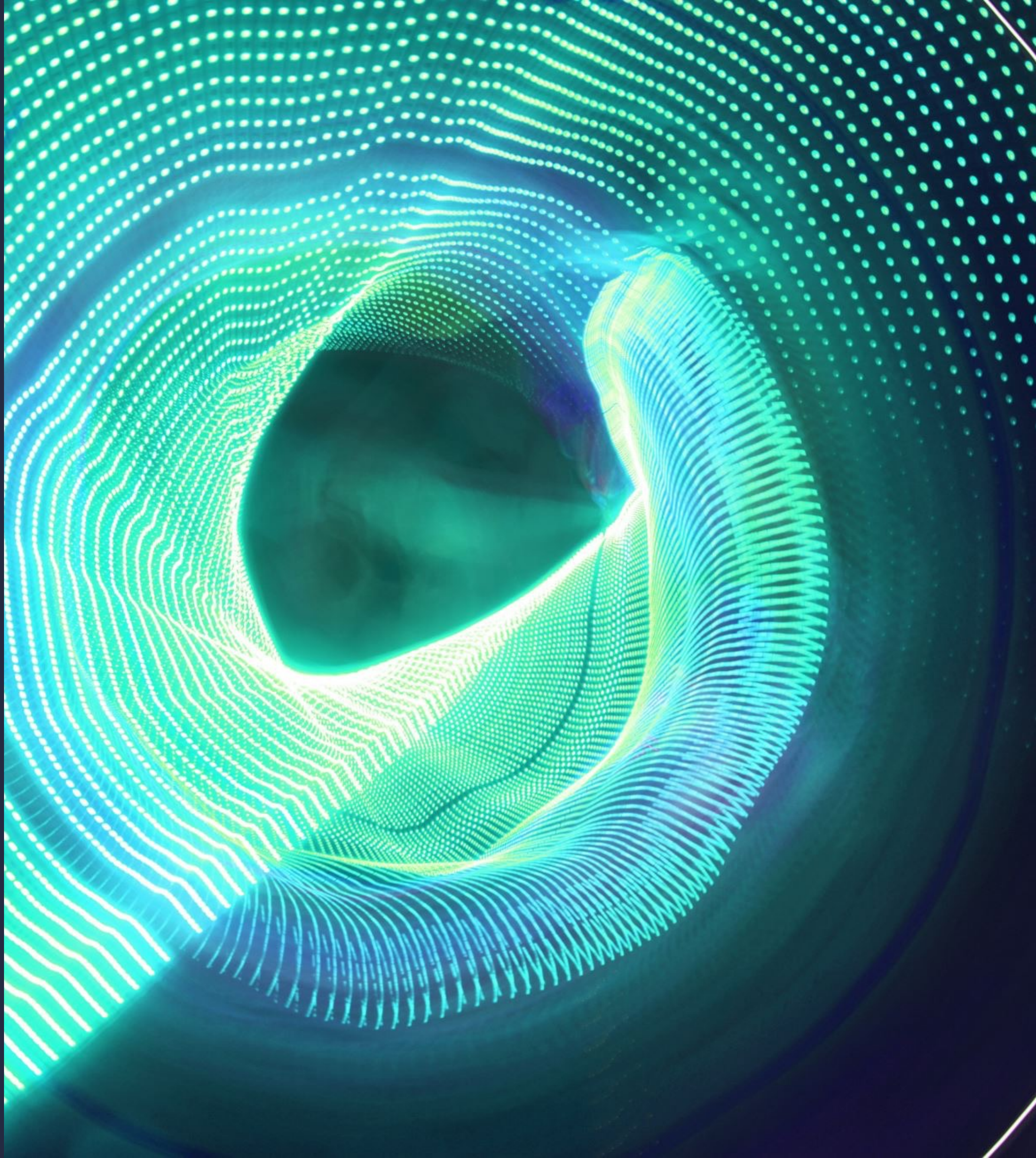
Independence + Joint, Marginal and Conditional Probability



The General Multiplication Rule (a.k.a. the Chain Rule)

- When we calculate probabilities involving one event AND another event occurring, we multiply their probabilities.
- In some cases, the first event happening impacts the probability of the second event. We call these **dependent events**.
  - In that case => **$P(A,B) = P(A) \times P(B|A)$**

