Revision of Probabilities: Bayes' Rule

- An extension to the conditional law of probabilities
- Enables revision of original probabilities with new information

$$P(X_i|Y) = \frac{P(Y|X_i)P(X_i)}{P(Y|X_1)P(X_1) + P(Y|X_2)P(X_2) + \cdots P(Y|X_n)P(X_n)}$$

- **Starting with a Question:**

  Ever Changed your mind after getting new information? That's essentially what Bayes' Theorem helps us do, but with math.

- **The Basic Idea:**

  Bayes' Theorem calculates the probability of something happening based on new evidence. Imagine guessing if it will rain then seeing dark clouds. Your guess just got better , right?

- **Breaking it Down:**

  Prior: What we initially thought (e.g. 30% chance to rain)

  Likelihood: New evidence (seeing dark clouds)

  Evidence (Marginalization): Overall chance of seeing those clouds

  Posterior: Updated belief about rain after seeing the clouds.

- **Bayesian Thinking:**

  It's like updating a software. You start with version 1.0 (Prior). New updates (Likelihood) come in. After the update, you have version 1.1 (Posterior).

  Bayes' theorem is the update process.

$$\underbrace{\boxed{P(A|B)}}_{\text{posterior}} = \underbrace{\boxed{P(A)}}_{\text{prior}} \times \frac{\boxed{P(B|A)}\ \text{likelihood}}{\boxed{P(B)}\ \text{marginal}}$$

Bayes' Rule is the most important rule in data science. It is the mathematical rule that describes how to update a belief, given some evidence.
In other words – it describes the act of learning.

Posterior = Prior x (Likelihood over Marginal probability)

LIKELIHOOD

The probability of "B" being True, given "A" is True

PRIOR

The probability "A" being True. This is the knowledge.

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

POSTERIOR

The probability of "A" being True, given "B" is True

MARGINALIZATION

The probability "B" being True.

There are four parts:

Posterior probability
(updated probability after the evidence is considered)

Prior probability
(the probability before the evidence is considered)

Likelihood
(probability of the evidence, given the belief is true)

Marginal probability
(probability of the evidence, under any circumstance)

Bayes' Rule in detail

•Bayes' Rule tells you how to calculate a conditional probability with information you already have.

•It is helpful to think in terms of two events – a hypothesis (which can be true or false) and evidence (which can be present or absent).

•However, it can be applied to any type of events, with any number of discrete or continuous outcomes.

$$P(Hypothesis \mid Evidence) = P(Hypothesis) \times \frac{P(Evidence \mid Hypothesis)}{P(Evidence)}$$

Bayes' Rule is the **posterior (or "updated") probability**.
This is a conditional probability. It is the probability of the hypothesis being true, if the evidence is present.

Think of the **prior (or "previous") probability** as your belief in the **hypothesis** before seeing the new evidence. If you had a strong belief in the hypothesis already, the prior probability will be large.

The prior is multiplied by a fraction. Think of this as the **"strength"** of the evidence. The posterior probability is greater when the top part (numerator) is big, and the bottom part (denominator) is small.

The numerator is the **likelihood**. This is another conditional probability. It is the probability of the evidence being present, given the hypothesis is true.
This is not the same as the posterior!

Remember, the "probability of the evidence being present given the hypothesis is true" is not the same as the "probability of the hypothesis being true given the evidence is present".

Now look at the denominator. This is the **marginal probability** of the evidence. That is, it is the probability of the evidence being present, whether the hypothesis is true or false. The smaller the denominator, the more "convincing" the evidence.

- **Why It’s a Big Deal:**

  In the face of uncertainty, being able to update our beliefs with evidence is powerful. From medical diagnoses to finance prediction, updating probabilities can be a game changer.

- **Relation to Machine Learning:**

  In Bayesian Machine learning, ML algorithms often need to update prediction as they receive more data. Bayes’ theorem provides the mathematical muscle for this.

- **A Common Misunderstanding:**

  Often confuse the likelihood with the posterior. Remember, Likelihood is just the new evidence, while the posterior is our updated belief after considering this evidence.

Worked example of Bayes' Rule

Your neighbour is watching their favorite football team. You hear them cheering, and want to estimate the probability their team has scored.
(If you want to think of it as a ML problem, team scoring is the target and we have one feature here, neighbour cheering for predicting whether team scored or not.)

