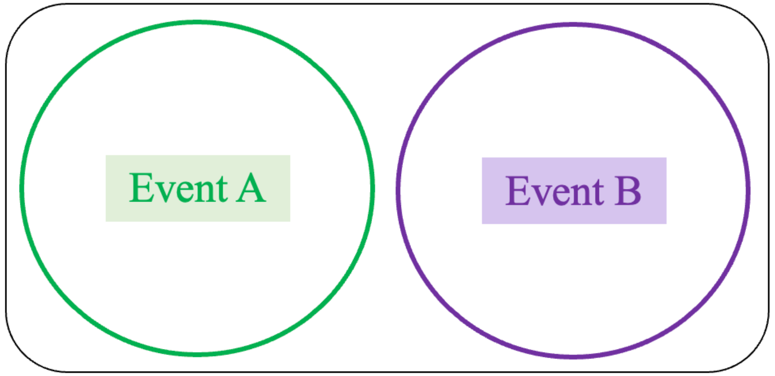
Basic Probability

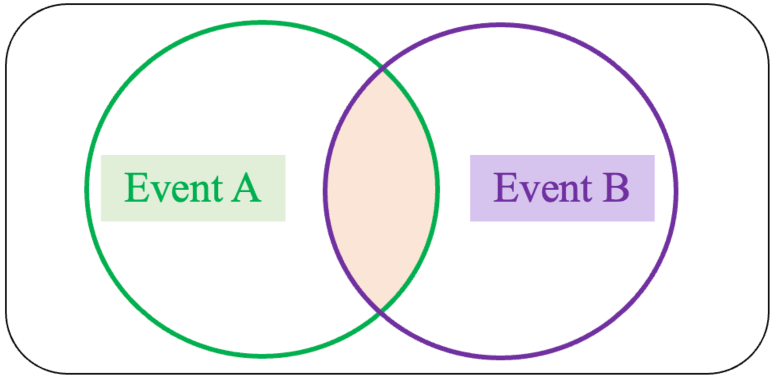
Quang-Vinh Dinh
Ph.D. in Computer Science

