

# Conditional Probability

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

**Problem:** You roll a standard six-sided die. Your outcome are  $\{1, 2, 3, 4, 5, 6\}$

- 1) what is the probability of rolling a 5?
- 2) what is the probability of rolling a 5, **given that** the result was a number greater than 3?

Are above answer same or different ?

If they are same, why ?

If they are different, why ?



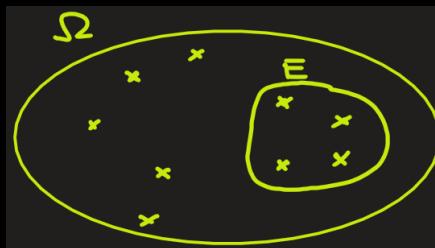
# Probability: Definition (Review)

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



Probability of an event E:

$$P(E) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$



Or,

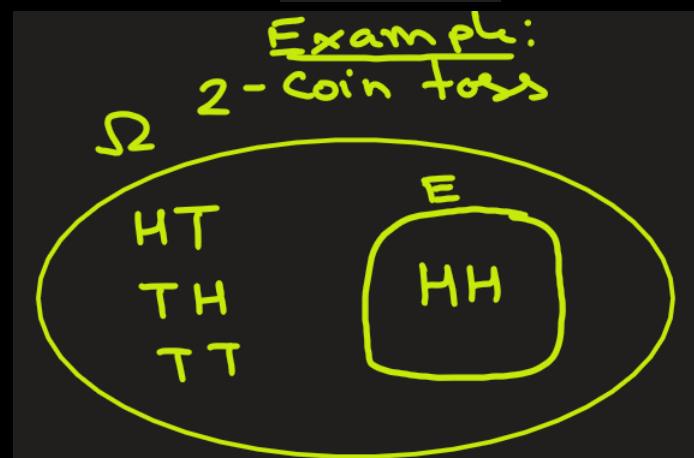
$$P(E) = \frac{|E|}{|\Omega|}$$

$$P(E) = \frac{|E|}{|\Omega|}$$

Here,

$|E|$  = number (size) of elements in event  $E$ ,

$|\Omega|$  = total number (size) of elements in sample space



Example: In 2-coin toss, what is the probability of getting **both heads**?

Ans: Here sample space is  $\Omega = \{HH, HT, TH, TT\}$ . Size is  $|\Omega| = 4$

The event E is both toss are heads:  $E = \{HH\}$ . Size is  $|E| = 1$

$$P(E) = 1/4.$$

# Conditional Probability,

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

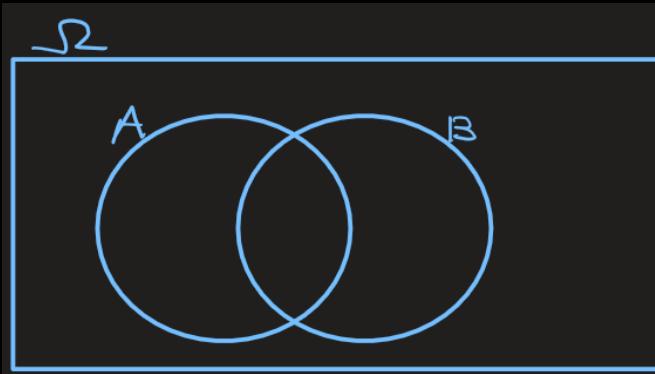


The **conditional probability** of an event **B**, given that another event **A** has already occurred, is the **probability that B happens under the condition that A is known to have occurred**.

It is written as  $P(B|A)$ .

It is defined mathematically as:

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



where,

$P(A \cap B)$ : Probability that both A and B occur together.

$P(A)$ : Probability that A occurs and it is  $P(A) > 0$ .

This formula tells us how to **update** our probability of B when we **know A has happened**.

Here A is also referred to as an **evidence**. Another name for  $P(B)$  is the **prior probability of B** and  $P(B|A)$  is the **posterior probability of B** ( prior means before updating based on the evidence A, and posterior means after updating based on the evidence)

# Conditional Probability

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

**Problem:** You roll a standard six-sided die

- 1) what is the probability of rolling a 5?
- 2) what is the probability of rolling a 5, **given that** the result was a number greater than 3?

Ans: Here  $\Omega = \{1, 2, 3, 4, 5, 6\}$

- 1) B is an event where you get 5:  $B = \{5\}$   
 $P(B) = 1/6$

- 2) Now we are given an additional **evidence** / information:  
the result was a number greater than 3.

