GATE BM 2023 Quetion 9

Snehil Singh EE22BTECH11050*

Question :
Out of 1000 individuals in a town, 100 unidentified individuals are covid positive. Due to lack of adequate covid-testing kits, the health authorities of the town devised a strategy to identify these covid-positive individuals. The strategy is to:

1) Collect saliva samples from all 1000 individuals and randomly group them into sets of 5.
2) Mix the samples within each set and test the mixed sample for covid.
3) If the test done in (ii) gives a negative result, then declare all the 5 individuals to be covid negative.
4) If the test done in (ii) gives a positive result, then all the 5 individuals are separately tested for covid.

Given this strategy, no more than testing kits will be required to identify all the 100 covid positive individuals irrespective of how they are grouped

1) 700
2) 600
3) 800
4) 1000

Solution :
Given :

Parameter	Value
Number of individuals	1000
Strength of each group	5
Number of groups	200
Number of Covid positive individuals	100

At first we ll have to examine all the groups and then test all 5 individuals of each group who tested positive.

Using binomial expansion to find the probability of a group testing positive

Parameter	Value	Expression
N	1000	Total number of individuals
n	5	Number of individuals in group
p	0.1	Probability of individual testing positive
q	0.9	Probability of individual testing negative

Probability of a group of 5 with i covid negative individuals is given as
the PMF:

$$Pr(X = i) = ^n C_i (q)^i (p)^{n-i} \tag{1}$$

$$= ^5 C_i (0.9)^i (0.1)^{5-i} \tag{2}$$

Let us assume X is a random variable which denotes the number of individuals tested negative in a group.

$$Pr = \begin{cases} {}^5C_0(0.9)^0(0.1)^5 \ \{X = 0\} \\ {}^5C_1(0.9)^1(0.1)^4 \ \{X = 1\} \\ {}^5C_2(0.9)^2(0.1)^3 \ \{X = 2\} \\ {}^5C_3(0.9)^3(0.1)^2 \ \{X = 3\} \\ {}^5C_4(0.9)^4(0.1)^1 \ \{X = 4\} \\ {}^5C_5(0.9)^5(0.1)^0 \ \{X = 5\} \end{cases} \tag{3}$$

Random Variable	**Value**	**Expression**	**Probability**
X	0	${}^5C_0(0.9)^0(0.1)^5$	0.00001
	1	${}^5C_1(0.9)^1(0.1)^4$	0.00045
	2	${}^5C_2(0.9)^2(0.1)^3$	0.0081
	3	${}^5C_3(0.9)^3(0.1)^2$	0.0729
	4	${}^5C_4(0.9)^4(0.1)^1$	0.32805
	5	${}^5C_5(0.9)^5(0.1)^0$	0.59049

Probability of group having atmost k negative tested individuals
the CDF :

$$F_X(k) = Pr(X \leq k) \tag{4}$$
$$= \Sigma_{i=0}^{k} {}^nC_i(q)^i(p)^{n-i} \tag{5}$$
$$= \Sigma_{i=0}^{k} {}^5C_i(0.9)^i(0.1)^{5-i} \tag{6}$$
$$\tag{7}$$

If a group is tested positive, atleast one of the 5 individuals should be covid positive which means atmost 4 individuals are covid negative.
Finding probability of a group testing positive

$$Pr = F_X(4) \tag{8}$$
$$= \Sigma_{i=0}^{4} {}^5C_i(0.9)^i(0.1)^{5-i} \tag{9}$$
$$= {}^5C_5(0.1)^5 + {}^5C_4(0.1)^4(0.9) + {}^5C_3(0.1)^3(0.9)^2 + {}^5C_2(0.1)^2(0.9)^3 + {}^5C_1(0.1)^1(0.9)^4 \tag{10}$$
$$= (1)\,(0.00001) + 5\,(0.9)\,(0.0001) + 10\,(0.81)\,(0.001) + 10\,(0.729)\,(0.01) + 5\,(0.6561)\,(0.1) \tag{11}$$
$$= 0.00001 + 0.00045 + 0.0081 + 0.0729 + 0.32805 \tag{12}$$
$$= 0.4096 \tag{13}$$

Parameter	Value
Number of tests in first round of testing	A
Number of groups tested positive in first round	b
Number of tests in second round of testing	B
Total number of tests in both rounds	T