Objectives

Events

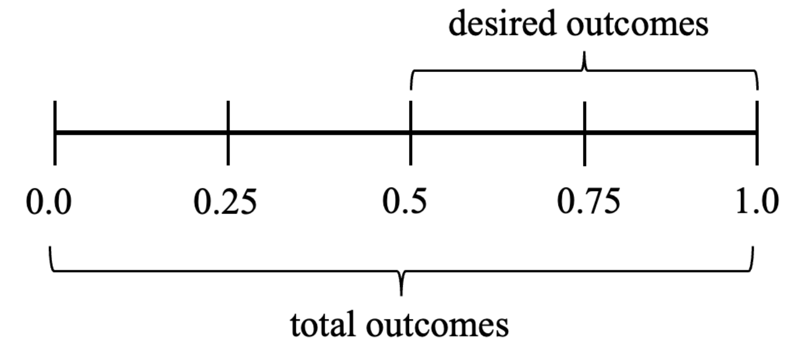


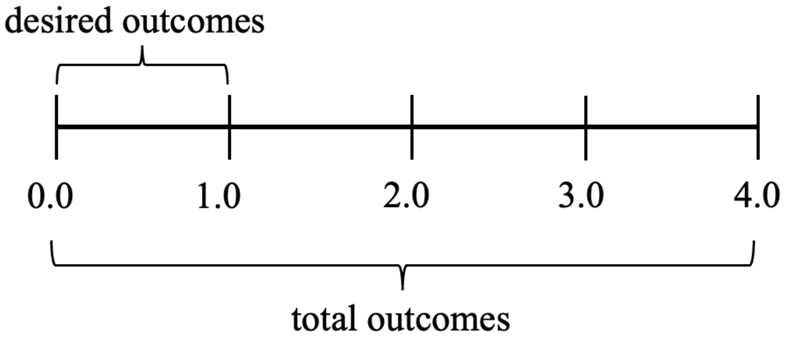
Mutually exclusive



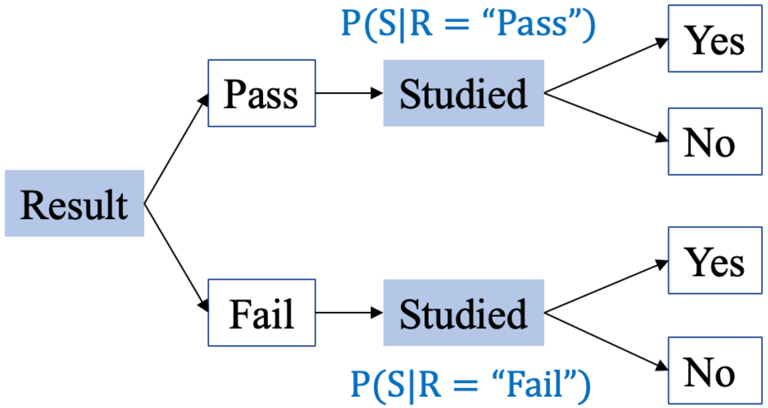
Non-mutually exclusive

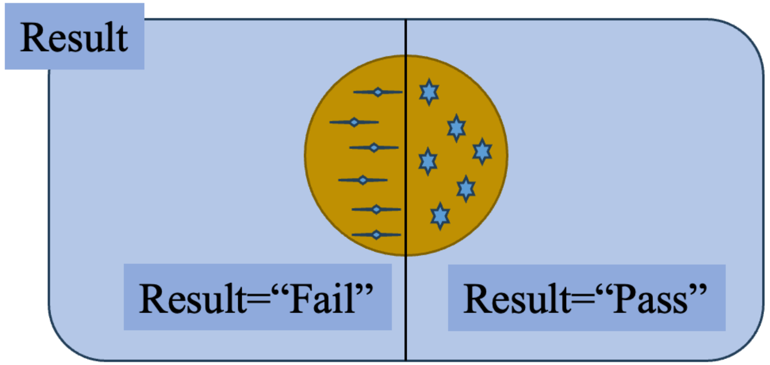
Probability





Bayes’ Theorem





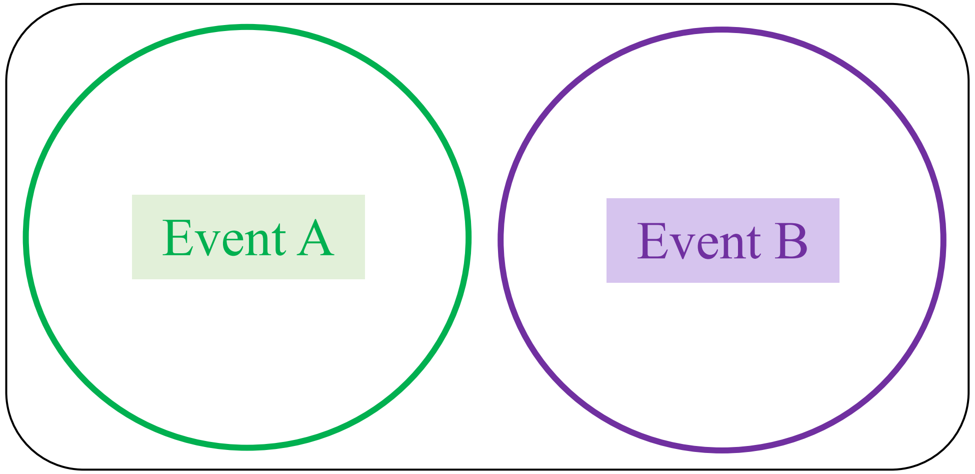
Outline

SECTION 1
Events

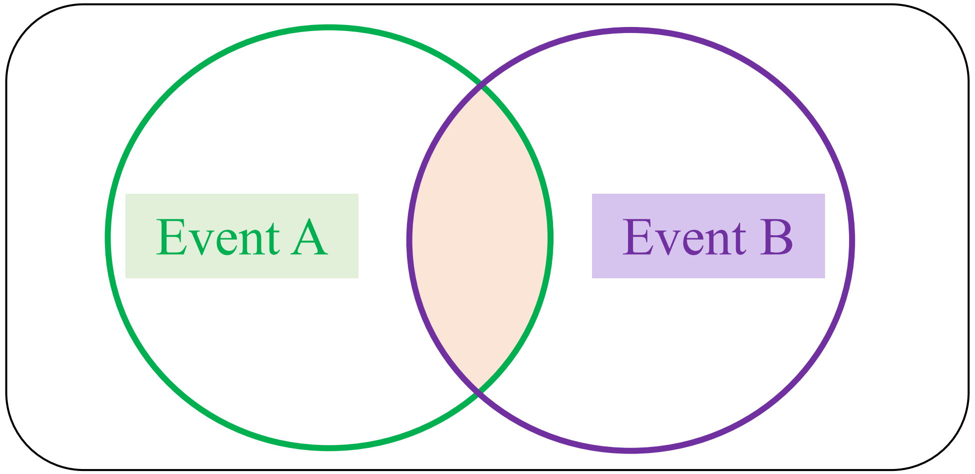
SECTION 2
Probability

SECTION 3
Bayes' Theorem

SECTION 4
Simple Classification



Mutually exclusive



Non-mutually exclusive

Basic Probability

❖ **Some concepts**

Toss a coin