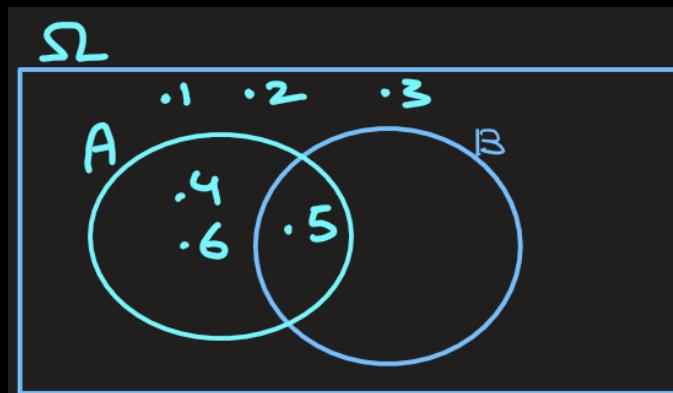
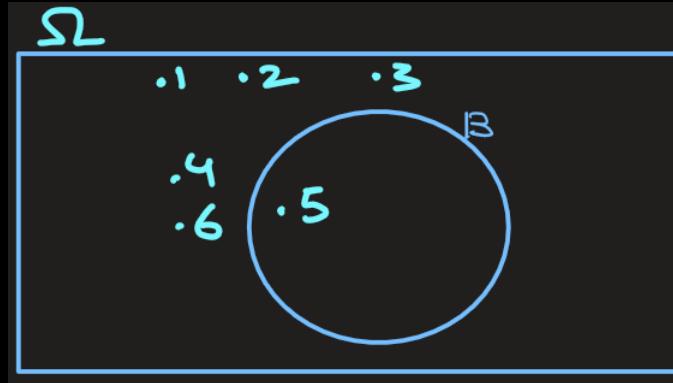
Let A is an event the result was a number greater than 3. We have to find  $P(B | A)$ .

Here     $A = \{4, 5, 6\}$   
 $P(A) = 3/6 = 1/2$

$A \cap B = \{5\}$   
 $P(A \cap B) = 1/6$

$$P(B | A) = (1/6) / (1/2) = 1/3$$

Conclusion: The original probability of rolling a 5 is  $1/6$ . Knowing the roll was greater than 3 changes the possible outcomes, and therefore the probability. The new probability increases to  $1/3$ .



# Conditional Probability

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

**Problem:** Two dice are rolled.

- 1) What is the probability at least one die shows a 5 ?
- 2) What is the probability at least one die shows a 5, **given that** the sum is 8 ?
- 3) What is the probability at least one die shows a 5, **given that** the sum is 15 ?

Ans: Here sample space  $\Omega$  is shown in the figure.

1) Let

B: at least one die shows a 5 .

$$\begin{aligned} B &= \{(5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ &\quad (6,5), (5,5), (4,5), (3,5), (2,5), (1,5)\} \end{aligned}$$

$$P(B) = 12/36 = 1/4 = 0.25 = 25\%$$

2) Now we are given an additional **evidence** / information: the sum is 8

Let A be the event : the sum is 8. We have to find  $P(B|A)$ .

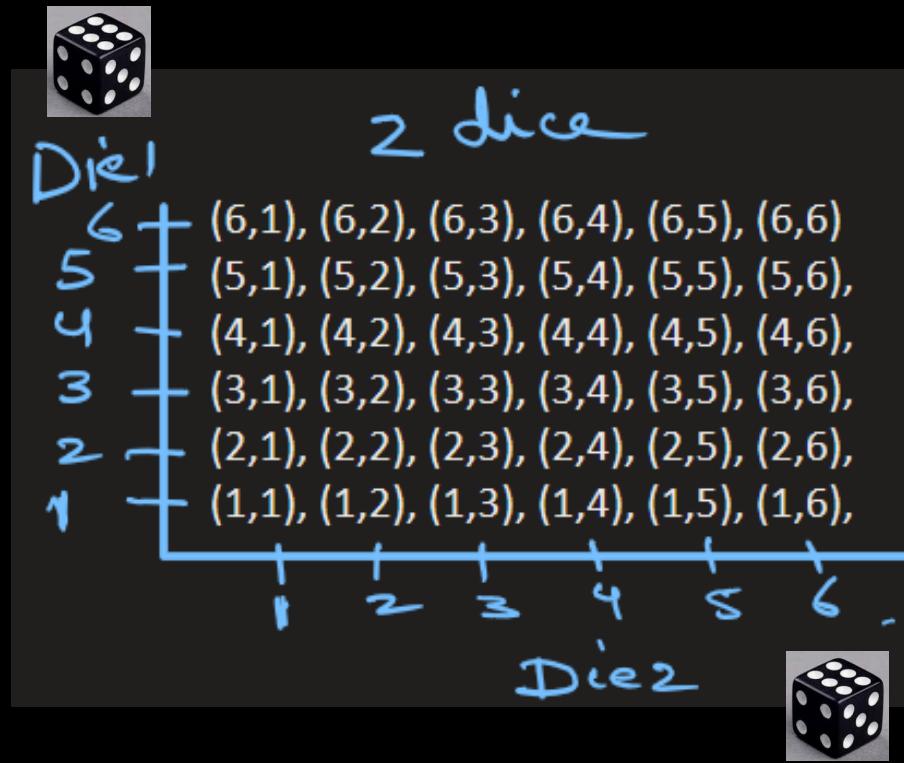
$$\begin{aligned} A &= \{(2,6),(3,5),(4,4),(5,3),(6,2)\} \\ P(A) &= 5/36 \end{aligned}$$