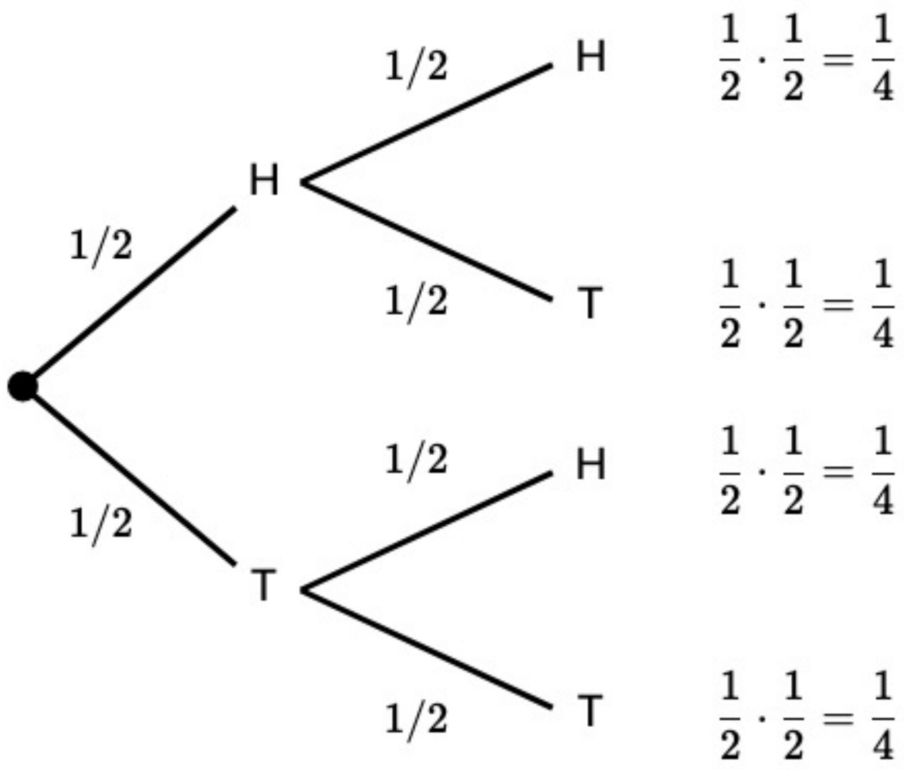
- In other cases, the first event happening does not impact the probability of the second. We call these **independent events**.
  - In that case => **$P(A,B) = P(A) \times P(B)$**

Independent Events: Flipping a coin twice

- What is the probability of flipping a fair coin and getting "heads" twice in a row? That is, what is the probability of getting heads on the first flip AND heads on the second flip?
- Imagine we had 100 people simulate this and flip a coin twice. On average, 50 people would get heads on the first flip, and then 25 of them would get heads again. So 25 out of the original 100 people – ¼ of them – would get heads twice in a row.
- The number of people we start with doesn't really matter. Theoretically, 1/2 of the original group will get heads, and ½ of that group will get heads again. To find a fraction of a fraction, we multiply.

Independent Events

We can represent this concept with a tree diagram :



Independent Events

- We multiply the probabilities along the branches to find the overall probability of one event AND the next even occurring.

- For example, the probability of getting 2 "tails" in a row would be:

- $P(T \text{ and } T) = ½ . ½ = 1/4$

Practice Problem 1

- Suppose that we are going to roll two fair 6-sided dice.

- Find the probability that both dice show a 3.

Practice Problem 2

- On a multiple choice test, problem 1 has 4 choices and problem 2 has 3 choices. Each problem has only 1 correct answer. What is the probabilty of randomly guessing the correct answer to both problems ?

Dependent events: Drawing cards

- We can use a similar strategy even when we are dealing with dependent events.

- Consider drawing two cards, without replacement, from a standard deck of 52 cards. That means we are drawing the first card, leaving it out, and then drawing the second card.

- What is the probability that both cards selected are black?

