and 30 white squares in total. There are two white and two black squares left to cover. A T-shaped tile occupies either 3 black squares and a white square, or 3 white squares and a black square. So we cannot cover the remaining four squares with a T-shaped tile.

12. To guard against login attempts by bots, Alphabet Bank of India uses a captcha challenge on their net banking login page. The captcha text is a sequence of five letters from the English alphabet.

(a) An early version of the captcha text consisted of a sequence of five *distinct* letters picked uniformly at random from the *lower-case* English alphabet, $\{a, b, c, \ldots, z\}$. Two examples of such captcha text are: qfhdk and smgft. An example of such captcha text which has its first two letters in increasing alphabetical order is: vxztr.

What is the probability that the captcha text generated in this way has **the first two letters** in *increasing alphabetical order*, and why?

Solution: $\boxed{\frac{1}{2}}$

One way to see this is as follows: construct a bijection between strings that have their first two letters in alphabetic order, and those which don't. That is: let xyZ be a captcha text, where x, y stand for some one letter each, and Z stands for the remaining three letters. The bijection maps xyZ to yxZ, and exactly one of these two strings has its first two letters in alphabetical order. So the number of captcha texts with the desired property is exactly half the total number of captcha texts.

(A solution which explicitly computes the fraction of favourable sequences over all possible sequences, and evaluates to half, is also valid.)

(b) After a couple of years the bank noticed that some bots had become very good at cracking the captcha. So they updated their software so that it now generates the captcha text by picking five *distinct* letters uniformly at random from the English alphabet, where the case of each letter can be either upper or lower. That is, the letters are now picked from the set $\{a, b, c, \ldots, z, A, B, C, \ldots, Z\}$. Note that the same letter does *not* appear in both lower and upper case in the same captcha text.

Two examples of the captcha text generated by this version are: QFhDk and sMgfT. Examples of such captcha text with the first two letters in increasing alphabetical order are: VxPTr and vXptR. Note that the case of the letter does *not* matter when deciding the alphabetical order.

What is the probability that the captcha text generated in *this* way has the first two letters in increasing alphabetical order, and why?

Solution: $\boxed{\frac{1}{2}}$

One way to see this is as follows: construct a bijection between strings that have their first two letters in alphabetic order, and those which don't. That is: let xyZ be a captcha text, where x, y stand for some one letter each, and Z stands for the remaining three letters. The bijection maps xyZ to yxZ, and exactly one of these two strings has its first two letters in alphabetical order. So the number of captcha texts with the desired property is exactly half the total number of captcha texts.

(A solution which explicitly computes the fraction of favourable sequences over all possible sequences, and evaluates to half, is also valid.)

13. Among 40 bicycle gears, 28 are broken or rusted but not both, 6 are non-defective (i.e. neither rusted nor

broken), and the number of broken gears equals the number of rusted ones. How many gears are rusted?

Solution: $\boxed{20}$

Let U = number of bicycle gears, B = number of broken gears, R = number of rusted gears. We are given that $|U| = 40$, $|B - R| \bigcup |R - B| = 28$, $|B \bigcup R|^c = 6$ and $|B| = |R|$. So $|B \bigcup R| = 40 - 6 = 34$, which implies $|B| + |R| - |B \cap R| = 34$ i.e. $2|R| - |B \cap R| = 34$. Since $|B \cap R| + |B - R| + |R - B| = |B \cup R|$, we must have $|B \cap R| = 34 - 28 = 6$. Thus, $2|R| = 34 + 6 = 40$, so that $|R| = 20$.

(The solution may also be based on a Venn diagram with these numbers.)

14. Consider the following code, in which `A` and `B` are arrays indexed from `0` and `lenA` and `lenB` are the numbers of elements in `A` and `B`, respectively.

```
function foo(A, B, lenA, lenB) {
  acc = 0;

  for i = 0 to (lenA - 1) {
    acc = A[i]^acc;
  }
  print(acc);

  for i = 0 to (lenB - 1) {
    acc = B[i]^acc;
  }
  print(acc);
}
```

Here, `a^b` 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.