$$\begin{aligned} A \cap B &= \{(5,3), (3,5)\} \\ P(A \cap B) &= 2/36 \end{aligned}$$

$$P(B|A) = (2/36) / (5/36) = 2/5 = 40\%$$

Conclusion: The extra piece of information, i.e. evidence A, changed the probability of B. Here it increased.

- 3) Let A be the event : the sum is 15. This is not possible  $\rightarrow P(A) = 0$ . So, the probability  $P(B|A)$  is not defined.



# Conditional Probability

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

**Problem:** Two dice are rolled

- 1) What is the probability least one die shows a 6?
- 2) What is the probability least one die shows a 6, **given that** sum is even?

Ans:

1) Let  
B: least one die shows a 6

B = {(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)  
(1,6), (2,6), (3,6), (4,6), (5,6)}

P(B) = 11/36 = 30.5 %

2) Now we are given an additional information: sum is even  
Let A be the event : sum is even. We have to find P(B|A).

A = {18 elements in total}

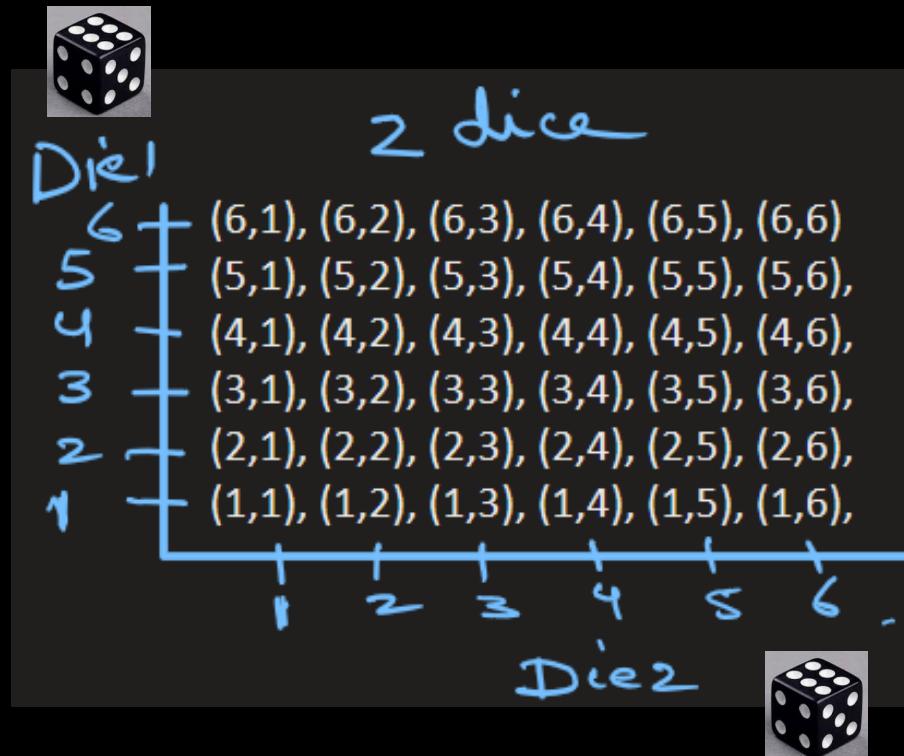
P(A) = 18/36

A  $\cap$  B = {(6,2), (6,4), (6,6), (2,6), (4,6)}

P(A  $\cap$  B) = 5/36

P(B|A) = (5/36) / (18/36) = 27.8 %

Conclusion: The extra piece of information , i.e. evidence A, changed the probability of B. Here it decreased.



