Probability and Statistics
(Conditional Probability_Bayes Theorem)

Prepared By: Dr.Mydhili K Nair, Professor, ISE Dept, RIT
For: Machine Learning Elective Class
Target Audience: Sem 6 Students
Term: Feb to June 2019

Sl#	Probability / Statistical Concept	Sl#	Probability / Statistical Concept
1	Sample Space and Events, Probability Space and Random Variable	8	Law of Large Numbers,Estimation Bias
2	Probability Density Function (PDF) / Cumulative Distribution Function(CDF), Conditional Probability	9	Bounded Moments,Density Estimation,Sampling, Sampling Bias - Importance Sampling, Stratified Sampling
3	Independent and Identically Distributed(IID) and Conditional Independence	10	Markov Chains and Monte Carlo Methods, Metropolis-Hasting, Correlation (and how it doesn't imply causation)
4	Bayes Rule,Likelihood,Maximum Likelihood Estimate	11	The Normal (Gaussian) distribution and standard deviations
5	Prior, Evidence, Posterior,Maximum A Posteriori Estimate(MAP)	12	Bias and Variance, Hypothesis tests such as Student's t-tests and z-tests
6	Generative Models, Discriminative Models, Parametric and Non-Parametric Models	13	The difference between frequentist and Bayesian statistics
7	Population Expectation, Sample Expectation/Variance/Covariance	14	Cross-validation (holdout and k-fold)

Dilemma at the movies

This person dropped their ticket in the hallway.

Do you call out

> "Excuse me, ma'am!"

or

> "Excuse me, sir!"

You have to make a guess.

Dilemma at the movies

What if they're standing in line for the men's restroom?

Bayesian inference is a way to capture common sense.

It helps you use what you know to make better guesses.

Case Study #2A: Conditional Probability and Data Dependency

Researchers surveyed recent graduates of two different universities about their annual incomes. The following two-way table displays data for the 300 graduates who responded to the survey.

Annual income	University A	University B	TOTAL
Under $20,000	36	24	60
$20,000 to 39,999	109	56	165
$40,000 and over	35	40	75
TOTAL	180	120	300

P(>$40K | Uni B)=40/300

P(Uni B) 120/300

= ⅓ = 0.33 = **33%**

P(>$40K) = 75/300

= 0.25 = **25%**

Dependant - Far Apart Values

Suppose we choose a random graduate from this data.

Are the events "income is $40,000 and over" and "attended University B" independent?

Case Study #2B: Conditional Probability and Data Dependency

Annual income	University A	University B	TOTAL
Under $20,000	36	24	60
$20,000 to 39,999	109	56	165
$40,000 and over	35	40	75
TOTAL	180	120	300

Suppose we choose a random graduate from this data.

Are the events "income under $20,000" and "attended University B" independent?

P(<$20K | Uni B)=24/300

P(Uni B) 120/300

= 0.2 = **20%**

P(<$20K) = 60/300
= 0.2 = **20%**
In-Dependant - Close Values

Case Study #2C: Conditional Probability and Data Dependency

James is interested in the weather conditions and whether the downtown train he sometimes takes runs on time. For a year, James records whether each day is sunny, cloudy, rainy, or snowy, as well as whether this train arrives on time or is delayed. His results are displayed in the table below.

	On time	Delayed	Total
Sunny	167	3	170
Cloudy	115	5	120
Rainy	40	15	55
Snowy	8	12	20
Total	330	35	365

For these days - are the events "Snowy" and "Delayed" **independant**?

P(Delayed) = 35/365 = 0.096 that is, <10%

P(Delayed | Snowy) = 12/20 = 0.6 that is 60%

$$\frac{P(\text{Delayed} \mid \text{Snowy})}{P(\text{Snowy})} = \frac{12/365}{20/365} = 0.6$$

Since the values obtained are really far apart, it means that the two events are **not** independent of each other. **They are dependant!!**

Case Study #2D: Conditional Probability and Data Dependency

Researchers surveyed one hundred students on which superpower they would most like to have. The two-way table below displays data for the sample of students who responded to the survey.

Superpower	Male	Female	TOTAL
Fly	30	10	40
Invisibility	12	32	44
Other	10	6	16
TOTAL	52	48	100

P(Fly) = 40/100 = 0.4 that is 40%

P(Male | Fly) = 30/40 = 0.75 that is 75%

$$\frac{P(\text{Male} \mid \text{Fly})}{P(\text{Fly})} = \frac{30/100}{40/100} = 0.75$$

Find the probability that the student was male, given the student chose to fly as their superpower.

Case Study #2E: Conditional Probability and Data Dependency

$P(A \mid B) \nLeftrightarrow P(B \mid A)$

Researchers surveyed one hundred students on which superpower they would most like to have. The two-way table below displays data for the sample of students who responded to the survey.

Superpower	Male	Female	TOTAL
Fly	30	10	40
Invisibility	12	32	44
Other	10	6	16
TOTAL	52	48	100

$P(Fly) = 40/100 = 0.4$ that is 40%

$P(Male \mid Fly) = 30/40 = 0.75$ that is 75%

$$\frac{P(Male \mid Fly)}{P(Fly)} = \frac{30/100}{40/100} = 0.75$$

$$\frac{P(Fly \mid Male)}{P(Male)} = \frac{30/100}{52/100} = 0.58$$

Find the probability that the student was male, given the student chose to fly as their superpower.

