Cancer rate

Even if 100% of patients with pancreatic cancer have a certain symptom, when someone has the same symptom, it does not mean that this person has a 100% chance of getting pancreatic cancer. Assume the incidence rate of pancreatic cancer is 1/100000, while 1/10000 healthy individuals have the same symptoms worldwide, the probability of having pancreatic cancer given the symptoms is only 9.1%, and the other 90.9% could be "false positives" (that is, falsely said to have cancer; "positive" is a confusing term when, as here, the test gives bad news).

Based on incidence rate, the following table presents the corresponding numbers per 100,000 people.

Symptom \ Cancer	Yes	No	Total
Yes	1	10	11
No	0	99989	99989
Total	1	99999	100000

Which can then be used to calculate the probability of having cancer when you have the symptoms:

$$\begin{aligned} P(\text{Cancer}|\text{Symptoms}) &= \frac{P(\text{Symptoms}|\text{Cancer})P(\text{Cancer})}{P(\text{Symptoms})} \\ &= \frac{P(\text{Symptoms}|\text{Cancer})P(\text{Cancer})}{P(\text{Symptoms}|\text{Cancer})P(\text{Cancer}) + P(\text{Symptoms}|\text{Non-Cancer})P(\text{Non-Cancer})} \\ &= \frac{1 \times 0.00001}{1 \times 0.00001 + (10/99999) \times 0.99999} = \frac{1}{11} \approx 9.1\% \end{aligned}$$

1000

$P(C) = \frac{1}{1000} = 0.001$ 1   999 $P(\bar{C}) = 0.999$

$P(+|C) = 1$ $P(-|C) = 0$

$P(+|\bar{C}) = 5\%$

$= \frac{5}{100} \times 999$

$\approx \frac{5}{100} \times 1000$

≈ 50

$999 - 50 = 949$

out of $1+50 = 51$ persons who are tested positive, only $\underline{1}$ has condition.

$\therefore P(C|+) = \frac{1}{51} \approx \frac{1}{50} \approx 2\%$

$$P(C|+) = \frac{P(+|C)\,P(C)}{P(+)} = \frac{P(+|C)\,P(C)}{P(+|C)\,P(C) + P(+|\bar{C})\,P(\bar{C})}$$

$$= \frac{1 \times 0.001}{1 \times 0.001 + 0.05 \times 0.999} = \boxed{1.96\%}$$

Defective item rate

A factory produces an item using three machines—A, B, and C—which account for 20%, 30%, and 50% of its output, respectively. Of the items produced by machine A, 5% are defective; similarly, 3% of machine B's items and 1% of machine C's are defective. If a randomly selected item is defective, what is the probability it was produced by machine C?

Condition / Machine	Defective	Flawless	Total
A	10	190	200
B	9	291	300
C	**5**	495	500
Total	**24**	976	1000

Once again, the answer can be reached without using the formula by applying the conditions to a hypothetical number of cases. For example, if the factory produces 1,000 items, 200 will be produced by Machine A, 300 by Machine B, and 500 by Machine C. Machine A will produce 5% × 200 = 10 defective items, Machine B 3% × 300 = 9, and Machine C 1% × 500 = 5, for a total of 24. Thus, the likelihood that a randomly selected defective item was produced by machine C is 5/24 (~20.83%).

This problem can also be solved using Bayes' theorem: Let X_i denote the event that a randomly chosen item was made by the i th machine (for i = A,B,C). Let Y denote the event that a randomly chosen item is defective. Then, we are given the following information:

$$P(X_A) = 0.2, \quad P(X_B) = 0.3, \quad P(X_C) = 0.5.$$

If the item was made by the first machine, then the probability that it is defective is 0.05; that is, $P(Y \mid X_A) = 0.05$. Overall, we have

$$P(Y|X_A) = 0.05, \quad P(Y|X_B) = 0.03, \quad P(Y|X_C) = 0.01.$$

To answer the original question, we first find P(Y). That can be done in the following way:

$$P(Y) = \sum_i P(Y|X_i)P(X_i) = (0.05)(0.2) + (0.03)(0.3) + (0.01)(0.5) = 0.024.$$

Hence, 2.4% of the total output is defective.