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

Let `A = [1, 3, 3, 5, 5]` and `B = [9, 5, 5, 3, 3, 1]`. What are the two values printed by `foo(A,B,5,6)`, in order?

Solution: $\boxed{\text{1 and 9, in this order.}}$

Starting with the value $acc = 0$, the first `for` loop does a cumulative XOR over the values in array `A`. Since the numbers $3, 5$ appear twice in `A`, they cancel out and the resulting value of acc is 1. This 1 gets cancelled out by the XOR with the 1 in array B. Since the the numbers $3, 5$ appear twice in B, they cancel out as well. So the result is 9.

(An explanation that shows how the code runs on this particular input and returns the answer, is also valid.)

15. Consider the following code, in which A and B are arrays indexed from 0 and lenA and lenB are the numbers of elements in A and B, respectively.

```
function foo(A, B, lenA, lenB) {
  sum = lenA + lenB;
  i = 0;
  j = 0;

  for t = 0 to (sum - 1) {
    if (A[i] < B[j]) {
      i = i + 1;
    } else {
      if (A[i] > B[j]) {
        j = j + 1;
      } else {
        return A[i];
      }
    }
  }

  return (-1);
}
```

Let A = [2, 4, 6, 7, 8, 9, 10] and B = [1, 3, 5, 7, 9, 11, 13]. What does foo(A,B,7,7) return?

Solution: $\boxed{7}$

Given (i) two positive integer arrays A and B, each sorted in non-decreasing order, and (ii) the lengths of these two arrays as arguments, this function returns the smallest integer which is present in both A and B, or -1 if there is no common element. Since the smallest common element is 7 for these arrays, the code returns 7.

(An explanation that shows how the code runs on this particular input and returns the answer, is also valid.)

16. Consider the following code, in which A is an integer array of length n indexed from 0, and x is an integer.

```
function foo(A,x,n) {
  found = False;

  while(found != True) {
    i = randInt(0,n);
```

```
        if (A[i] == x) {
            found = True;
        }
    }
    return(i);
}
```

Here, randInt(0,n) returns an integer picked uniformly at random from the range $\{0, 1, \ldots, (n-1)\}$.

All integers present in array A are distinct, and integer x is present in array A. Suppose we make the call foo(A,x,n). What is the expected number of times that the call randInd(0, n) is made from within this call to foo(A,x,n)?

Solution: $\boxed{n}$

The probability of each call to randInd(0, n) succeeding (that is, returning an i such that A[i] == x) is $\frac{1}{n}$. This is because there are exactly n elements in A, of which exactly one element is equal to x. So the expected number of trials before one trial succeeds, is n.

17. A graph consists of a finite set of vertices, and edges between some (unordered) pairs of these vertices. The graphs in this question have no loops (an edge between a vertex and itself) or multiple edges (more than one edge between the same pair of vertices). A vertex v in a graph is said to be a *global* vertex if there is an edge between v and every other vertex in the graph.

    A graph G is constructed as follows: First, its vertex set is assigned to be the set $\{v_1, v_2, \ldots, v_{10}\}$. Next, between each pair of distinct vertices v_i, v_j, an edge is added with probability $\frac{1}{2}$.

    (a) What is the probability that vertex v_1 is a global vertex in graph G?

    Solution: $\boxed{\frac{1}{2^9}}$

    All the 9 possible edges incident with v_1 have to be added by the random process. The probability of adding one of these edges is $\frac{1}{2}$. The description implies that edges are added independently of one another. So the probability of adding all these 9 edges is $\frac{1}{2^9}$.

    (b) What is the expected number of global vertices in graph G?

    Solution: $\boxed{\frac{10}{2^9}}$

    From the previous question, the probability of each vertex being global is $\frac{1}{2^9}$. So the expected number of global vertices among 10 such vertices is $10 \times \frac{1}{2^9}$.

    Here is an alternative explanation, with more details: To each vertex, assign an indicator random variable that takes the value 1 exactly when the vertex is a global vertex. The number of global vertices is then the sum of these indicator random variables. The expected number of global vertices is the expectation of this sum. By linearity of expectation, this in turn is the sum of the expectations of the indicator random variables. The expectation of each indicator random variable is just the probability that it takes the value 1. And this probability is $\frac{1}{2^9}$. So the sum of expectations is $10 \times \frac{1}{2^9}$.