Find the probability that the student can use fly as super-power, given that the student is a male.

Quiz : Conditional Probability

Let I represent the event where the student chose invisibility as their superpower, and F represent the event where the student was female.

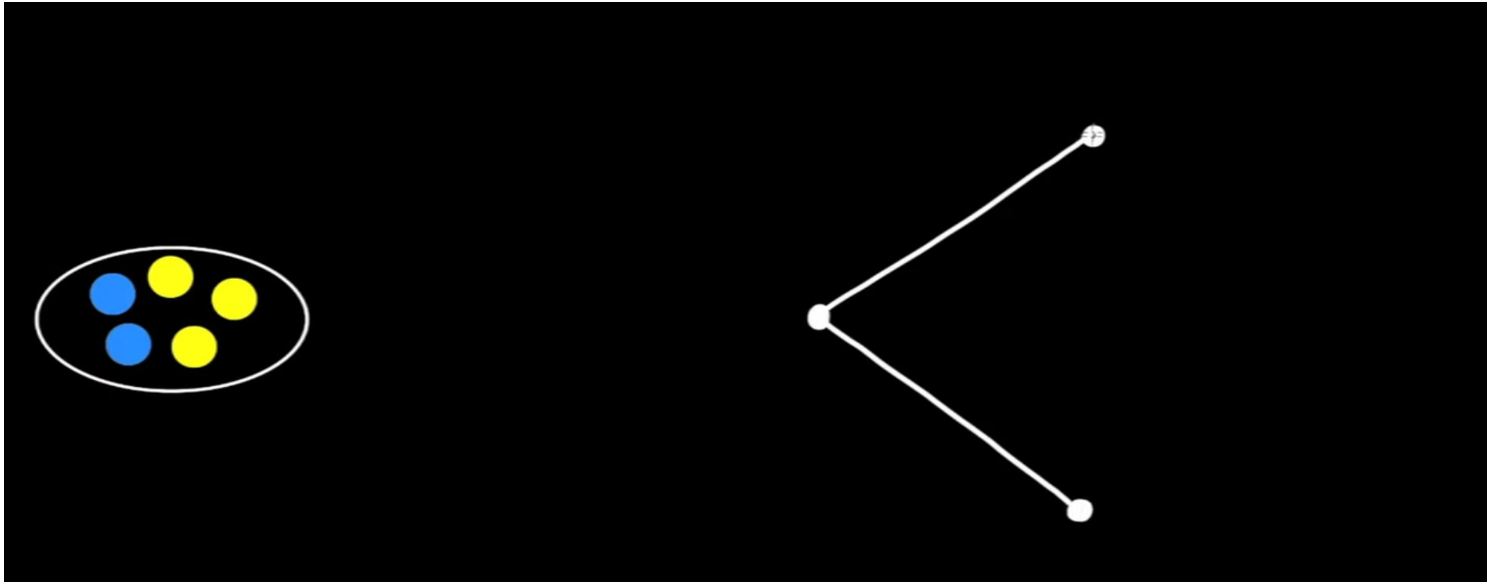
Interpret the meaning of $P\left(\mathrm{I} \mid \mathrm{F}\right) \approx 0.62$.

Choose 1 answer:

(A) About 62% of females chose invisibility as their superpower.

(B) About 62% of people who chose invisibility as their superpower were female.

Conditional Probability : Tree Diagrams



Work Out these probabilities.