We are given that Y has occurred, and we want to calculate the conditional probability of X_C. By Bayes' theorem,

$$P(X_C|Y) = \frac{P(Y|X_C)P(X_C)}{P(Y)} = \frac{0.01 \cdot 0.50}{0.024} = \frac{5}{24}$$

Given that the item is defective, the probability that it was made by machine C is 5/24. Although machine C produces half of the total output, it produces a much smaller fraction of the defective items. Hence the knowledge that the item selected was defective enables us to replace the prior probability $P(X_C) = 1/2$ by the smaller posterior probability $P(X_C \mid Y) = 5/24$.

22) Suppose you were interviewed for a technical role. 50% of the people who sat for the first interview received the call for second interview. 95% of the people who got a call for second interview felt good about their first interview. 75% of people who did not receive a second call, also felt good about their first interview. If you felt good after your first interview, what is the probability that you will receive a second interview call?

A) 66%
B) 56%

C) 75%

D) 85%
Solution: **(B)**

Let's assume there are 100 people that gave the first round of interview. The 50 people got the interview call for the second round. Out of this 95 % felt good about their interview, which is 47.5. 50 people did not get a call for the interview; out of which 75% felt good about, which is 37.5. Thus, the total number of people that felt good after giving their interview is (37.5 + 47.5) 85. Thus, out of 85 people who felt good, only 47.5 got the call for next round. Hence, the probability of success is (47.5/85) = 0.558.

Another more accepted way to solve this problem is the Baye's theorem. I leave it to you to check for yourself.

5. Example of Bayes Theorem and Probability trees

Let's take the example of the breast cancer patients. The patients were tested thrice before the oncologist concluded that they had cancer. The general belief is that 1.48 out of a 1000 people have breast cancer in the US at that particular time when this test was conducted. The patients were tested over multiple tests. Three sets of test were done and the patient was only diagnosed with cancer if she tested positive in all three of them.

Let's examine the test in detail.

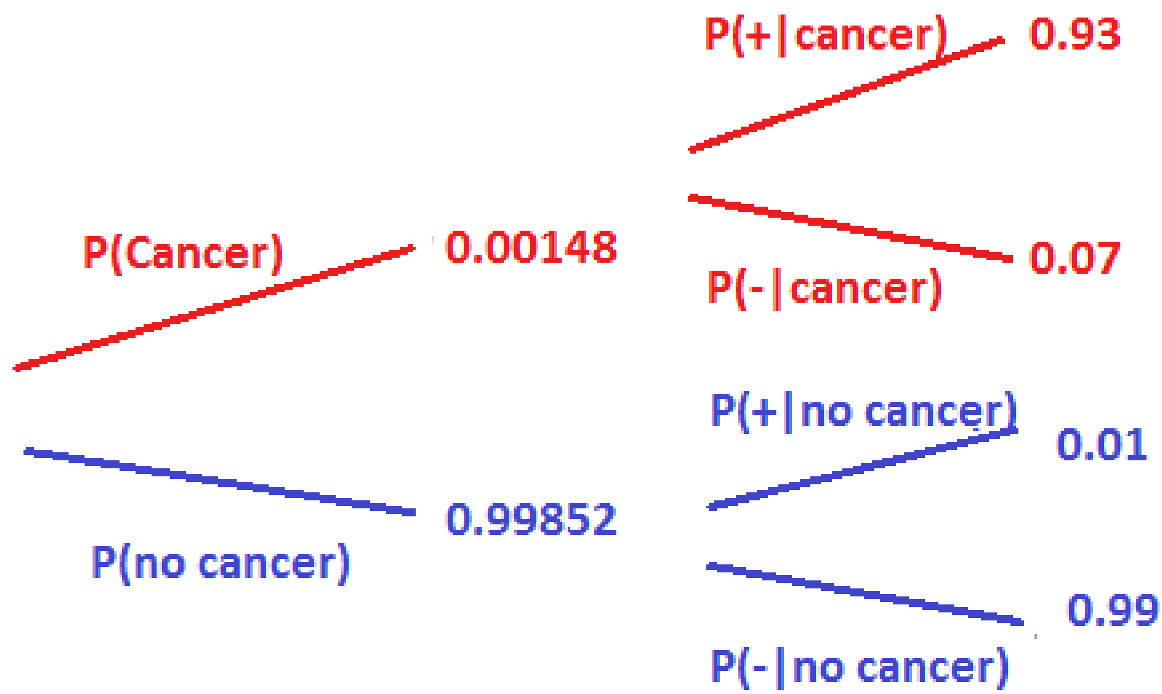
> Sensitivity of the test (93%) – true positive Rate
>
> Specificity of the test (99%) – true negative Rate