In the first round we will test all the 200 groups.
Number of tests in first round(A) :

$$A = 200$$

Out of these 200 tests,assuming the estimated number of groups tested positive is b
If a group is tested positive, atleast one of the 5 is covid positive, which impies the group can only have atmost 4 covid negative individual.
Calculating b :

$$\begin{aligned}
\implies Pr(\text{of group testing positive}) &= \frac{no.of groups tested positive}{total no of groups tested} && (14)\\
\implies F_X(4) &= \frac{b}{A} && (15)\\
\implies 0.4096 &= \frac{b}{200} && (16)\\
\implies b &= (0.4096)\,(200) && (17)\\
\implies b &= 81.92 && (18)
\end{aligned}$$

In all the groups tested positive, atleast one individual is covid positive, to find all the covid positive individuals, we will test all the 5 individuals of the positive groups seperately in round 2.
So estimated number tests in round 2 (B) :

$$\begin{aligned}
B &= 5b && (19)\\
&= (5)\,(81.92) && (20)\\
&= 409.6 && (21)
\end{aligned}$$

Total number of tests estimated (T) :

$$\begin{aligned}
T &= A + B && (22)\\
&= 200 + 409.6 && (23)\\
&= 609.6 && (24)
\end{aligned}$$

∴ So 609.6 is the expected number of tests required to detect all the covid positive individuals.
According to the options, no more than 700 tests will be required.
So, the correct answer is option A 700.

Simulation :

1) Including standard C libraries - `stdio.h` (for input/output), `stdlib.h` (for dynamic memory allocation), `time.h` (for time-related functions), and `math.h` (for mathematical functions).

2) Defining constants for the total number of individuals (`TOTAL_INDIVIDUALS`), the number of initially positive individuals (`POSITIVE_INDIVIDUALS`), and the group size used in testing (`GROUP_SIZE`).

3) Declared a function named `covidTestingStrategy` that simulates a COVID-19 testing strategy using group testing. The function takes an array of individuals, the total number of individuals, and the number of initially positive individuals as parameters.

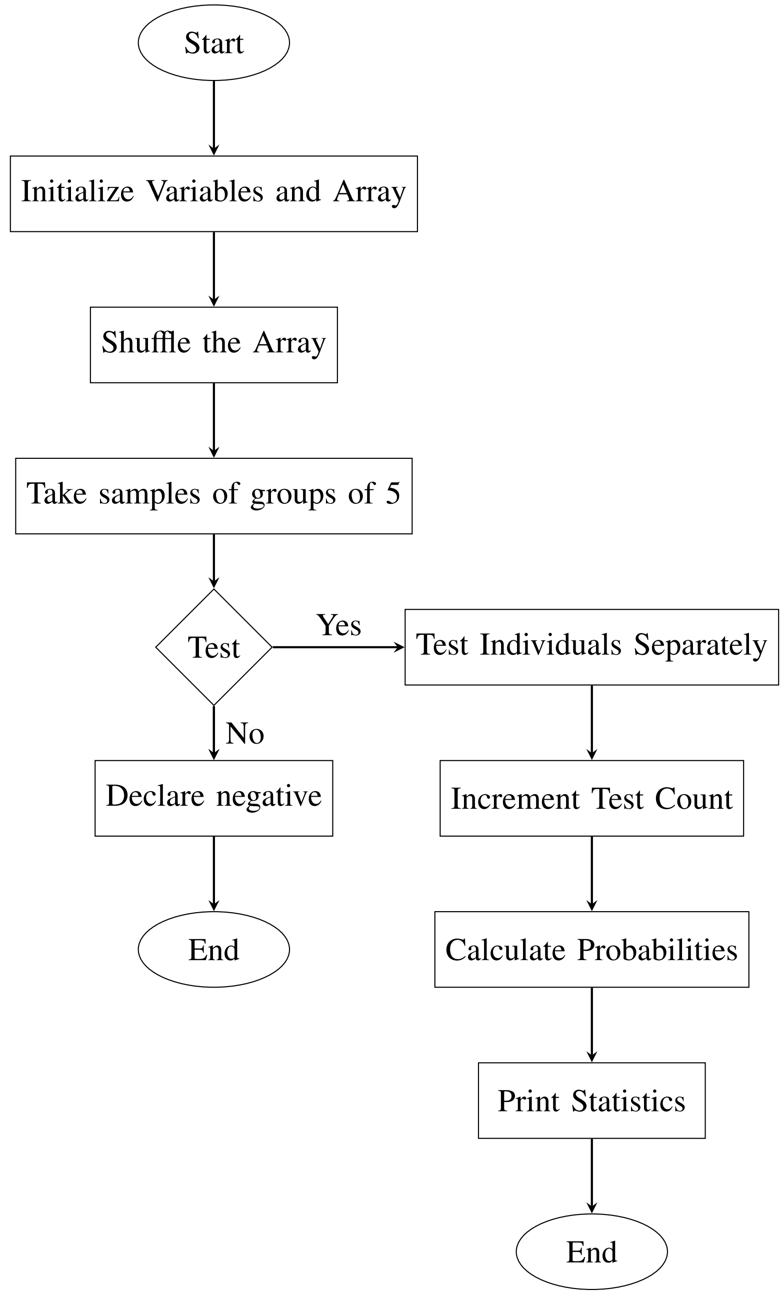
4) Initialized variables `testCount` and `groupTestTrueCount` to keep track of the total number of tests and the number of group tests that turned out to be true and shuffled the array of individuals randomly using the Fisher-Yates shuffle algorithm to have randomness into the order of testing.