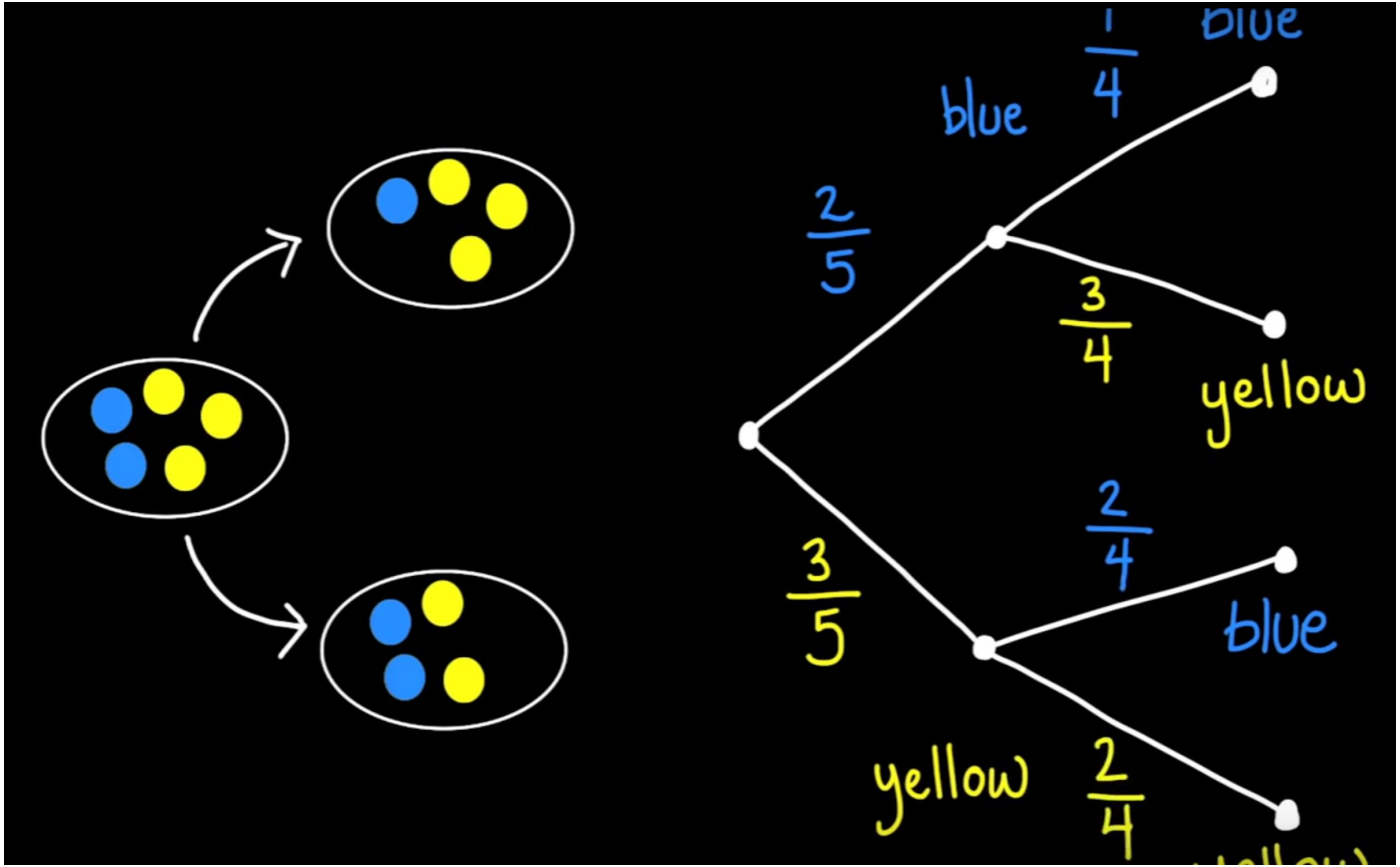
P(B1 and B2) =

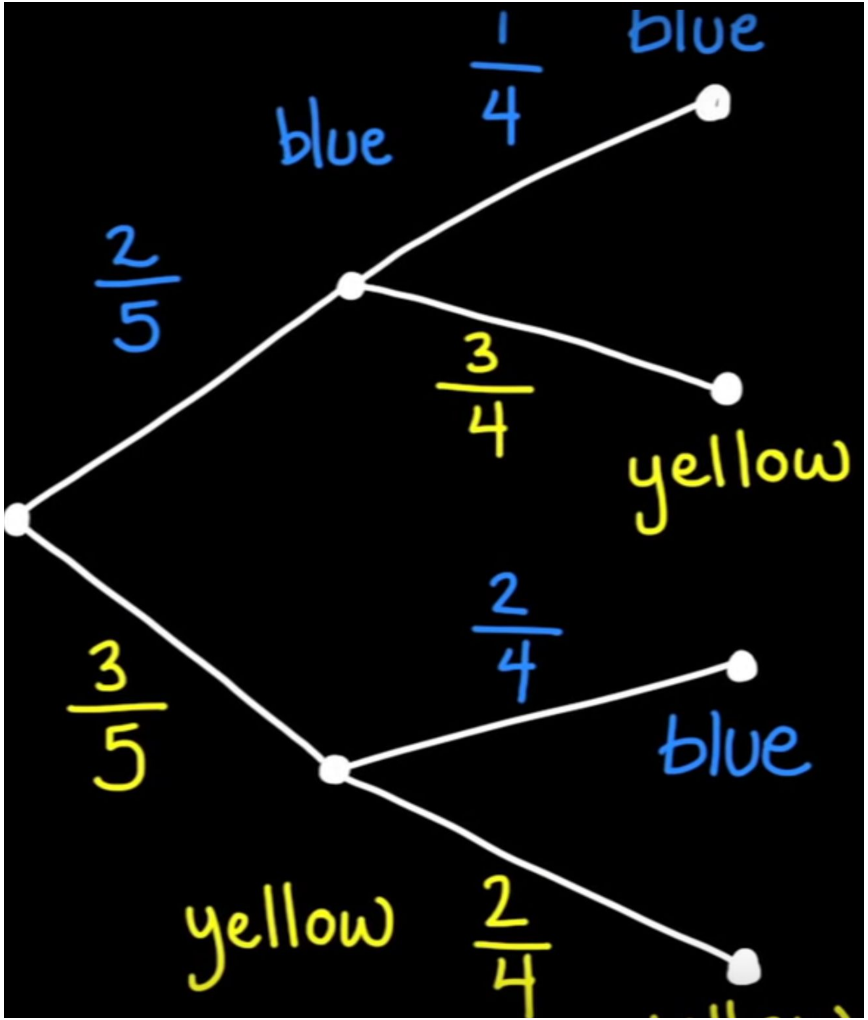
P(B1 and Y2) =

P(Y1 and B2) =

P(Y1 and Y2) =

Conditional Probability : Tree Diagrams



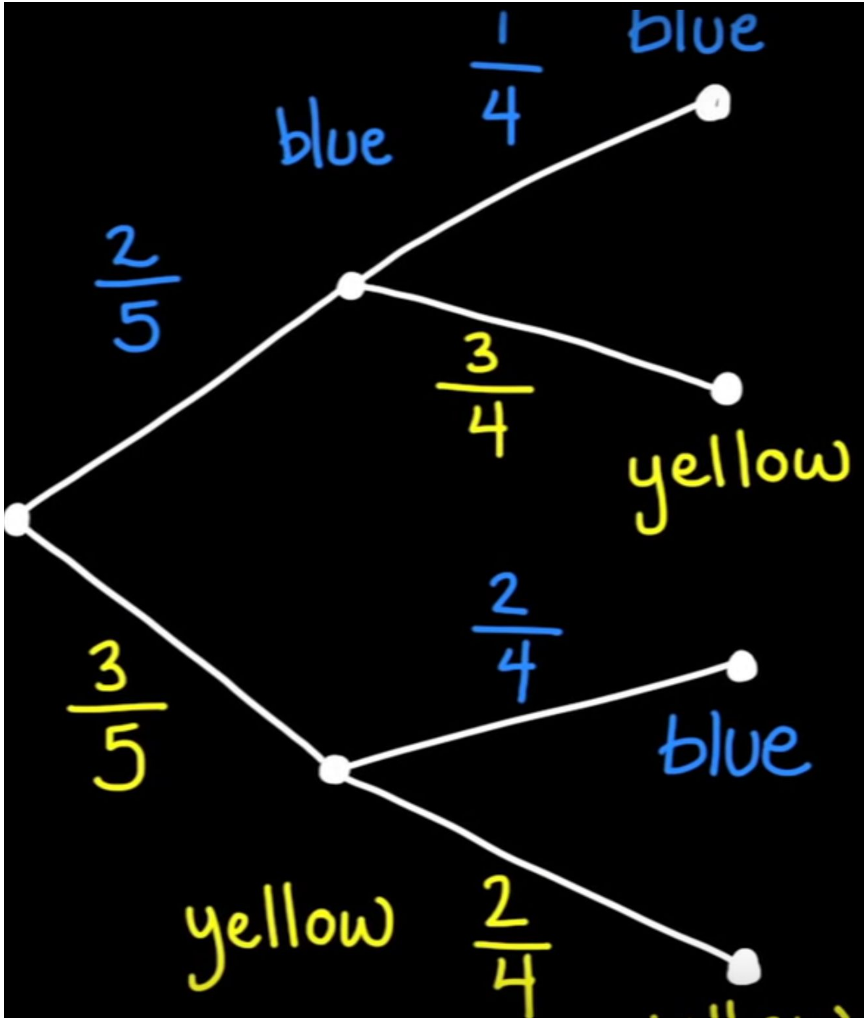


P(B1 and B2) = $\frac{2}{5} * \frac{1}{4} = 2/20 = 1/10$

P(B1 and Y2) = $\frac{2}{5} * \frac{3}{4} = 6/20 = 3/10$

P(Y1 and B2) = $\frac{3}{5} * 2/4 = 6/20 = 3/10$

P(Y1 and Y2) = $\frac{3}{5} * 2/4 = 6/20 = 3/10$



P(B1 and B2) = 1/10

P(B1 and Y2) = 3/10

P(Y1 and B2) = 3/10

P(Y1 and Y2) = 3/10

Question 1:

P(Y2) = P(B1 and Y2) + P(Y1 + Y2)
= 3/10 + 3/10
= 6/10
= 3/5

Question #2:

P(B1 | Y2)

$$= \frac{3/10}{(3/10 + 3/10)}$$

$$= \frac{1}{2}$$

Fraction of favourable to possible outcomes.

Bayes Theorem and Probability Trees - Case Study #1

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. Sensitivity (True Positive Rate) of the test is **93%** and Specificity (True Negative Rate) of the test is **99%.**

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.*

Why magic number : three tests?