Let's first compute the probability of having cancer given that the patient tested positive in the first test.

> P (has cancer | first test +)
>
> P (cancer) = 0.00148
>
> Sensitivity can be denoted as P (+ | cancer) = 0.93
>
> Specificity can be denoted as P (- | no cancer)

Since we do not have any other information, we believe that the patient is a randomly sampled individual. Hence our prior belief is that there is a 0.148% probability of the patient having cancer.

The complement is that there is a 100 – 0.148% chance that the patient does not have CANCER. Similarly we can draw the below tree to denote the probabilities.



Let's now try to calculate the probability of having cancer given that he tested positive on the first test i.e. P (cancer|+)

$$P\ (\text{cancer}\,|\,+) = \frac{P(cancer\ and +)}{P(+)}$$

P (cancer and +) = P (cancer) * P (+) = 0.00148*0.93

P (no cancer and +) = P (no cancer) * P(+) = 0.99852*0.01

To calculate the probability of testing positive, the person can have cancer and test positive or he may not have cancer and still test positive.

$$P\ (\text{CANCER}\,|\,+) = \frac{P(cancer\ and +)}{P(cancer\ and +) + P(no\ cancer\ and +)} = 0.12$$

This means that there is a 12% chance that the patient has cancer given he tested positive in the first test. This is known as the **posterior probability.**

Drug testing [edit]

Suppose, a particular test for whether someone has been using cannabis is 90% sensitive, meaning the true positive rate (TPR)=0.90. Therefore it leads to 90% true positive results (correct identification of drug use) for cannabis users.

The test is also 80% specific, meaning true negative rate (TNR)=0.80. Therefore the test correctly identifies 80% of non-use for non-users, but also generates 20% false positives, or false positive rate (FPR)=0.20, for non-users.

Assuming 0.05 prevalence, meaning 5% of people use cannabis, what is the probability that a random person who tests positive is really a cannabis user?

The Positive predictive value (PPV) of a test is the proportion of persons who are actually positive out of all those testing positive, and can be calculated from a sample as:

PPV = True positive / Tested positive

If sensitivity, specificity, and prevalence are known, PPV can be calculated using Bayes theorem. Let $P(\text{User} \mid \text{Positive})$ mean "the probability that someone is a cannabis user given that they test positive," which is what is meant by PPV. We can write:

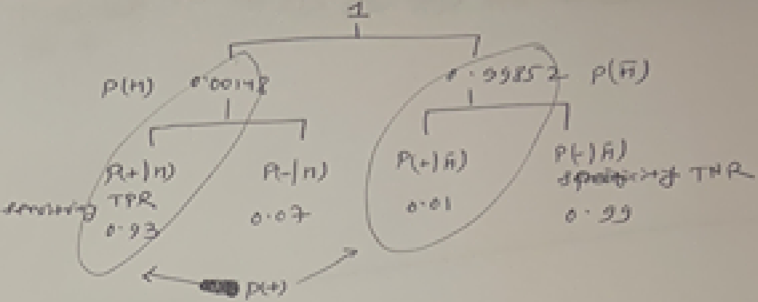
$$\begin{aligned} P(\text{User} \mid \text{Positive}) &= \frac{P(\text{Positive} \mid \text{User})P(\text{User})}{P(\text{Positive})} \\ &= \frac{P(\text{Positive} \mid \text{User})P(\text{User})}{P(\text{Positive} \mid \text{User})P(\text{User}) + P(\text{Positive} \mid \text{Non-user})P(\text{Non-user})} \\ &= \frac{0.90 \times 0.05}{0.90 \times 0.05 + 0.20 \times 0.95} = \frac{0.045}{0.045 + 0.19} \approx 19\% \end{aligned}$$

HIV is still a very scary disease to even get tested for. The US military tests its recruits for HIV when they are recruited. They are tested on three rounds of Elisa(an HIV test) before they are termed to be positive.