Sample space: S = {heads, tails}

Experiment: implementation of set of basic conditions for observing a certain phenomenon

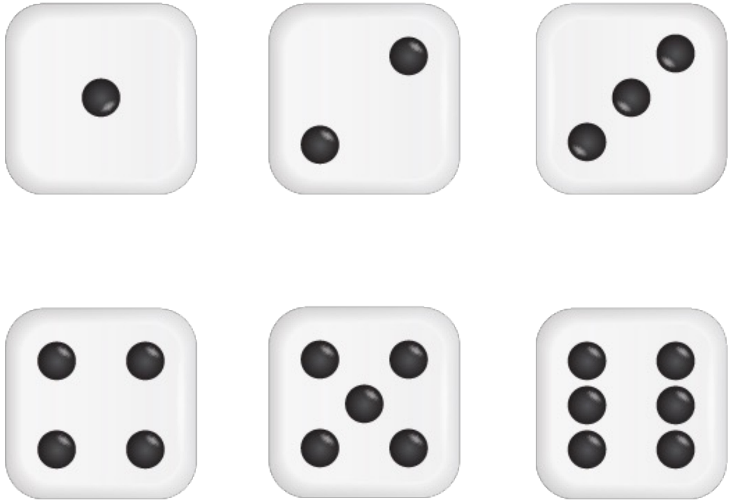
An outcome is a result of an experiment

The set of all possible outcomes is called the sample space

An event is a subset of the sample space

Roll a dice

Sample space: S = {1, 2, 3, 4, 5, 6}



❖ Event

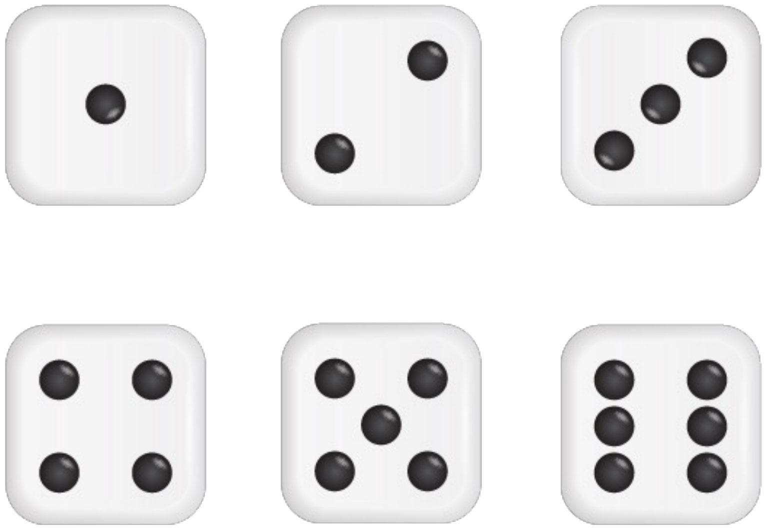
Certain event: An event that always occurs in an experiment, denoted by Ω

Impossible event: An event that never occurs when the experiment is executed, denoted by $\emptyset$.