# Conditional Probability

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

**Problem:** Imagine a **square dartboard** that goes from **0 to 1** along both the x-axis and y-axis.

Every dart throw lands **uniformly at random** on the board: the probability of landing in a region is **proportional to its area**.

- 1) What is the probability that it landed in the **bottom-left quarter** of the board (where  $x < 0.5$  and  $y < 0.5$ )?
- 2) What is the probability that it landed in the **bottom-left quarter** of the board (where  $x < 0.5$  and  $y < 0.5$ ), given that the dart landed **below the diagonal line**  $y = x$  ?

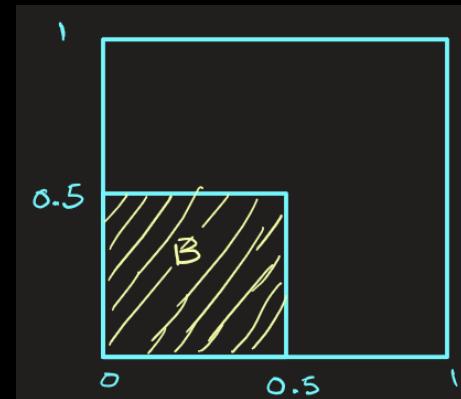
Ans: 1) Let B be the event that dart lands in the **bottom-left quarter**.

Since probability is proportional to area, so

$$P(B) = \frac{\text{Area of bottom-left quarter}}{\text{Total area}} = \frac{0.5 \times 0.5}{1} = 0.25$$

2) Now we are given an additional information: dart landed below diagonal line,  $y = x$

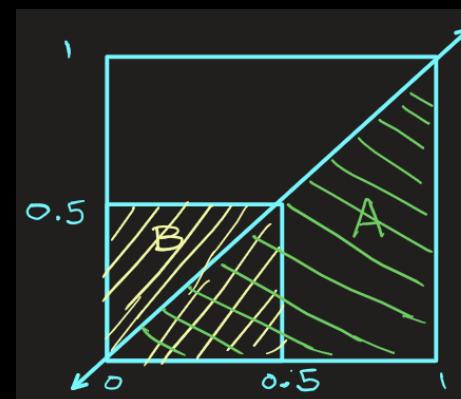
Let A be the event that dart landed below diagonal line  $y = x$ . We have to find  $P(B|A)$ .



$$P(A \cap B) = \frac{\text{Area of } A \cap B}{\text{Total area}} = \frac{0.5 \times 0.5 \times 0.5}{1} = 0.125$$

$$P(A) = \frac{\text{Area between diagonal and bottom}}{\text{Total area}} = \frac{0.5}{1} = 0.5$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.125}{0.5} = 0.25$$



Conclusion: The extra piece of information, i.e. evidence A, did not change the probability of B.

# EXTRA

Extra problems





# Conditional Probability

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



**Problem:** Conditional Probability in a Markov (2-gram) Language Model

Suppose you are developing a **Markov-based language model** for predicting the next word in a sentence.

You have following corpus:

The **big dog** barked loudly.

The **big cat** chased the mouse.

A **big dog** ran across the yard.

The small dog slept peacefully.

The tree sailed across the river.

The **big tree** was by the river.

The dog looked at the tree.

A small tree carried a **big dog**.

The small dog was barking at **big tree**.

The **big tree** had juicy fruits.

- 1) What is the probability that next word is “dog”, given that previous word is “big” ?, i.e.  $P(\text{“big dog”}) = ?$
- 2) What is the probability that next word is “tree”, given that previous word is “big” ?, i.e.  $P(\text{“big tree”}) = ?$

Side Note: These types of problems are solved by AI in sentence completion → “**big ?**”



# Conditional Probability



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

Ans:

1) Let

**Event B:** The next word is “**dog**.”

**Event A:** The previous word is “**big**.”

We have to find  $P(B|A)$ , i.e.  $P(\text{"big dog"})$ .