The prior probability of anyone having HIV is 0.00148. The true positive rate for Elisa is 93% and the true negative rate is 99%.

29) What is the probability that a recruit has HIV, given he tested positive on first Elisa test? The prior probability of anyone having HIV is 0.00148. The true positive rate for Elisa is 93% and the true negative rate is 99%.

A) 12%
B) 80%
C) 42%
D) 14%
Solution: **(A)**



$$P(H|+) = \frac{P(H)\cdot P(+|H)}{P(+)} = \frac{0.00148 \times 0.93}{0.00148 \times 0.93 + 0.99852 \times 0.01}$$

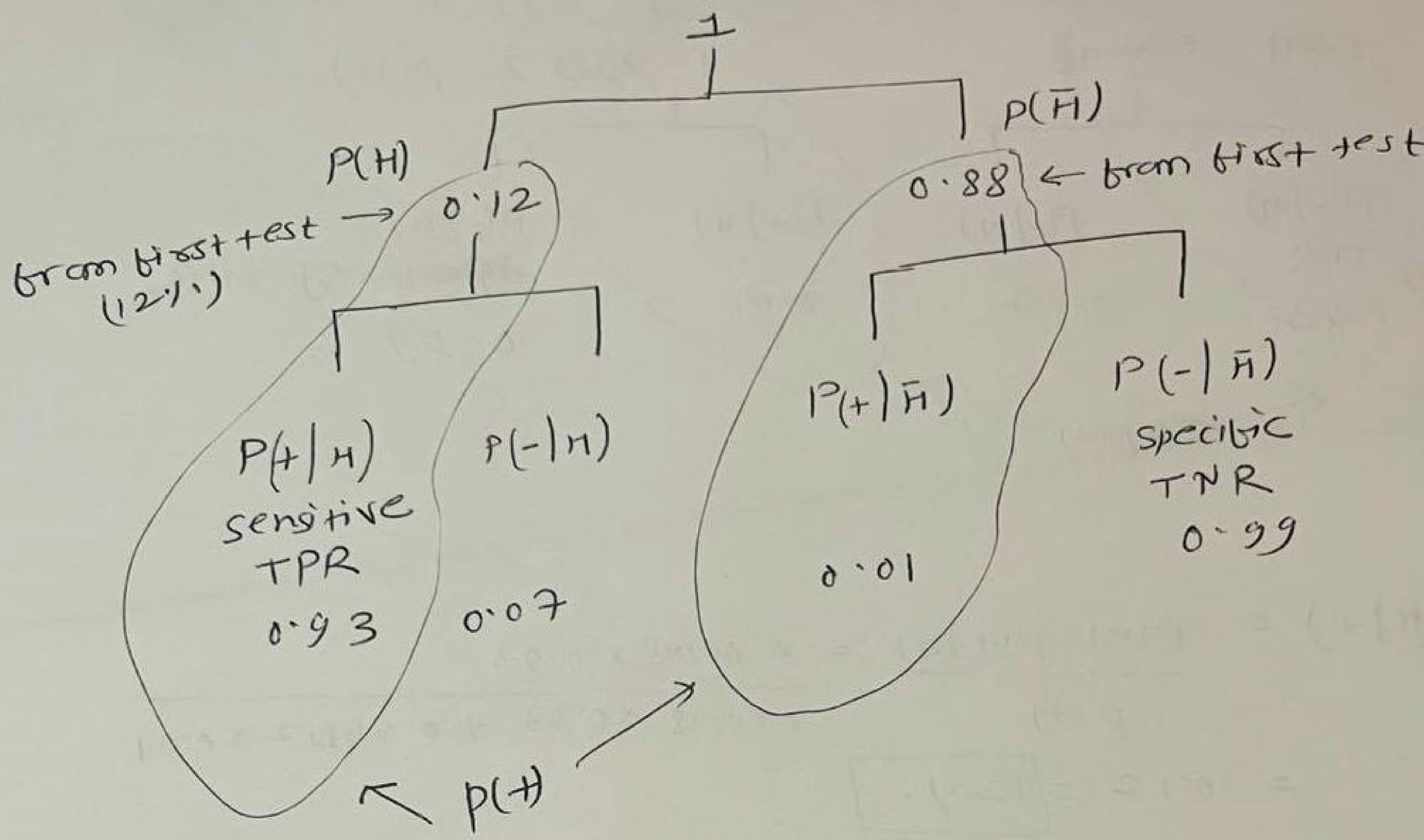
$$= 0.12 = \boxed{12\%}$$

30) What is the probability of having HIV, given he tested positive on Elisa the second time as well.

The prior probability of anyone having HIV is 0.00148. The true positive rate for Elisa is 93% and the true negative rate is 99%.

A) 20%
B) 42%
C) 93%
D) 88%
Solution: **(C)**

Bayes update



$$P(H|+) = \frac{P(H)\cdot P(+|H)}{P(+)} = \frac{0.12 \times 0.93}{0.12 \times 0.93 + 0.88 \times 0.01}$$

$$= 0.93 = \boxed{93\%}$$

37) About 30% of human twins are identical, and the rest are fraternal. Identical twins are necessarily the same sex, half are males and the other half are females. One-quarter of fraternal twins are both males, one-quarter both female, and one-half are mixed: one male, one female. You have just become a parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?

A) 50%

B) 72%
C) 46%
D) 33%

Solution: **(C)**

This is a classic problem of Bayes theorem.

P(I) denoted Probability of being identical and P(~I) denotes Probability of not being identical