Random event: An event that may or may not occur when performing the experiment

Random Experiment: An experiment whose outcomes are random events

For convenience, events are usually denoted with capital letters: A, B, C, . . .



Roll a dice:

Ω = “dots ≤ 6 and ≥ 1” is a certain event

$\emptyset$ = “7-dot” is an impossible event

A = “even-dot” is a random event

Experiment and Event

❖ Example

- A family with 2 children. Events:
  - A = "A family has 1 boy and 1 girl"
  - B = "A family has 3 children"
  - C = "A family has 2 children"

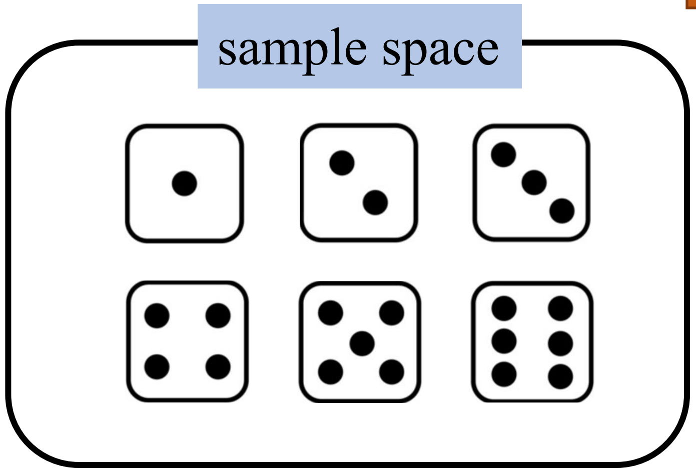
  Which event is certain, random, or impossible event?

- A box contains 8 balls: 6 blue and 2 red. Pick randomly 3 balls:
  - A = "get 3 blue balls"
  - B = "get 3 red balls"
  - C = "get 3 balls"

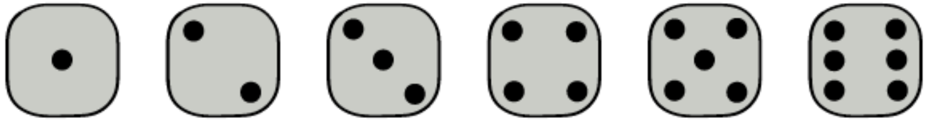
  Which event is certain, random, or impossible event?