Bayes Theorem and Probability Trees- Case Study #1

Sensitivity - True Positive - can be denoted as **P (+ | cancer) = 0.93**

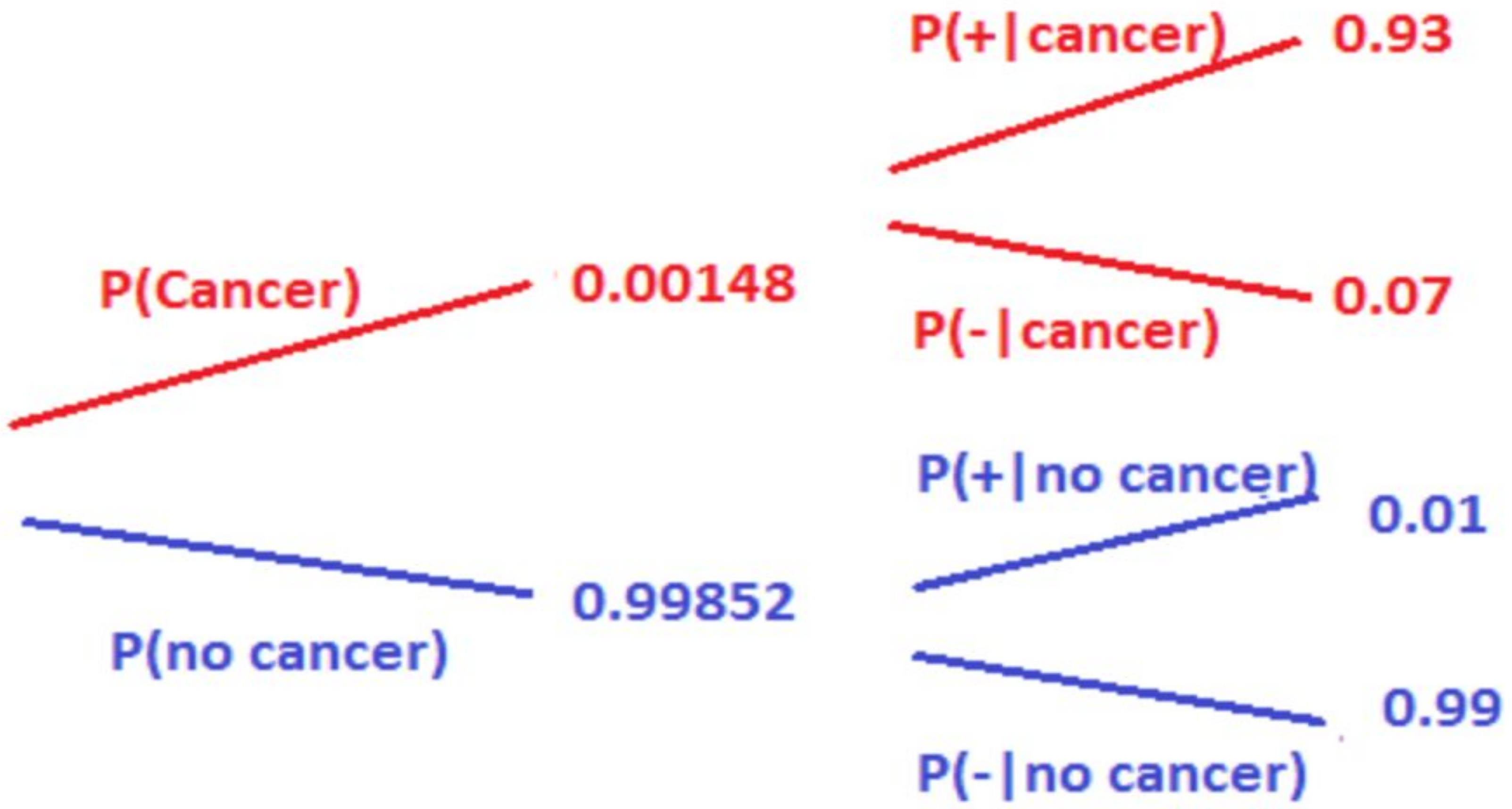
Specificity - True Negative - can be denoted as **P (- | no cancer) = 0.99**

P (cancer) = 0.00148

We need to find:

P (has cancer | first test +)

Bayes Theorem and Probability Trees- Case Study #1



Bayes Theorem and Probability Trees- Case Study #1

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

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

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

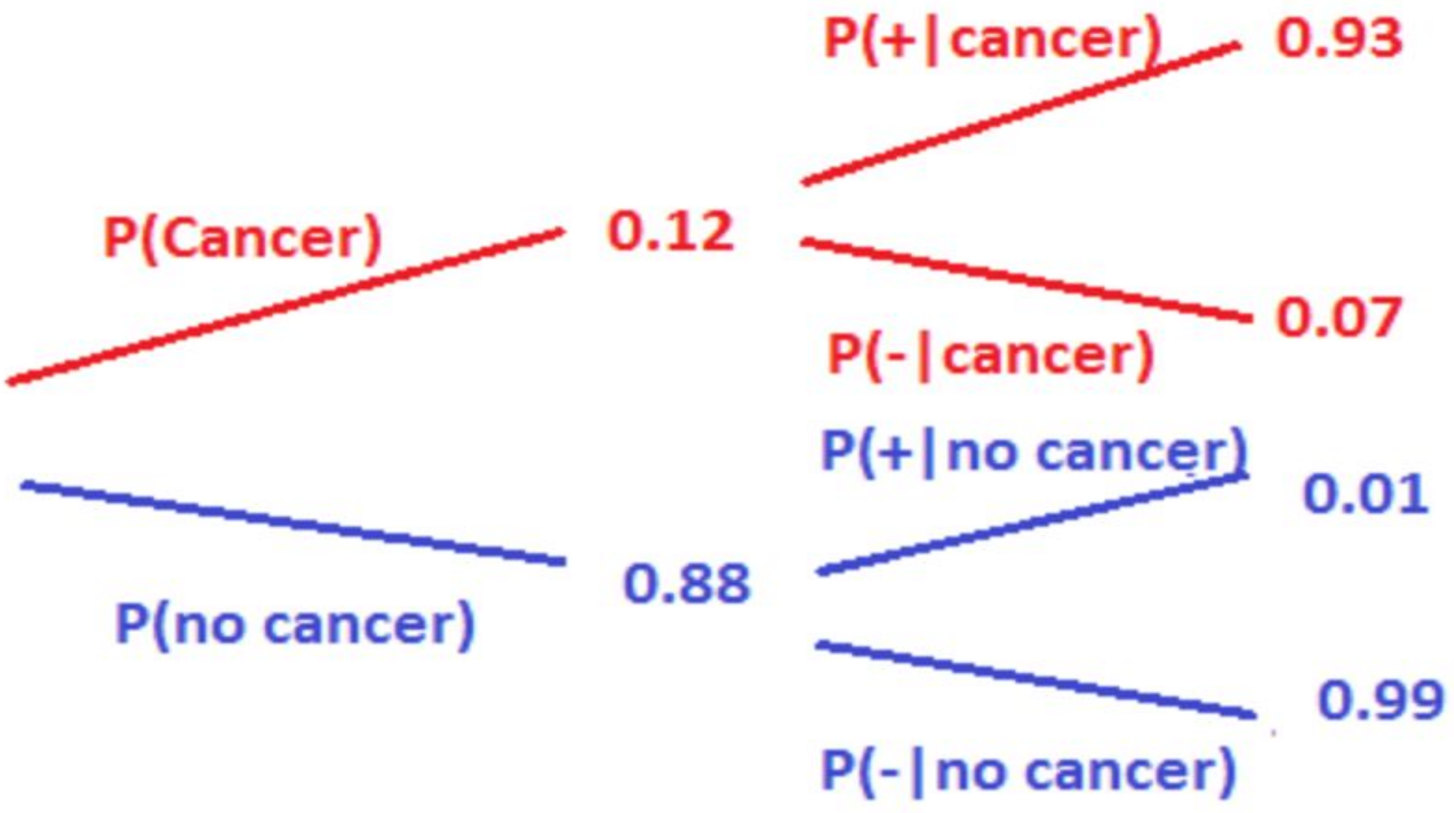
P (no cancer and +) = P (no cancer) * P(+) = 0.99852*0.01