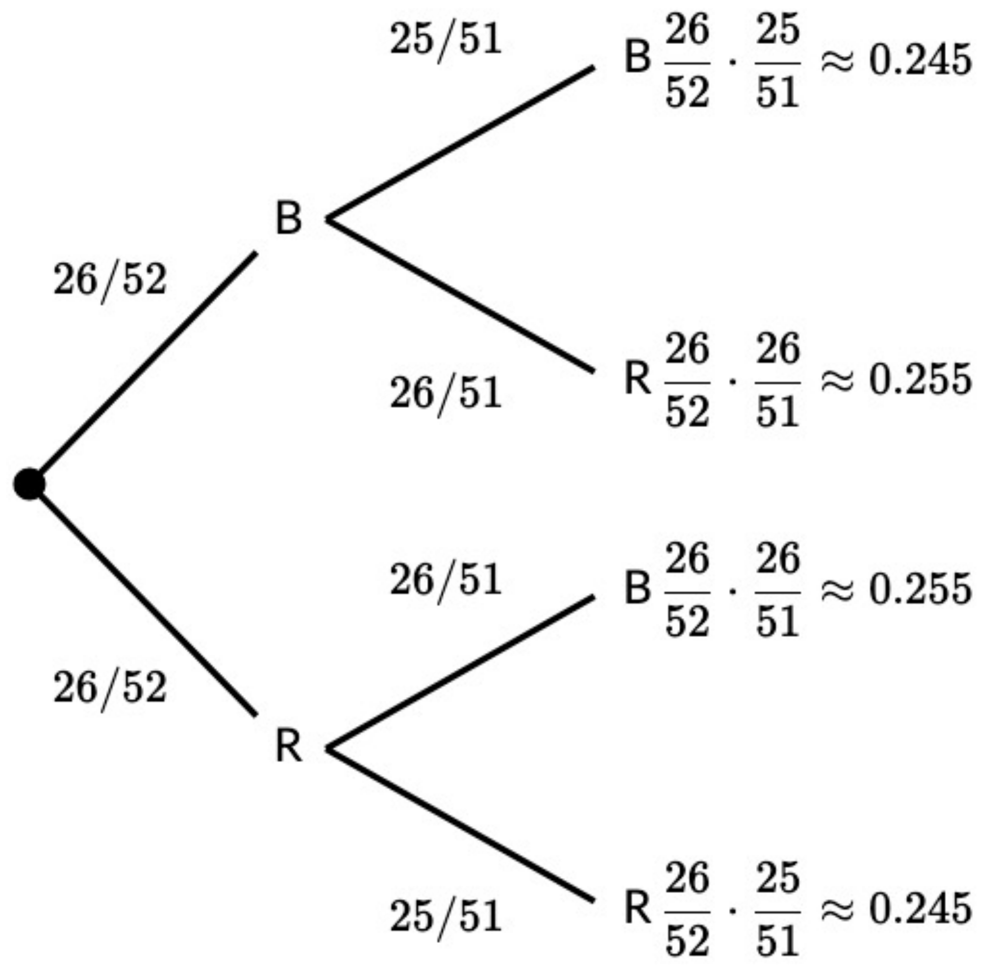
Dependent events: Drawing cards

- Half of the 52 cards are black, so the probability that the first card is black is 26/52.

- But the probability of getting a black card changes on the next draw, since the number of black cards and the total number of cards have both been decreased by 1.

Dependent events: Drawing cards

- Here's what the probabilities would look like in a tree diagram:



Dependent events: Drawing cards

- So the probability that both cards are black is:

$$P\text{ (both black)} = \frac{26}{52} . \frac{25}{51} \approx 0.245$$

Practice Problem 1

- A table of 5 students has 3 seniors and 2 juniors. The teacher is going to pick 2 students at random from this group to present homework solutions.

- Find the probability that both students selected are juniors.

Practice Problem 2

- The probability of winning the Nobel Prize if you have a PhD in Physics is 1 in a million [P(A|B = 0.000001]
- Only 1 in 10,000 people have a PhD in Physics [P(B) = 0.0001]

- What is the probability of a person both having a PhD in Physics and winning the Nobel Prize?
  1. Smaller than 1 in a million
  2. Greater than 1 in a million
  3. Impossible to tell

Independence with regard to Conditional Probability

- Two events A and B are independent if and only if :

$$P(A, B) = P(A)P(B)$$

- Meaning :

$$P(A) = P(A|B)$$

$$P(B) = P(B|A)$$

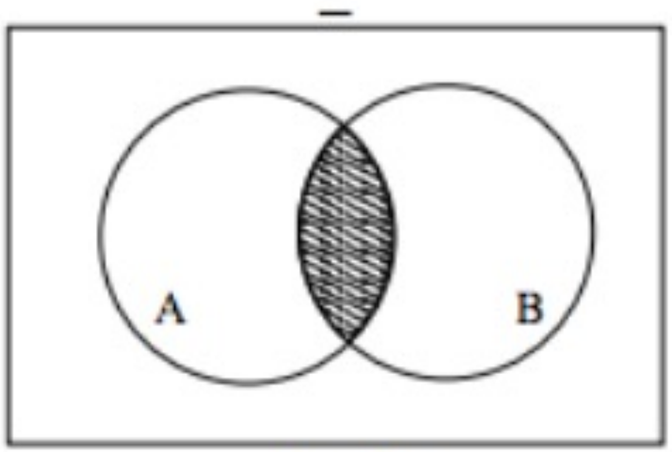
General Definition

- Given events A and B in Ω, with P (B) > 0 the conditional probability of A given B is

$$P(A|B) =_{\mathrm{DEF}} \frac{P(A \cap B)}{P(B)}$$

- P (A and B) or P(A,B) is the *joint* probability of A and B.
- The prob that a person is rich and famous –> joint
  The prob that a person is rich if they are famous –> conditional
  The prob that a person is famous if they are rich –> conditional

Conditional Probability with Venn Diagrams



$A \cap B$

$$
\begin{aligned}
P(A) &= \text{size of } A \text{ relative to } \Omega \\
P(A, B) &= \text{size of } A \cap B \text{ relative to } \Omega \\
P(A|B) &= \text{size of } A \cap B \text{ relative to } B
\end{aligned}
$$

NLP example

- We sample word bigrams (pairs) from a large text T
- *Sample space and events*:
  - $\Omega = \{(w1, w2) \in T\}$ => the set of bigrams in T
  - $A = \{(w1, w2) \in T \mid w1 = run\}$ => bigrams starting with *run*
  - $B = \{(w1, w2) \in T \mid w2 = amok\}$ => bigrams ending with *amok*
- *Probabilities* :
  - $P(A) = 10^{-3}$ => proba that a bigram starts with *run*
  - $P(B) = 10^{-6}$ => proba that a bigram ends with *amok*
  - $P(A, B) = 10^{-7}$ => proba of the bigram *run amok*
- So what is the probability of *amok* following *run*? Of *run* preceding *amok*?