Operations with Events

❖ Intersection of events



- In the experiment of rolling a single dice

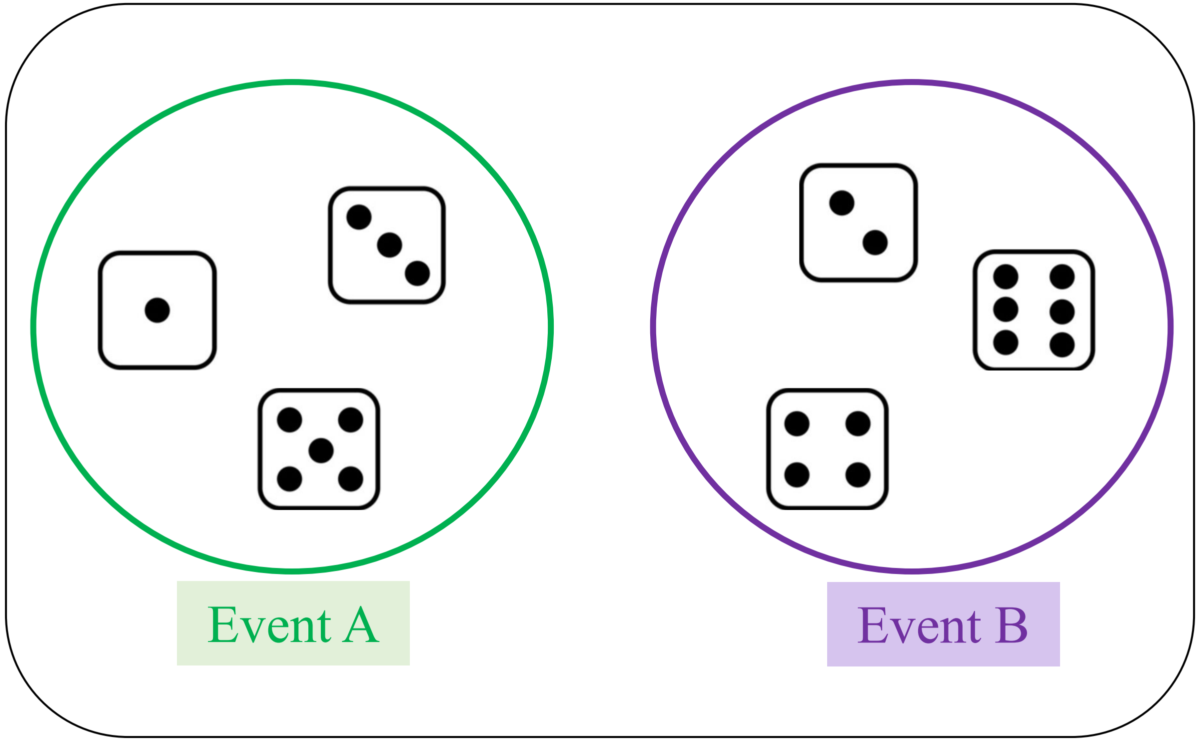


- Event A: “the number rolled is odd”

  => A = {1, 3, 5}

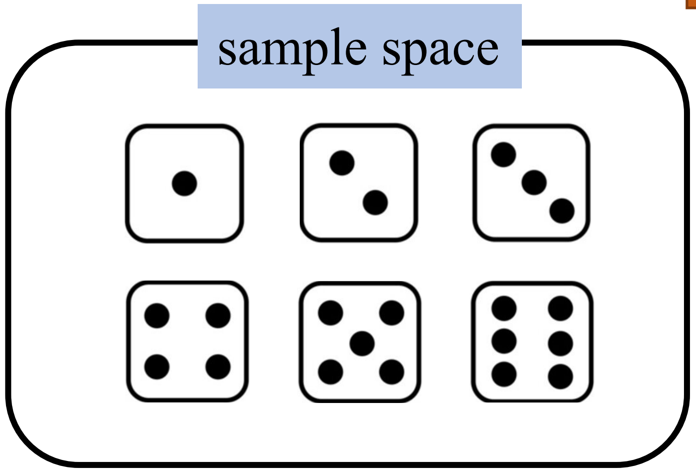
- Event B: “the number rolled is divisible by 2”