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.**

Case Study #1-Bayes Theorem and Probability Trees - Bayes Updating



Case Study #1-Bayes Theorem and Probability Trees - Bayes Updating

Let's calculate again the probability of having cancer given she tested positive in the second test.

P (cancer and +) = P(cancer) * P(+) = 0.12 * 0.93

P (no cancer and +) = P (no cancer) * P (+) = 0.88 * 0.01

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

$$P\ (CANCER|+) = \frac{P(cancer\ and +)}{P(cancer\ and +) + P(no\ cancer\ and +)} = 0.93$$

Now we see, that a patient who tested positive in the test twice, has a 93% chance of having cancer.

Specificity is given as **0.99**

Bayes Theorem - Case Study #2

Case Study: **99%** of people tested for a disease(e.g Cancer) is positively tested for it, meaning they have cancer. It also implies that **1%** is falsely tested for the cancer.

Question: If you are tested for the disease as "positive" what is the probability that you have cancer?

Values like **1.48** and **0.93** is not given here.

Sample Size of 1000 People Tested for a disease (e.g Cancer).
Hypothesis (**H**) : Has cancer --- 990/1000
Event (**E**) : Tested Positive --- 10/1000

Another Way to ask the Question:

a) What is the probability of you being one in the 10 people falsely identified for cancer?
b) What is the probability of you being one in the 990 people correctly identified for as not having cancer?

Solution: Bayes Theorm

P(H | E) : Given that I am tested positive for cancer, what is the probability of actually having cancer. **This is the question to solve.**

P(E | H) : Given that I may hypothetically have cancer cancer, what is the probability of being tested positive for cancer. Here value is is **0.99**

P(-H) : Probability of the hypothesis “Has Cancer” being untrue. Here value is is **0.99**

P(H) : Probability of the hypothesis “Has Cancer” being true. Guess value. Its also called “prior probability”. To start with its very low. Here we take **0.001**.