Step 1 – write down the posterior probability of a goal, given cheering

Step 2 – estimate the prior probability of a goal as 2%=0.2

Step 3 – estimate the likelihood probability of cheering, given there's a goal as 90% (perhaps your neighbour won't celebrate if their team is losing badly)
(Hence, probability of cheering given a goal is scored is 90% = 0.9)

The probability of cheering when the goal is not scored assumed to be around 1% = 0.01
So, probability of not goal is (100–2) / 100 = 0.98

Step 4 – estimate the marginal probability of cheering – this could be because:
- a goal has been scored (2% of the time, times 90% probability)
- or any other reason, such as the other team missing a penalty or having a player sent off (98% of the time, times perhaps 1% probability)

Our aim is to find the probability of a goal scored while the neighbours cheering, using Bayes theorem

$$P\left(\frac{Goal}{Cheer}\right) = \frac{P\left(\frac{Cheer}{Goal}\right) P(Goal)}{P(Cheer)}$$

$$P\left(\frac{Goal}{Cheer}\right) = \frac{0.9 * 0.02}{(0.9 * 0.02) + (0.98 * 0.01)}$$

$$= 64.7\%$$

Use cases for Bayes' Rule

Bayes' Rule has use cases in many areas:

- Understanding probability problems (including those in medical research)
- Statistical modelling and inference
- Machine learning algorithms (such as Naive Bayes, Expectation Maximization)
- Quantitative modelling and forecasting

- **Real World Applications:**
  - ➢ Spam filters adjusting based on emails you mark.
  - ➢ Predicting disease spread using new health data
  - ➢ Financial models adjusting to market changes.
- **In Conclusion:**
- Bayes' theorem is all about refining our beliefs with evidence. It's the art and science of evolving our understanding as we learn more. An elegant blend of intuition and mathematics.

Note: So, nest time adjust your plans based on new info, remember: you are thinking like a Bayesian.

Problem

- Machines A, B, and C all produce the same two parts, X and Y. Of all the parts produced, machine A produces 60%, machine B produces 30%, and machine C produces 10%. In addition
  - 40% of the parts made by machine A are part X.
  - 50% of the parts made by machine B are part X.
  - 70% of the parts made by machine C are part X.
- A part produced by this company is randomly sampled and is determined to be an X part.
- With the knowledge that it is an X part, revise the probabilities that the part came from machine A, B, or C.

Event	Prior $P(E_i)$	Conditional $P(X \mid E_i)$	Joint $P(X \cap E_i)$	Posterior
A	.60	.40	$(.60)(.40) = .24$	$\frac{.24}{.46} = .52$
B	.30	.50	.15	$\frac{.15}{.46} = .33$
C	.10	.70	.07	$\frac{.07}{.46} = .15$
			$P(X) = .46$	

Problem

- A particular type of printer ribbon is produced by only two companies, Alamo Ribbon Company and South Jersey Products.
- Suppose Alamo produces 65% of the ribbons and that South Jersey produces 35%.
- Eight percent of the ribbons produced by Alamo are defective and 12% of the South Jersey ribbons are defective
- A customer purchases a new ribbon. What is the probability that Alamo produced the ribbon? What is the probability that South Jersey produced the ribbon?

Revision of Probabilities with Bayes' Rule: Ribbon Problem

$$P(Alamo) = 0.65$$

$$P(SouthJersey) = 0.35$$

$$P(d|Alamo) = 0.08$$

$$P(d|SouthJersey) = 0.12$$

$$P(Alamo|d) = \frac{P(d|Alamo)\cdot P(Alamo)}{P(d|Alamo)\cdot P(Alamo) + P(d|SouthJersey)\cdot P(SouthJersey)}$$