P(Identical) = 0.3

P(not Identical)= 0.7

P(FF|I)= 0.5

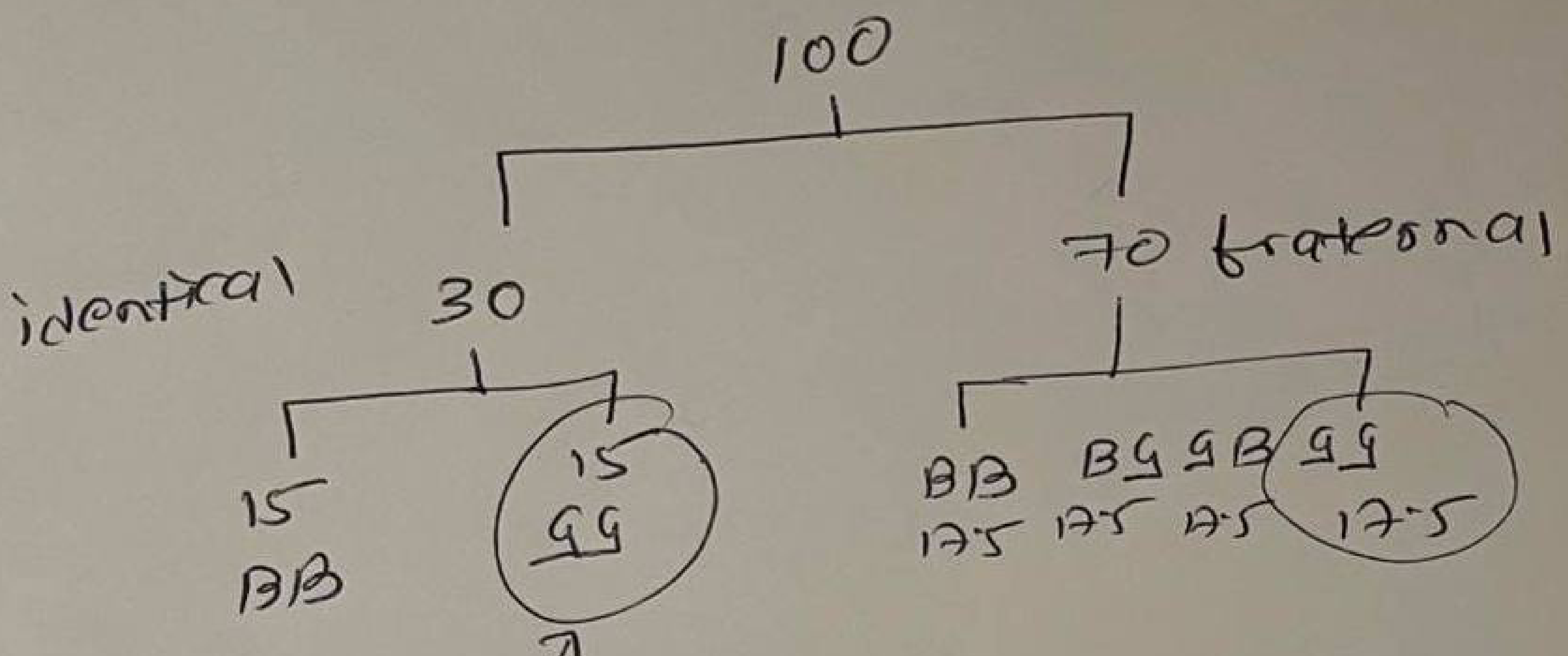
P(MM|I)= 0.5

P(MM|~I)= 0.25

P(FF|~I)= 0.25

P(FM|~I)= 0.25

P(I|FF) = 0.46



$$\text{Answer} = \frac{15}{15+17.5} = 0.4615 = 46.15\%$$

> *You are given a bag of 100 coins, with 99 fair ones flipping heads and tails with* 0.5 *probability each, and one unfair coin which flips heads with* 1.0 *probability. You pick a coin randomly and flip it 10 times, getting heads every single time. What is the probability that you picked the unfair coin?*

The question can be answered quite easily using Bayes' Theorem from probability theory. The extended formula from the theorem is:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}.$$

Here we let A be the even of having picked the unfair coin, and B be the event of having flipped 10 heads in a row. So we have $P(A) = 0.01$, $P(B|A) = 1.0$, $P(B|\neg A) = (\frac{1}{2})^{10}$, and $P(\neg A) = 0.99$. Putting it all together we get

$$P(A|B) = \frac{1.0 \times 0.01}{1.0 \times 0.01 + (\frac{1}{2})^{10} \times 0.99}$$

which is approximately 0.9118 . So there is a **91.18%** chance we picked the unfair coin.

$P(F) = 0.99 \quad P(U) = 0.01$

$P(H|F) = 0.5 \quad P(H|U) = 1.0$

$P(10H|F) = (\frac{1}{2})^{10} \quad P(10H|U) = 1.0$

$$P(U|10H) = ? = \frac{P(10H|U) \cdot P(U)}{P(10H)} = \frac{P(10H|U) \cdot P(U)}{P(10H|U) \cdot P(U) + P(10H|F) \cdot P(F)}$$