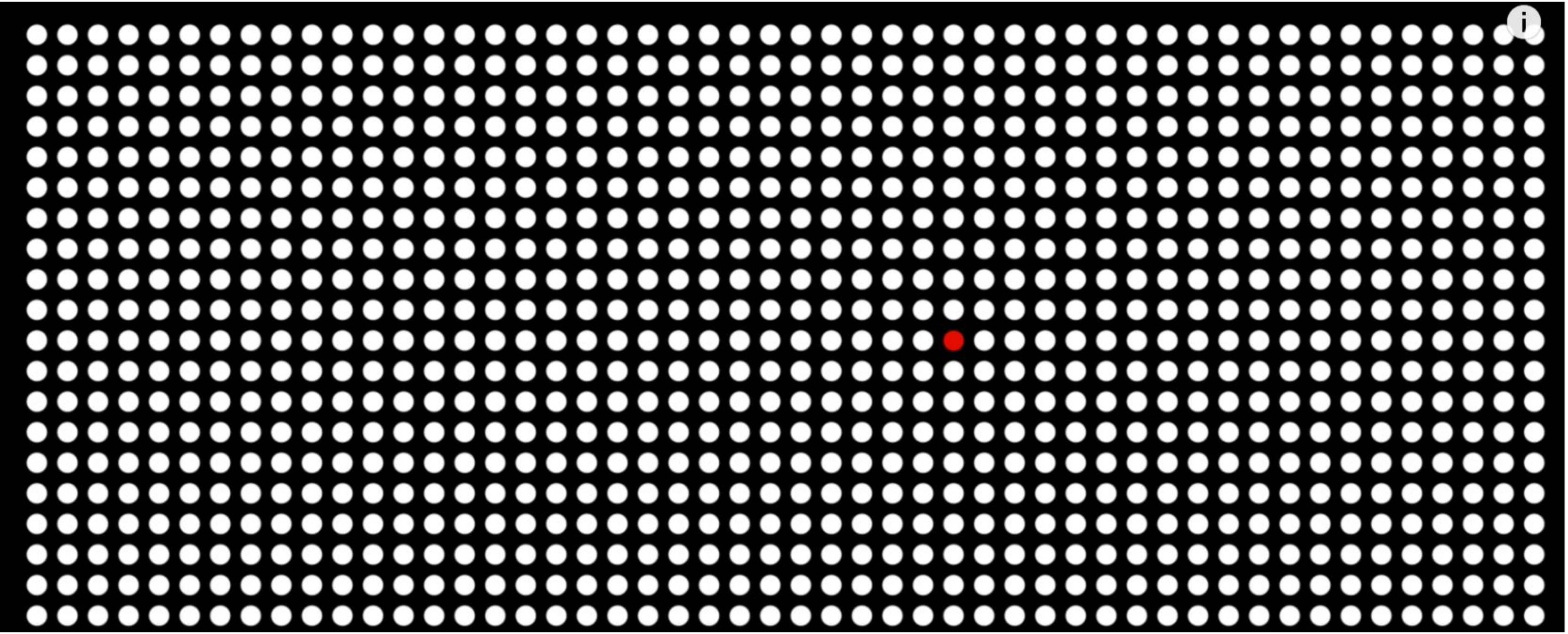
P(E |-H) : Probability of the hypothesis event “Tested Positive” in the conditional event of “Has Cancer” being untrue. Here value is is **0.01**

$$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)} = \frac{P(E \mid H) * P(H)}{P(H) * P(E \mid H) + P(-H) * P(E \mid -H)}$$

$$\frac{0.99 * \boxed{0.001}}{\boxed{0.001} * 0.99 + 0.99 * 0.01} = 0.09 = 9\%$$

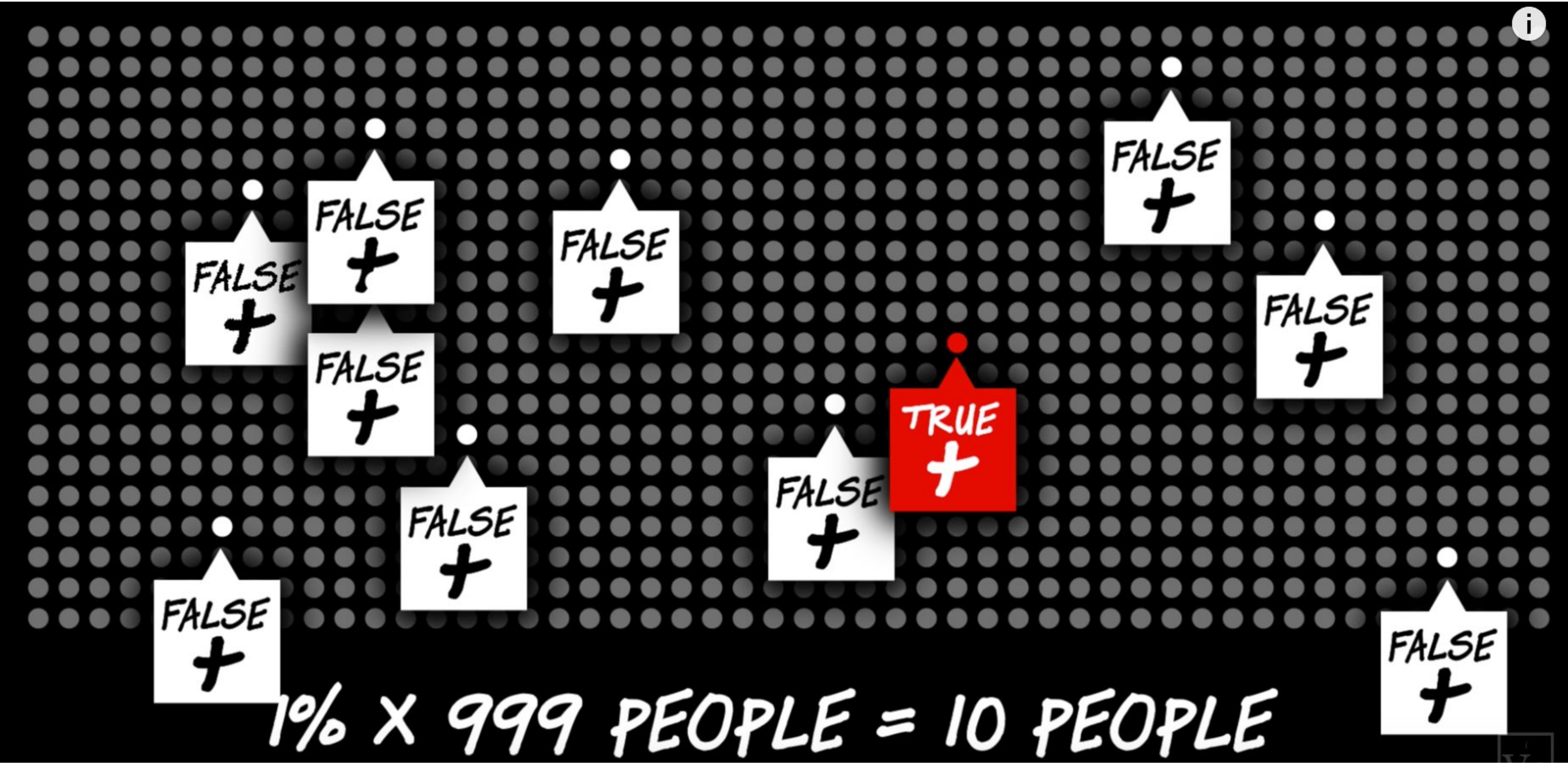
P(H) Value **0.001 is a guess** and is not given here.