NLP example

- So what is the probability of *amok* following *run*? Of *run* preceding *amok*?

- P (*amok* following *run*) = P(B|A) = 10^{-7}/ 10^{-3} = 0.0001

- P (*run* preceding *amok*) = P (A|B) = 10^{-7}/ 10^{-6} = 0.1

Marginal Probability

- Marginalization, or the law of total probability
- Best illustrated with an example

Marginal Probability

Joint probabilities for rain and wind :

P (*no wind*) = ?
P (*light rain*) = ?

	no wind	some wind	strong wind	storm
no rain	0.1	0.2	0.05	0.01
light rain	0.05	0.1	0.15	0.04
heavy rain	0.05	0.1	0.1	0.05

Marginal Probability

- P(*no wind*) = 0.1+0.05+0.05 = 0.2
- P(*light rain*) = 0.05 + 0.1 + 0.15 + 0.04 = 0.34

=> Marginal because we're interested in the total probability, we've removed any dependence regarding the wind or the rain (depending on the case).

	no wind	some wind	strong wind	storm
no rain	0.1	0.2	0.05	0.01
light rain	0.05	0.1	0.15	0.04
heavy rain	0.05	0.1	0.1	0.05

Marginal Probability

- P(*no wind*) = 0.2
- P(*light rain*) = 0.34
- Combine to get conditional probabilities:
  - P (no wind | light rain) = 0.05/0.34 = 0.147
  - P (light rain | no wind) = 0.05/0.2 = 0.25

	no wind	some wind	strong wind	storm
no rain	0.1	0.2	0.05	0.01
light rain	0.05	0.1	0.15	0.04
heavy rain	0.05	0.1	0.1	0.05

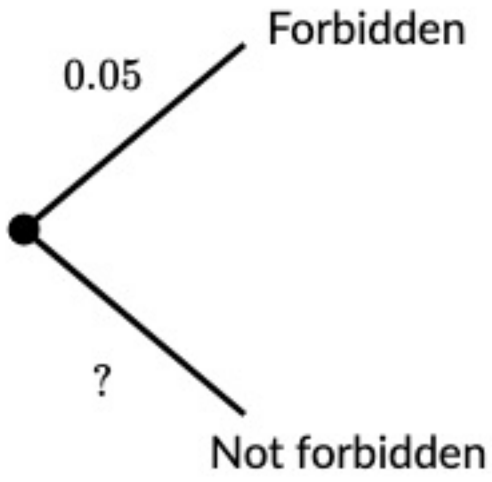
Practice Case: Airport Bags

- An airport screens bags for forbidden items, and an alarm is supposed to be triggered when a forbidden item is detected.
- Suppose that 5%, percent of bags contain forbidden items.
- If a bag contains a forbidden item, there is a 98%, percent chance that it triggers the alarm.
- If a bag doesn't contain a forbidden item, there is an 8%, percent chance that it triggers the alarm.

Given that a randomly chosen bag triggers the alarm, what is the probability that it contains a forbidden item?