$$= \frac{1 \times 0.01}{1 \times 0.01 + (\frac{1}{2})^{10} \times 0.99} = 0.9118 = 91.18\%$$

1

I encountered this problem:

> We have three coins, two fair coins and one coin with heads on each face. One coin is picked randomly among those three coins and is flipped two times. We see the sequence: H, H. What is the probability of obtaining heads if we flip this coin one more time?

To tackle it, here's my reasoning:

1. Find the probabilities of the coin being fair and unfair conditional on the sequence: H, H.
2. Find the probability of getting a heads in the additional flip conditional on the sequence: H, H.

1) Using Bayes formula we have:

$$\begin{aligned} P(fair|HH) &= \frac{P(HH|fair)P(fair)}{P(HH|fair)P(fair) + P(HH|unfair)P(unfair)} \\ &= \frac{\frac{1}{4}\frac{2}{3}}{\frac{1}{4}\frac{2}{3} + 1 \cdot \frac{1}{3}} \\ &= \frac{1/6}{1/6 + 1/3} \\ &= \frac{1}{3} \end{aligned}$$

We get at the same time:

$$P(unfair|HH) = 1 - 1/3 = 2/3$$

2) Let's denote A the event of obtaining of heads at the additional flip. We have:

$$P(A) = P(A|fair)P(fair) + P(A|unfair)P(unfair)$$

and conditioning on the sequence H, H we get:

$$\begin{aligned} P(A|HH) &= P(A|fair, HH)P(fair|HH) + P(A|unfair, HH)P(unfair|HH) \\ &= P(A|fair)P(fair|HH) + P(A|unfair)P(unfair|HH) \end{aligned}$$