  => B = {2, 4, 6}

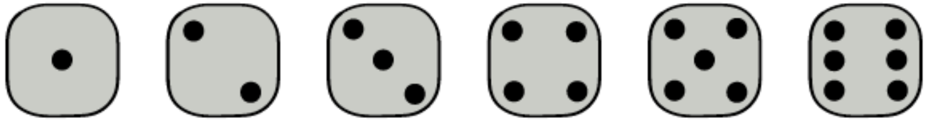


Operations with Events

❖ Intersection of events



- In the experiment of rolling a single dice

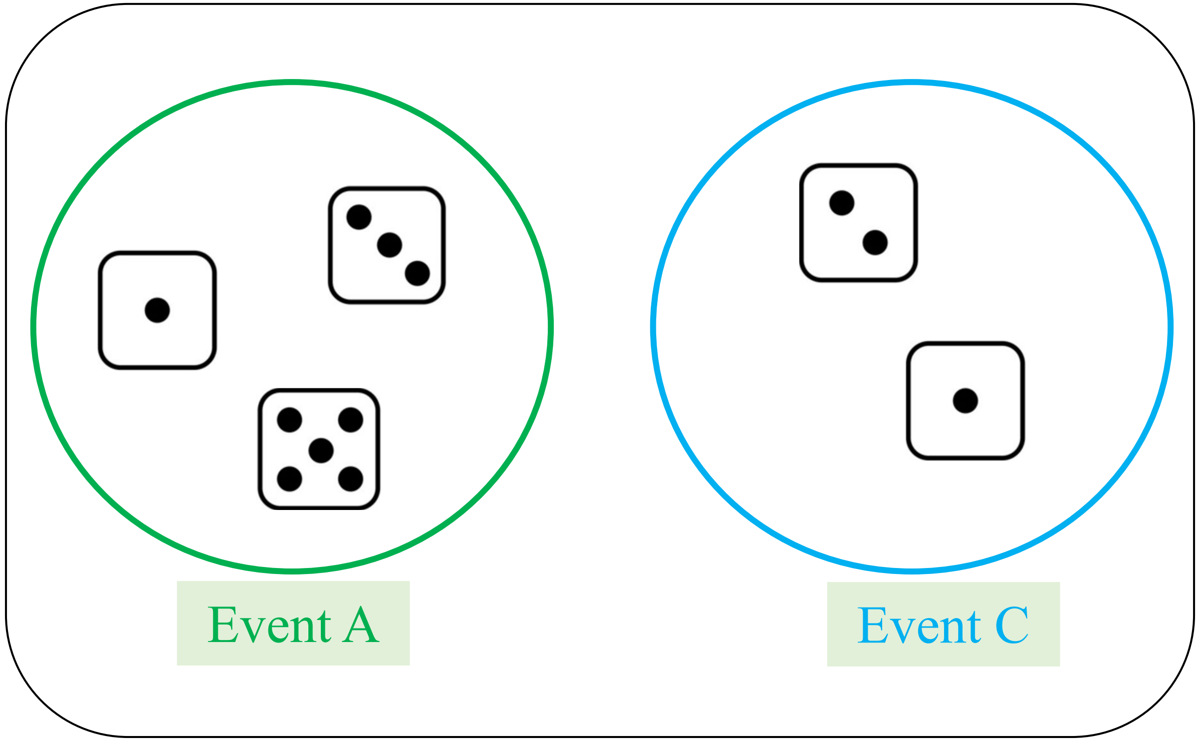


- Event A: “the number rolled is odd”

  => A = {1, 3, 5}

- Event C: “the number rolled is smaller than 3”