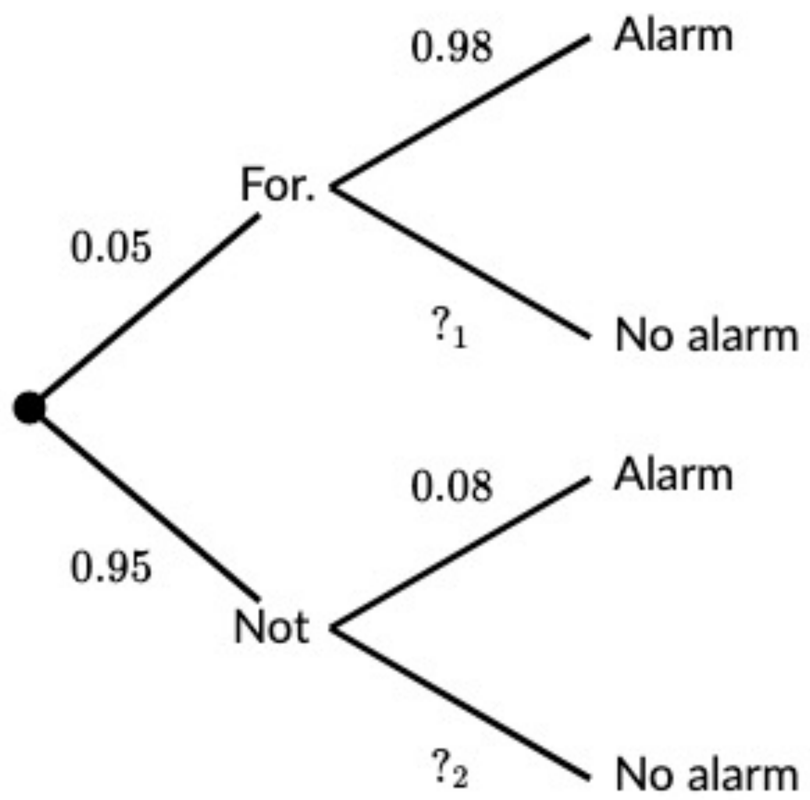
Practice Case: Airport Bags

- The chance that the alarm is triggered depends on whether or not the bag contains a forbidden item, so we should first distinguish between bags that contain a forbidden item and those that don't.



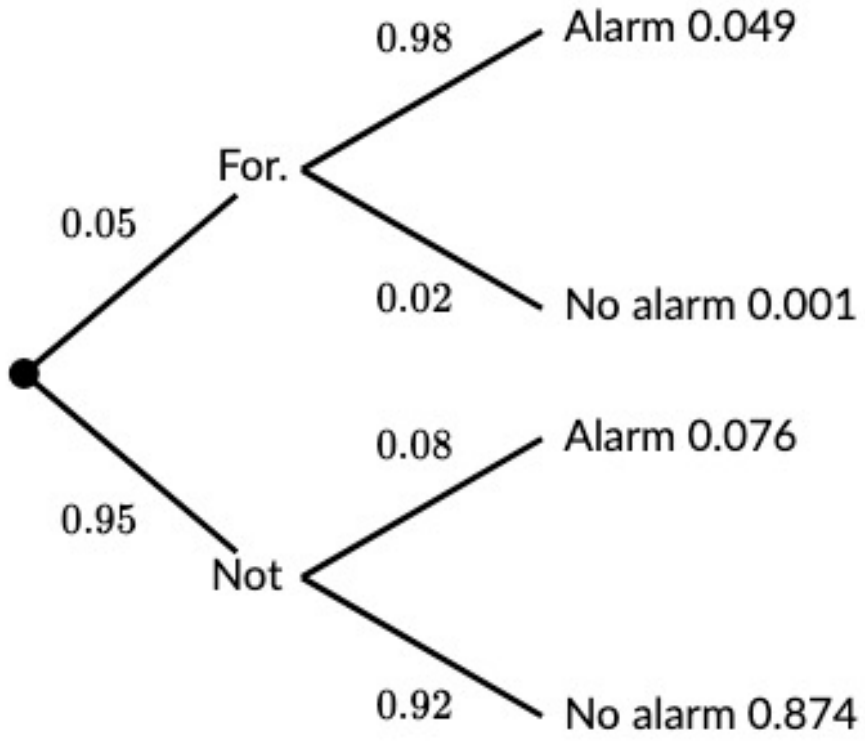
Practice Case: Airport Bags

- "If a bag contains a forbidden item, there is a 98%, percent chance that it triggers the alarm."
- "If a bag doesn't contain a forbidden item, there is an 8%, percent chance that it triggers the alarm. »



Practice Case: Airport Bags

- We multiply the probabilities along the branches to complete the tree diagram.
- Here's the completed diagram:



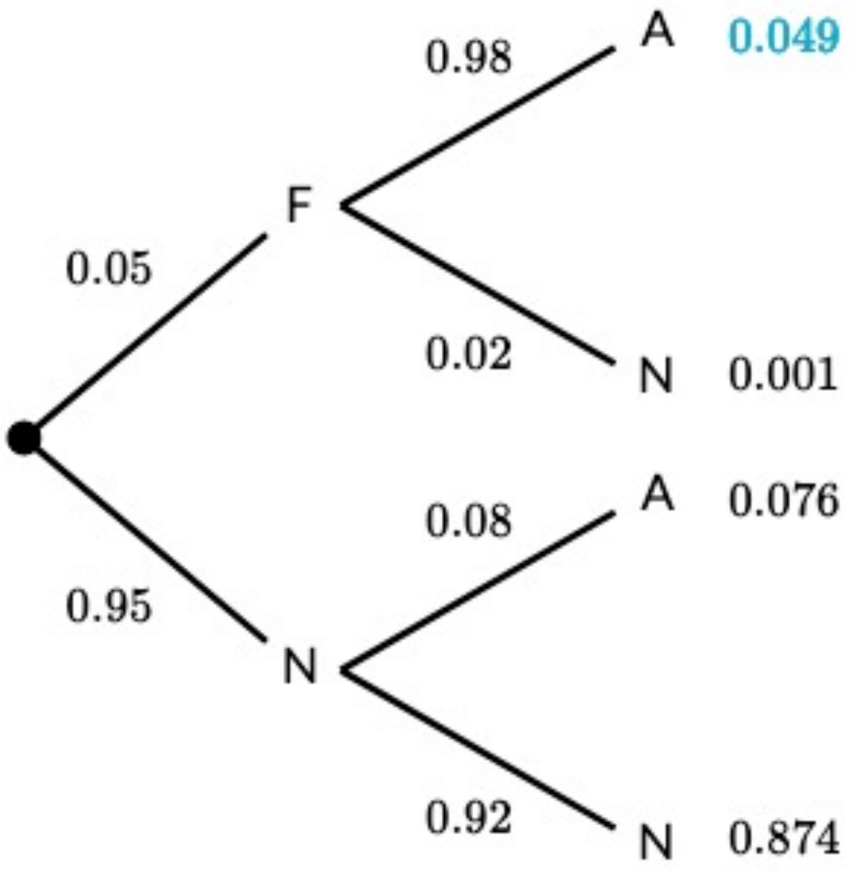
Practice Case: Airport Bags

- **"Given a randomly chosen bag triggers the alarm, what is the probability that it contains a forbidden item?"**
- Use the probabilities from the tree diagram and the conditional probability formula:

$$P\,(\text{forbidden} \mid \text{alarm}) = P(F \cap A) \,/\, P(A)$$

Practice Case: Airport Bags

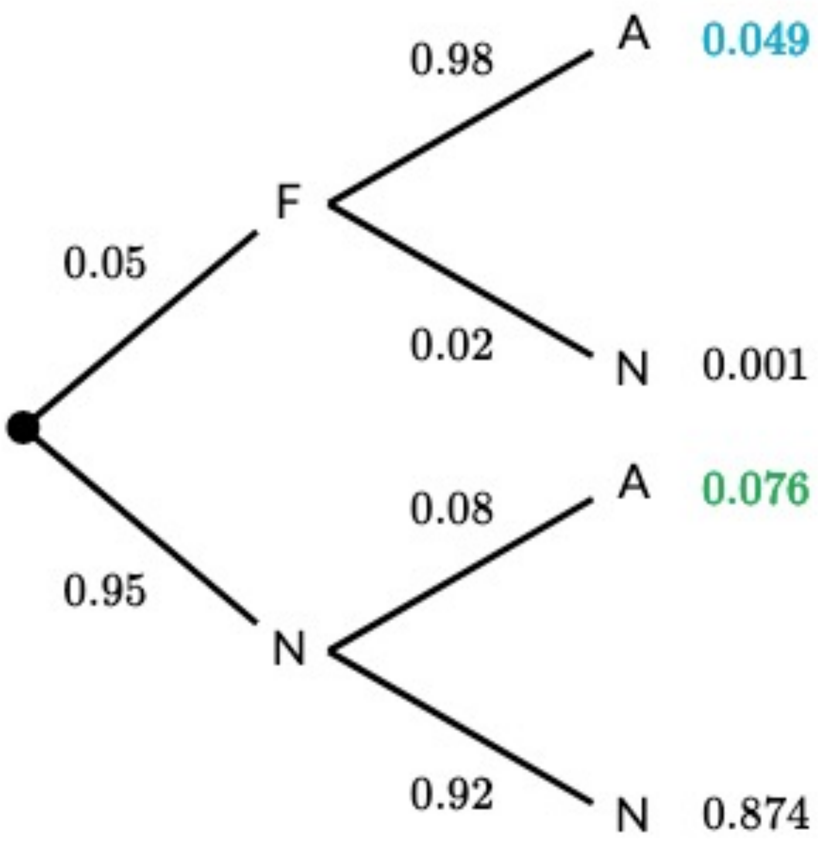
$P(F \cap A) = 0.05 \times 0.98 = 0.049$



Practice Case: Airport Bags

There are two situations where bags can trigger the alarm, so we add those two probabilities together to get the marginal probability :

$P(A) = P(F \cap A) + P(N \cap A) = 0.125$



Practice Case: Airport Bags

- Finally : **Given that a randomly chosen bag triggers the alarm, what is the probability that it contains a forbidden item?**:
- $P(F|A) = P(F \cap A) / P(A)$