Baye's Theorm - Picturization-1

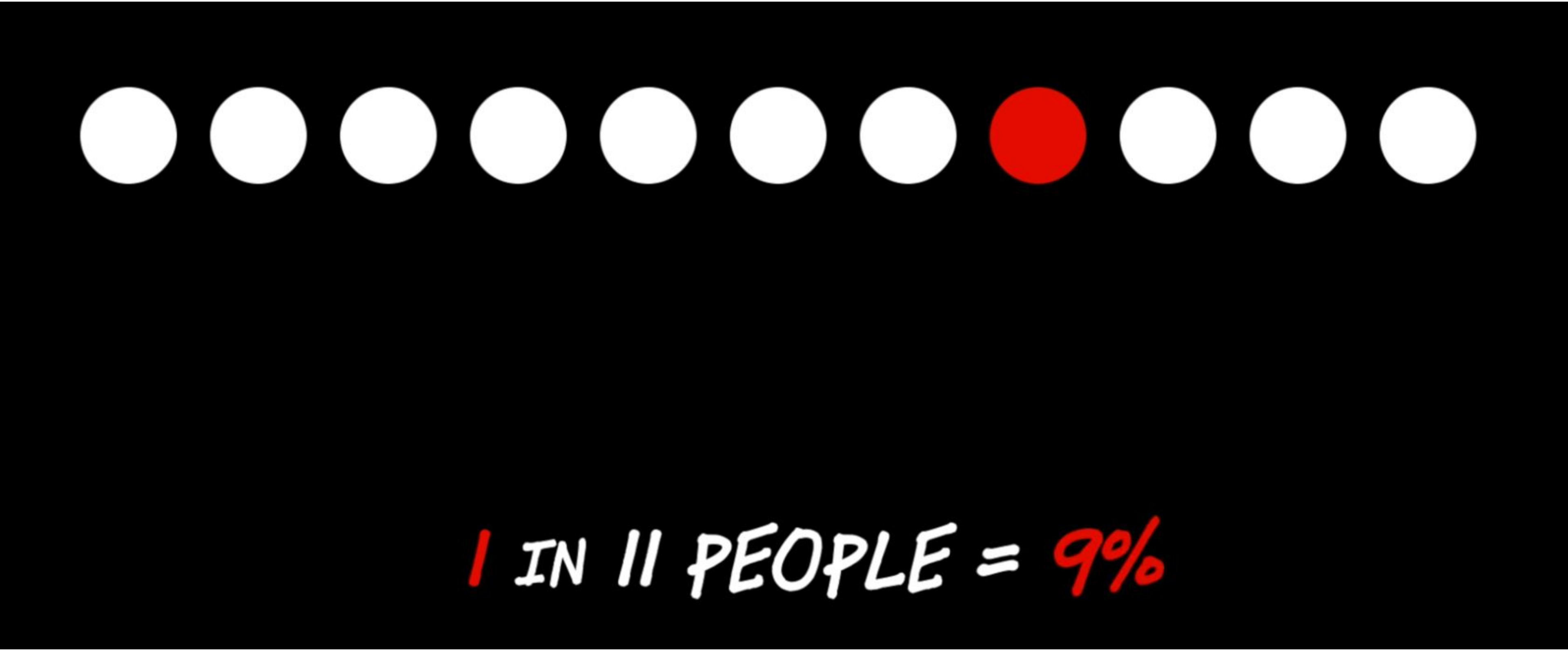


Sample Size of 1000 People Tested for a disease (e.g Cancer)

Baye's Theorm - Picturization-2



Baye's Theorm - Picturization-3



Solution: Bayes Theorm

Posterior Probability

$$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)} = \frac{P(E \mid H) * P(H)}{P(H) * P(E \mid H) + P(-H) * P(E \mid -H)}$$

$$\frac{0.99 * 0.09}{0.09 * 0.99 + 0.91 * 0.01} = 0.907 = 91\%$$

Bayes Theorm Practical Application - Spam Mail Filtering

$$P(Spam \mid Word) = \frac{P(Word \mid Spam) * P(Spam)}{P(Word)} =$$

$$\frac{P(Word \mid Spam) * P(Spam)}{P(Spam) * P(Word \mid Spam) + P(-Spam) * P(Word \mid -Spam)}$$

What does “Bayesian inference” even mean?

Inference = Educated guessing

Thomas Bayes = A nonconformist Presbyterian minister in London back when the United States were still The Colonies.

He wrote two books. One was about theology, and one was about probability.

Bayesian inference = Guessing in the style of Bayes