$$= \frac{(0.08)(0.65)}{(0.08)(0.65)+(0.12)(0.35)} = 0.553$$

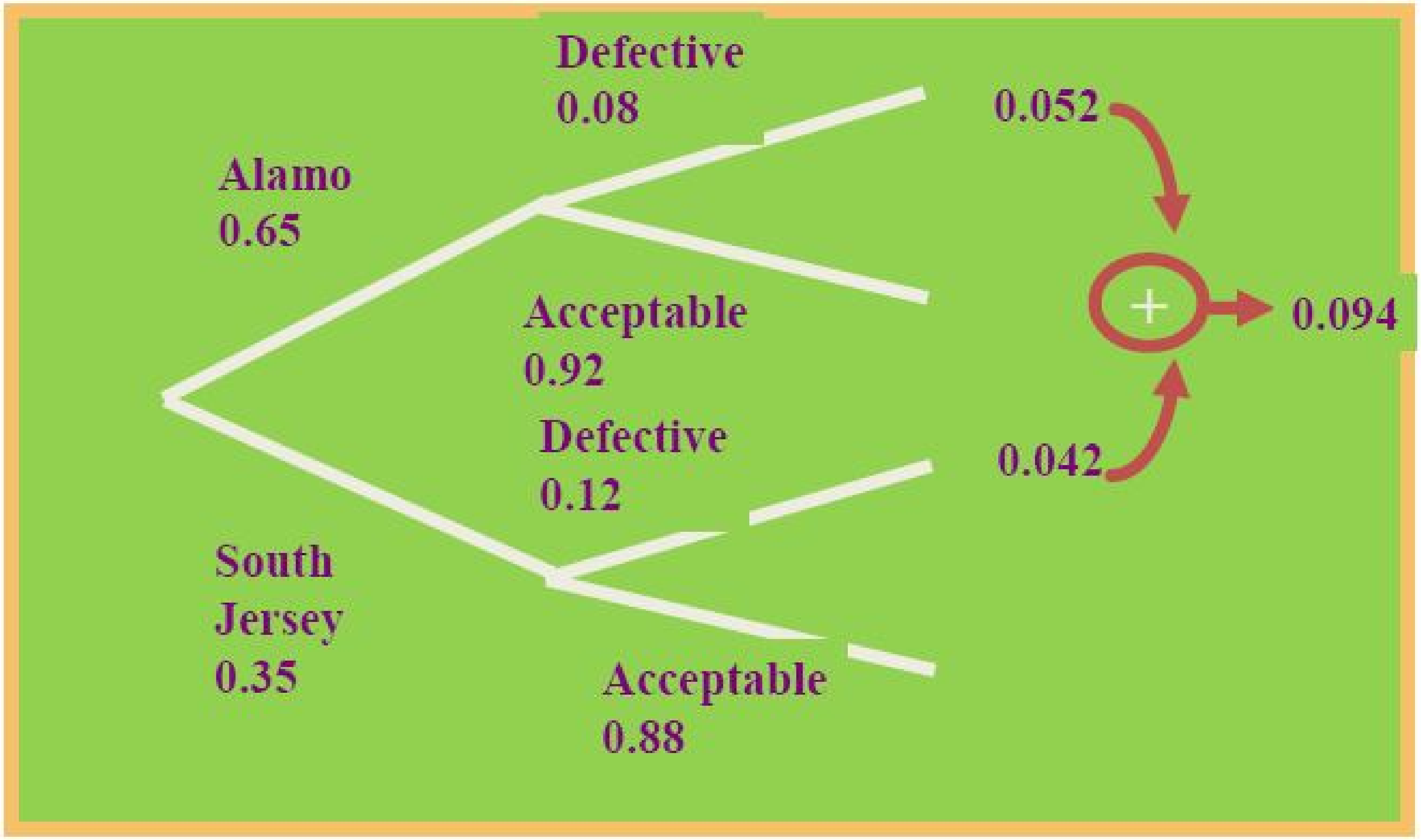
$$P(SouthJersey|d) = \frac{P(d|SouthJersey)\cdot P(SouthJersey)}{P(d|Alamo)\cdot P(Alamo) + P(d|SouthJersey)\cdot P(SouthJersey)}$$

$$= \frac{(0.12)(0.35)}{(0.08)(0.65)+(0.12)(0.35)} = 0.447$$

Revision of Probabilities with Bayes' Rule: Ribbon Problem

Event	Prior Probability $P(E_i)$	Conditional Probability $P(d \mid E_i)$	Joint Probability $P(E_i \cap d)$	Revised Probability $P(E_i \mid d)$
Alamo	0.65	0.08	0.052	$\frac{0.052}{0.094}$ =0.553
South Jersey	0.35	0.12	0.042 0.094	$\frac{0.042}{0.094}$ =0.447

Revision of Probabilities with Bayes' Rule: Ribbon Problem



Example: A manufacturer produces cars in two factories. Ten percent cars produced in factory A and 5% cars produced in factory B are defective. Factory A produces 100000 cars and factory B produces 50000 cars annually, compute the following:
(a) Probability of purchasing a defective car
(b) If the car is defective, probability that it was manufactured in factory A.

Solution:

Probability that car is defective and was produced in factory A
P(D|A)= 1/10

Probability that car is defective and was produced in factory B
P(D|B)= 1/20

Probability that car purchased was manufactured at factory A
P(A) = 2/3

Probability that car purchased was manufactured at factory B
P(B) =1/3

Probability that car purchased is defective

$$P(D) = P(D|A)P(A) + P(D|B)P(B) = \frac{1}{10} \times \frac{2}{3} + \frac{1}{20} \times \frac{1}{3} = \frac{1}{12}$$

Probability that purchased car is defective was produced at factory A

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{1}{10} \times \frac{2}{3} \div \frac{1}{12} = \frac{4}{5}$$

Example

- In a class 75% students are boys. On a particular day 30% of boys and 25% of girls habitually wear jeans. A student is selected randomly, what is the probability that selected student is boy wearing jeans?
- $P(B) =$
- $P(G) =$
- $P(J|B) =$
- $P(J|G) =$
- $P(B|J) = \dfrac{P(B)P(J|B)}{P(B)P(J|B) + P(G)P(J|G)}$