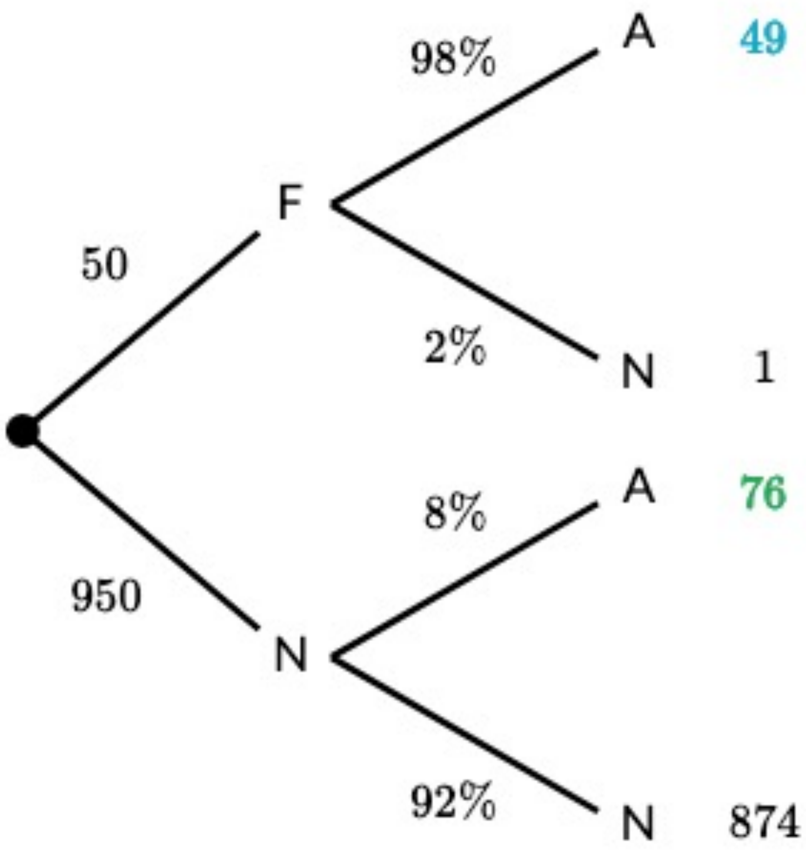
  $= 0.049 / (0.076 + 0.049)$

  $= 0.049 / 0.125$

  $= 0.392$
- Why does this make sense ? Why is it so low ?

Practice Case: Airport Bags

- In this problem, with a sample of 1000 bags, we can clearly see there are more bags without forbidden items that trigger the alarm…
- So when all we know is that a random bag triggers the alarm, it's actually more likely to NOT contain a forbidden item.



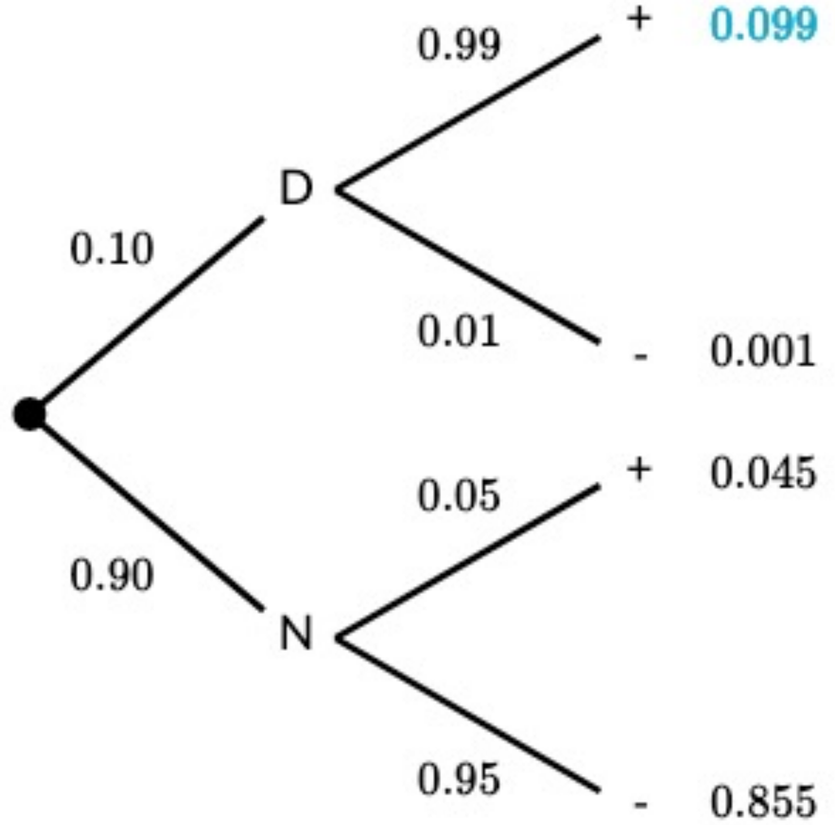
Practice Case : Testing for a Disease

- A hospital is testing patients for a certain disease. If a patient has the disease, the test is designed to return a "positive" result. If a patient does not have the disease, the test should return a "negative" result. No test is perfect though.
- 99% of patients who have the disease will test positive.
- 5% of patients who don't have the disease will also test positive.
- 10% of the population in question has the disease.

If a random patient tests positive, what is the probability that they have the disease?