5) We mixed the samples within each group using Bitwise "OR" and tested them.
   a) If the mixed sample is negative (`groupResult == 0`), declares all 5 individuals in the group as COVID negative with one test.
   b) If the mixed sample is positive, tests each individual separately, and increments the test count accordingly.
6) Calculated the probability of the group test being true based on the initial probability of an individual being positive and the group size and Calculated the expected number of tests for one set, total expected number of sets, and total expected number of tests based on the calculated probabilities.

7) Printed various statistics related to the testing strategy.

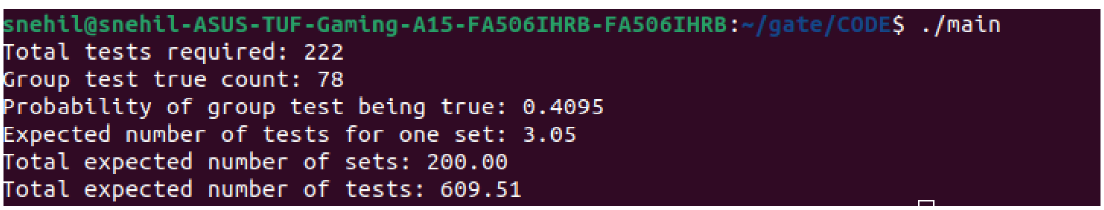
8) Seed the random number generator with the current time. Initialized an array representing individuals, marking the first `POSITIVE_INDIVIDUALS` as positive. Calls the `covidTestingStrategy` function with the array and other parameters.

Flowchart to explain the function `covidTestingStrategy` is given below :


Start
Initialize Variables and Array
Shuffle the Array
Take samples of groups of 5
Test
Yes
Test Individuals Separately
No
Declare negative
Increment Test Count
End
Calculate Probabilities
Print Statistics
End

```c
// Custom power function for integer exponentiation
double customPow(double base, int exponent) {
    double result = 1.0;

    for (int i = 0; i < exponent; i++) {
        result *= base;
    }

    return result;
}

// Function to simulate the testing strategy
void covidTestingStrategy(int individuals[], int total, int positive) {
    int testCount = 0;
    int groupTestTrueCount = 0;

    // Shuffle the individuals randomly
    for (int i = total - 1; i > 0; --i) {
        int j = rand() % (i + 1);
        // Swap individuals[i] and individuals[j]
        int temp = individuals[i];
        individuals[i] = individuals[j];
        individuals[j] = temp;
    }

    // Perform testing in groups
    for (int i = 0; i < total; i += GROUP_SIZE) {
        int groupResult = 0;

        // Mix the samples within each set
        for (int j = i; j < i + GROUP_SIZE; ++j) {
            groupResult |= individuals[j];
        }

        // Test the mixed sample for covid
        if (groupResult == 0) {
            // Declare all 5 individuals to be covid negative
            testCount++;
        } else {
            // Test each individual separately if the mixed sample is positive
            for (int j = i; j < i + GROUP_SIZE; ++j) {
                if (individuals[j] == 1) {
                    // Individual is covid positive
                    testCount++;
                }
            }
            groupTestTrueCount++;
        }
    }

    // Probability of the group test being true
    double p = (double)positive / total;
    double groupTestTrueProbability = 1.0 - customPow(1.0 - p, GROUP_SIZE);

    // Expected number of tests for one set
    double expectedTestsOneSet = 1 + GROUP_SIZE * groupTestTrueProbability;

    // Total expected number of sets
    double totalExpectedSets = (double)total / GROUP_SIZE;

    // Total expected number of tests
    double totalExpectedTests = totalExpectedSets * expectedTestsOneSet;
```

Fig. 1: Input



```
snehil@snehil-ASUS-TUF-Gaming-A15-FA506IHRB-FA506IHRB:~/gate/CODE$ ./main
Total tests required: 222
Group test true count: 78
Probability of group test being true: 0.4095
Expected number of tests for one set: 3.05
Total expected number of sets: 200.00
Total expected number of tests: 609.51
```

Fig. 2: Output