We then get:

$$P(A|HH) = \frac{1}{2}\frac{1}{3} + 1 \cdot \frac{2}{3}$$
$$P(A|HH) = \frac{5}{6}$$

2

Your answer is correct. A shorter way of writing it: Let H_n denote the event that we get H at the n-th coin toss. Hence, we want the probability

$$P(H_3|(H_1 \cap H_2)) = \frac{P(H_1 \cap H_2 \cap H_3)}{P(H_1 \cap H_2)}$$

$$= \frac{P(H_1 \cap H_2 \cap H_3|fair) \cdot P(fair) + P(H_1 \cap H_2 \cap H_3|unfair) \cdot P(unfair)}{P(H_1 \cap H_2|fair) \cdot P(fair) + P(H_1 \cap H_2|unfair) \cdot P(unfair)}$$

$$= \frac{(\frac{1}{2})^3 \cdot \frac{2}{3} + 1^3 \cdot \frac{1}{3}}{(\frac{1}{2})^2 \cdot \frac{2}{3} + 1^2 \cdot \frac{1}{3}} = \frac{5}{6}.$$

39) Jack is having two coins in his hand. Out of the two coins, one is a real coin and the second one is a faulty one with Tails on both sides. He blindfolds himself to choose a random coin and tosses it in the air. The coin falls down with Tails facing upwards. What is the probability that this tail is shown by the faulty coin?

A) 1/3
B) 2/3
C) 1/2
D) 1/4
Solution: **(B)**

We need to find the probability of the coin being faulty given that it showed tails.

P(Faulty) = 0.5

P(getting tails) = 3/4

P(faulty and tails) =0.5*1 = 0.5

Therefore the probability of coin being faulty given that it showed tails would be 2/3