18. The sum of two positive integers a and b is 48 and their least common multiple is 189. Find a and b.

Solution: $\boxed{a = 21 \text{ and } b = 27}$

Let d denote the g.c.d. of a and b. Then there are integers m and n such that $a = dm$, $b = dn$ and g.c.d.$(m, n) = 1$. In that case $dmn = 189$ and $d(m + n) = 48$.

g.c.d.$(m, n) = 1$ implies g.c.d.$(dmn, d(m + n)) = d$, so d must be g.c.d.$(189, 48) = 3$. Then $m + n = 24$ and $mn = 63$. Together with the fact that g.c.d.$(m, n) = 1$, this implies that $m = 7$, $n = 9$. So $a = 21$ and $b = 27$.

19. Recall that for an $n \times n$ matrix A, $\det(A) = \det(A^T)$. Here A^T is the transpose of matrix A. An $n \times n$ matrix A is said to be skew-symmetric if $a_{i,j} = -a_{j,i}$, where $a_{i,j}$ denotes the element in the i^{th} row and j^{th} column, for all $1 \leq i, j \leq n$ and all elements are real.

(a) Is the following statement true or false? Why?

If n is odd, then the determinant of an $n \times n$-skew symmetric matrix is 0.

Solution: $\boxed{\text{This statement is true.}}$

One way to see this is as follows: $\det(-A) = (-1)^n \det(A) = -\det(A)$ since n is odd. Here $A^T = -A$, and so $\det(A^T) = \det(-A) = -\det(A)$. But we have been told that $\det(A^T) = \det(A)$. Thus $\det(A) = -\det(A)$, which implies $\det(A) = 0$.

(b) Give an example of a 2×2 skew-symmetric matrix whose determinant is non-zero.

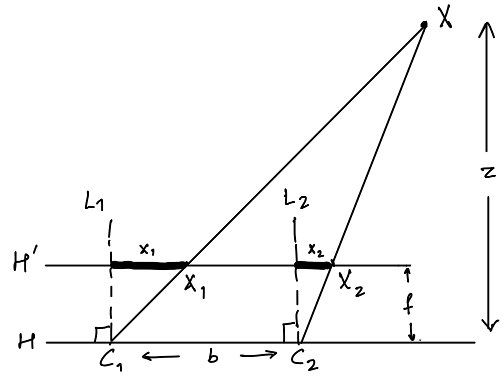
Solution:

One such example:

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

20. An object X is viewed from a camera placed at point C_1. The camera is then moved to the point C_2 and X is again viewed from this new position. Suppose H is the plane containing the points C_1 and C_2. Let H' be a plane parallel to H lying between the object X and the camera positions. Let f denote the distance between H and H'. Let X_1, X_2 be the points of intersection of H' with the line segments joining X with C_1 and C_2 respectively.

Let x_1 be the distance between the reference axis L_1 and X_1. Similarly let x_2 be the distance between the reference axis L_2 and X_2. The axes L_1 and L_2 are both perpendicular to the camera plane H. The difference $|x_1 - x_2|$ is called the *disparity* in views. Find a formula for the *depth* z i.e. distance of the object from the camera plane in terms of f, the distance b between the cameras, and the disparity.



Solution: $\boxed{z = \frac{b \times f}{|x_1 - x_2|}}$

Let x denote the perpendicular distance between L_1 and X. Then

$$\frac{x_1}{f} = \frac{x}{z} \quad \text{and} \quad \frac{x_2}{f} = \frac{x - b}{z}.$$

$$\text{So} \quad \frac{x_2}{f} = \frac{x - b}{z} = \frac{x}{z} - \frac{b}{z} = \frac{x_1}{f} - \frac{b}{z}$$

$$\text{implying that} \quad \frac{|x_1 - x_2|}{f} = \frac{b}{z}$$

so we get $z = \dfrac{b \times f}{|x_1 - x_2|}$.