Total bigrams starting with “**big**”  $\rightarrow 7$

{ big dog, big cat, big dog, big boat, big dog, big boat, big house }

Count of (“**big dog**”) = 3

In a first-order Markov model, the probability of one word depends only on the previous word. Hence, the model estimates:

$$P(A \cap B) = \frac{\text{count of ("big dog")}}{\text{total bigrams}} = \frac{3}{\text{total bigrams}}$$

$$P(A) = \frac{\text{count of bigrams ("big -")}}{\text{total bigrams}} = \frac{7}{\text{total bigrams}}$$

$$P(\text{dog}|\text{big}) = \frac{(3/\text{total bigrams})}{(7/\text{total bigram})} = \frac{3}{7} = 0.43$$



Ans:

2) Let

**Event B:** The next word is “**tree**.”

**Event A:** The previous word is “**big**.”

We have to find  $P(B|A)$ , i.e.  $P(\text{“big tree”})$ .

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



Total bigrams starting with “**big**”  $\rightarrow 7$

{ big dog, big cat, big dog, big tree, big dog, big tree, big house }

Count of (“**big tree**”) = 2

In a first-order Markov model, the probability of one word depends only on the previous word. Hence, the model estimates:

$$P(A \cap B) = \frac{\text{count of (“big tree”)}}{\text{Total bigrams}} = \frac{2}{7}$$

$$P(A) = \frac{\text{Count of bigrams (“big —”)}}{\text{total bigrams}} = \frac{7}{7}$$

$$P(\text{tree}|\text{big}) = \frac{(2/\text{total bigrams})}{(7/\text{total bigram})} = \frac{2}{7} = 0.2857$$

**Conclusion:**

$$P(\text{dog}|\text{big}) = 0.43 , P(\text{tree}|\text{big}) = 0.2857$$

The next word should be dog because of higher probability


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

# Conditional Probability

**Problem:** Medical Diagnosis (Bayesian Networks)

Suppose you're building a disease prediction model.

Based on past data we have,

- 0.95% of the population both have the disease and test positive
- 5% of all tested people test positive

If a patient tests positive, what is the chance they really have the disease ?

Ans: Let

B: Patient has the disease

A: Patient has a positive test result

$P(A \cap B) = 0.0095$  (0.95% of the population both have the disease and test positive)

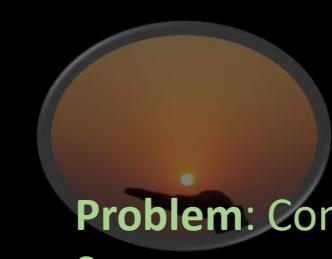
$P(A) = 0.05$  (5% of all tested people test positive)

Then,

$$P(B|A) = P(A \cap B) / P(A) = 0.0095 / 0.05 = 0.19$$

So, if a patient tests positive, there's a 19% chance they really have the disease.

This is the Bayesian update idea in medical AI.



# Conditional Probability



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

**Problem:** Conditional Probability in a Markov (N-gram) Language Model

Suppose you are developing a **Markov-based language model** for predicting the next word in a sentence.

We know from corpus statistics:

2% of all word pairs are “big dog” →  $P(A \cap B) = 0.02$

5% of all word pairs start with “big” →  $P(A) = 0.05$

What is the probability that next word is “dog”, given that previous word is “big” ?

Ans: Let

**Event B:** The next word is “**dog**.”

**Event A:** The previous word is “**big**.”

In a first-order Markov model, the probability of one word depends only on the previous word. Hence, the model estimates:

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

$$P(\text{"dog"} | \text{"big"}) = \frac{P(\text{"big dog"})}{P(\text{"big"})} = \frac{0.02}{0.05} = 0.4$$

There is a **40% probability** that the word “**dog**” follows “**big**.” It predicts probability for all words that follows big, and then it assigns the word that has highest probability.

# Application of probability in AI ML (p1)



## Probability in Clustering — Gaussian Mixture Models (GMM)

### Example:

When clustering customers, a GMM says:

- Each cluster is a **probability distribution** (usually Gaussian).
- Each point belongs to each cluster with some probability.

So, customer A might belong:

- 70% to Cluster 1 (young students)
- 30% to Cluster 2 (young professionals)

Unlike K-Means, GMM gives **soft assignments**.

# Application of probability in AI ML (p2)

In RL, like training robots, the agent's policy is

$$\pi(a|s) = \text{Prob}(\text{take action } a \mid \text{in state } s)$$

The policy samples actions probabilistically-  
helping agents to explore



3) Reinforcement learning

agents policy

$$\pi(a|s) = P(\text{take action } a \mid \text{in state } s)$$

$$a = \{\text{F}, \text{B}, \text{L}, \text{R}\}$$
$$\frac{N}{N}, S, E, W$$

$$\pi(N|s=17) = 0.6$$

$$\pi(S|s=17) = 0.1$$

$$\pi(E|s=17) = 0.1$$

$$\pi(W|s=17) = 0.2$$