Practice Case : Testing for a Disease

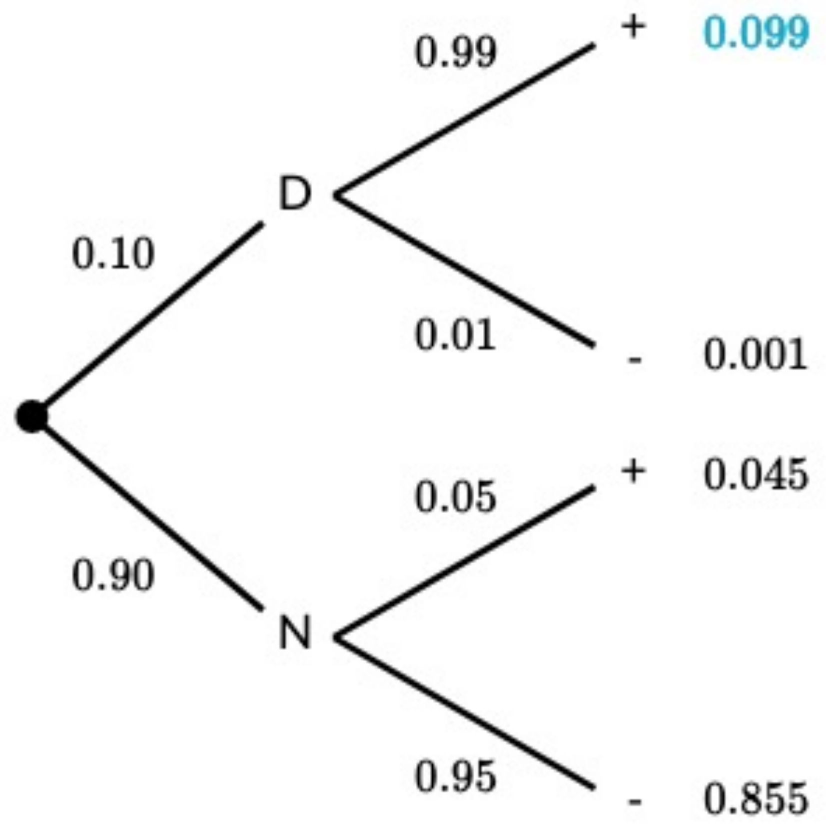
Tree diagram :



Practice Case : Testing for a Disease

Since 10% of patients have the disease, and 99% of those patients test positive, we can multiply those probabilities:

$P(D \cap +) = 0.10 \times 0.99 = 0.099$



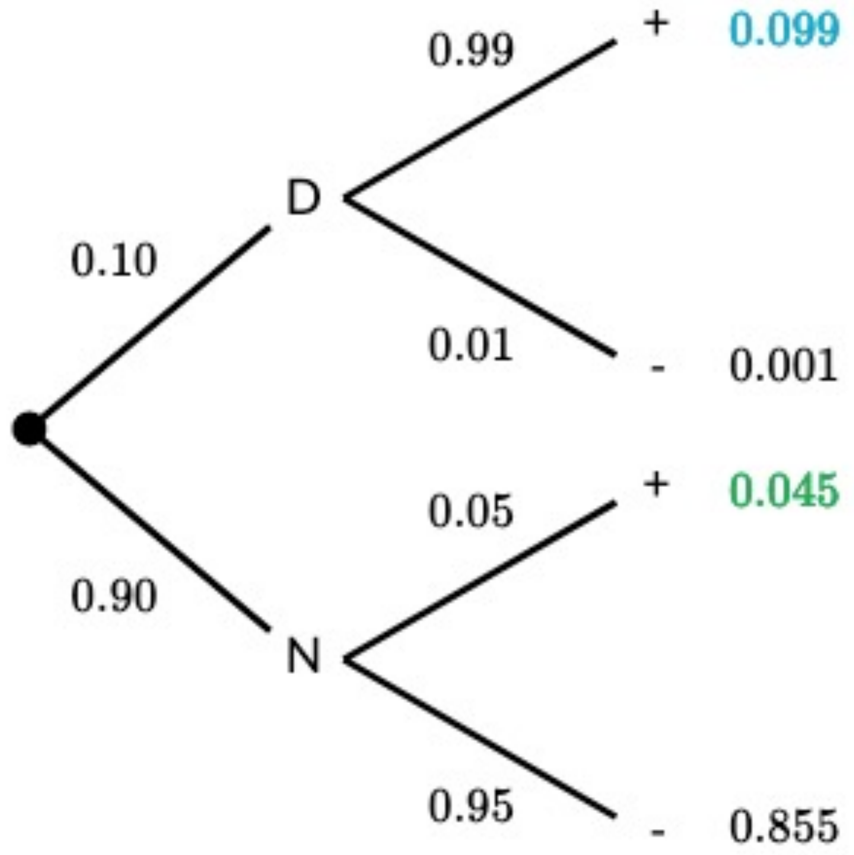
Practice Case : Testing for a Disease

There are two situations where patients can test positive, so we add those two probabilities together:

$P(+) = P(D \cap +) + P(N \cap +)$

$= 0.099 + 0.045$

$= 0.144$



Practice Case : Testing for a Disease

If a random patient tests positive, what is the probability that they have the disease?

$P(D \mid +) = P(+)\ P(D \cap +)$

$= 0.099 / (0.099 + 0.045)$

$= 0.099 / 0.144$

$= 0.6875$