$P(R) = 0.5 \quad P(F) = 0.5$

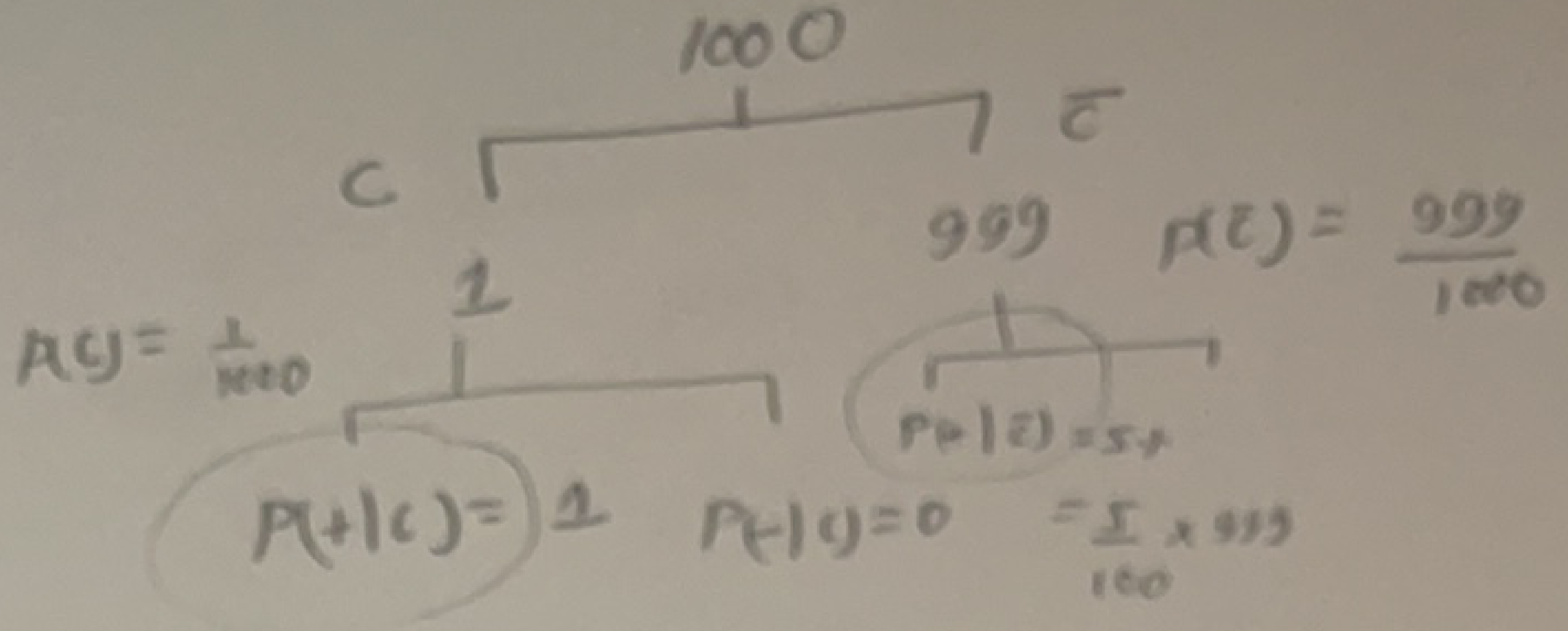
$P(T|R) = 0.5 \quad P(T|F) = 1$

$$P(F|T) = ? = \frac{P(T|F) \cdot P(F)}{P(T)} = \frac{P(T|F) \cdot P(F)}{P(T|F) \cdot P(F) + P(T|R) \cdot P(R)}$$

$$= \frac{1 \times 0.5}{1 \times 0.5 + 0.5 \times 0.5} = \frac{0.5}{0.5 + 0.25} = \frac{0.25 \times 2}{0.25 \times 3} = \boxed{\frac{2}{3}}$$

6 | A test has a true positive rate of 100% and a false positive rate of 5%. There is a population with a one-in-thousand rate of having the condition the test tests for. Considering only that you have a positive test, what is the probability of having that condition? (Topic: Classification Rates)

Let's suppose you are being tested for a disease — if you have the illness the test will end up saying you have the illness. However, if you don't have the illness, 5% of the time the test will end up saying you have the illness and 95% of the time the test will determine that you do not have the illness. Therefore, there is a 5% error in the case that you do not have the illness. Out of 1000 people, 1 person who has the disease will get true positive result. Out of the remaining 999 people, 5% will also get a (false) positive result. Close to 50 people will get a positive result for the disease. This means that out of 1000 people, 51 people will be tested positive for the disease even though only one person has the illness. There is only a 2% probability of you having the disease even if the test is positive.



$$P(C|+) = \frac{P(+|C) \cdot P(C)}{P(+|C) \cdot P(C) + P(+|\bar{C}) \cdot P(\bar{C})} = \frac{1 \times 0.001}{1 \times 0.001 + 0.05 \times 0.999} = 1.96\%$$

Q38. You have two coins, one of which is fair and comes up heads with a probability 1/2, and the other which is biased and comes up heads with probability 3/4.
You randomly pick coin and flip it twice, and get heads both times.
What is the probability that you picked the fair coin?

P(F) = 1/2, P(U) = 1/2, P(H|F) = 1/2, P(H|U) = 3/4, P(2H|F) = (1/2)^2, P(2H|U) = (3/4)^2
P(2H) = P(2H|F) * P(F) + P(2H|U) * P(U)
P(F|2H) = P(2H|F) * P(F) / P(2H) = 4/13

19) A test has a true positive rate of 100% and false positive rate of 5%. There is a population with a 1/1000 rate of having the condition the test identifies. Considering a positive test, what is the probability of having that condition?

Let's suppose you are being tested for a disease, if you have the illness the test will end up saying you have the illness. However, if you don't have the illness- 5% of the times the test will end up saying you have the illness and 95% of the times the test will give accurate result that you don't have the illness. Thus there is a 5% error in case you do not have the illness.

Out of 1000 people, 1 person who has the disease will get true positive result.

Out of the remaining 999 people, 5% will also get true positive result.

Close to 50 people will get a true positive result for the disease.

This means that out of 1000 people, 51 people will be tested positive for the disease even though only one person has the illness. There is only a 2% probability of you having the disease even if your reports say that you have the disease.