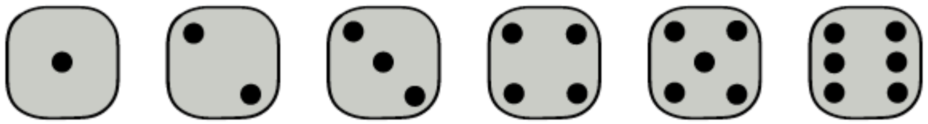
  => C = {1, 2}

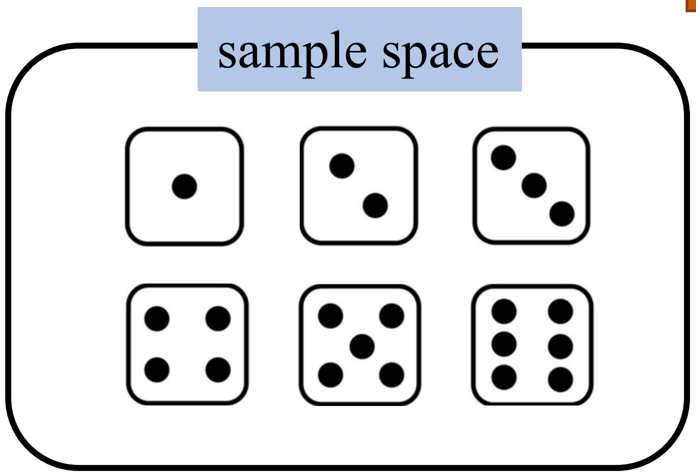


Operations with Events

❖ Intersection of events

- In the experiment of rolling a single dice





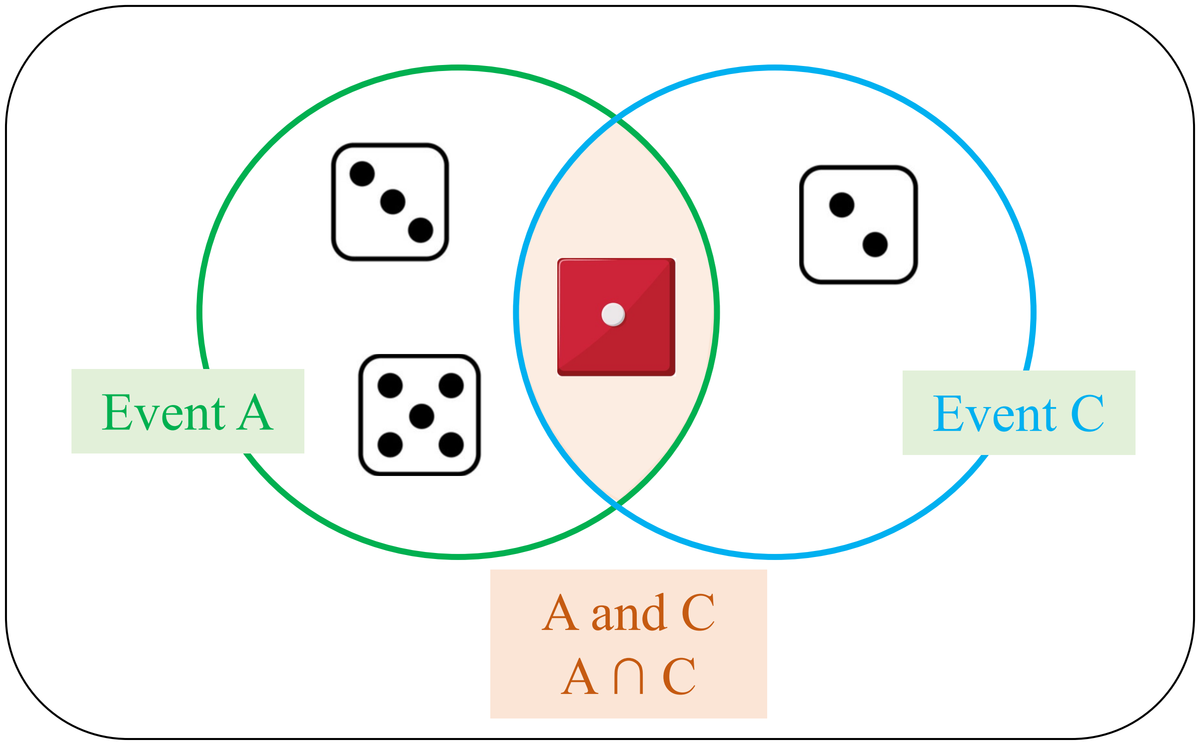
- Event A: “the number rolled is odd”

  => A = {1, 3, 5}

- Event C: “the number rolled is smaller than 3”

  => C = {1, 2}

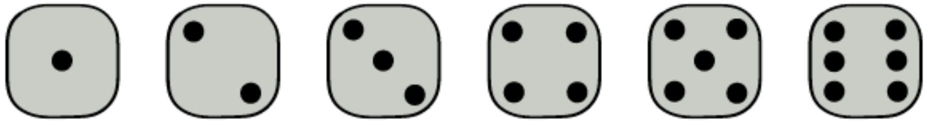
- $A \cap C = \{1\}$



Operations with Events

❖ Intersection of events: Quiz

- In the experiment of rolling a single dice



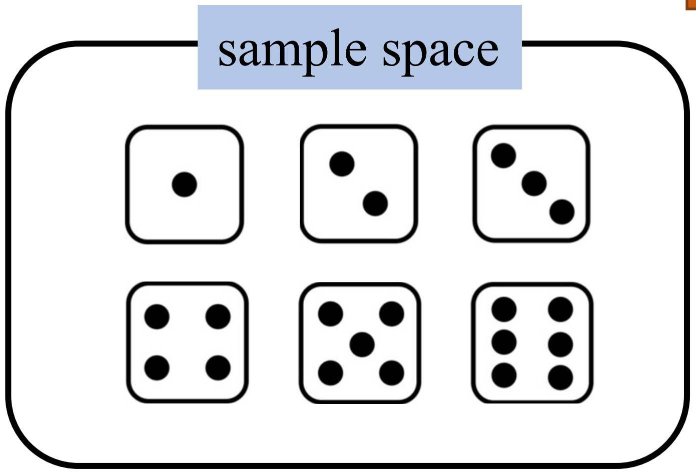
- Event A: "the number rolled is odd"

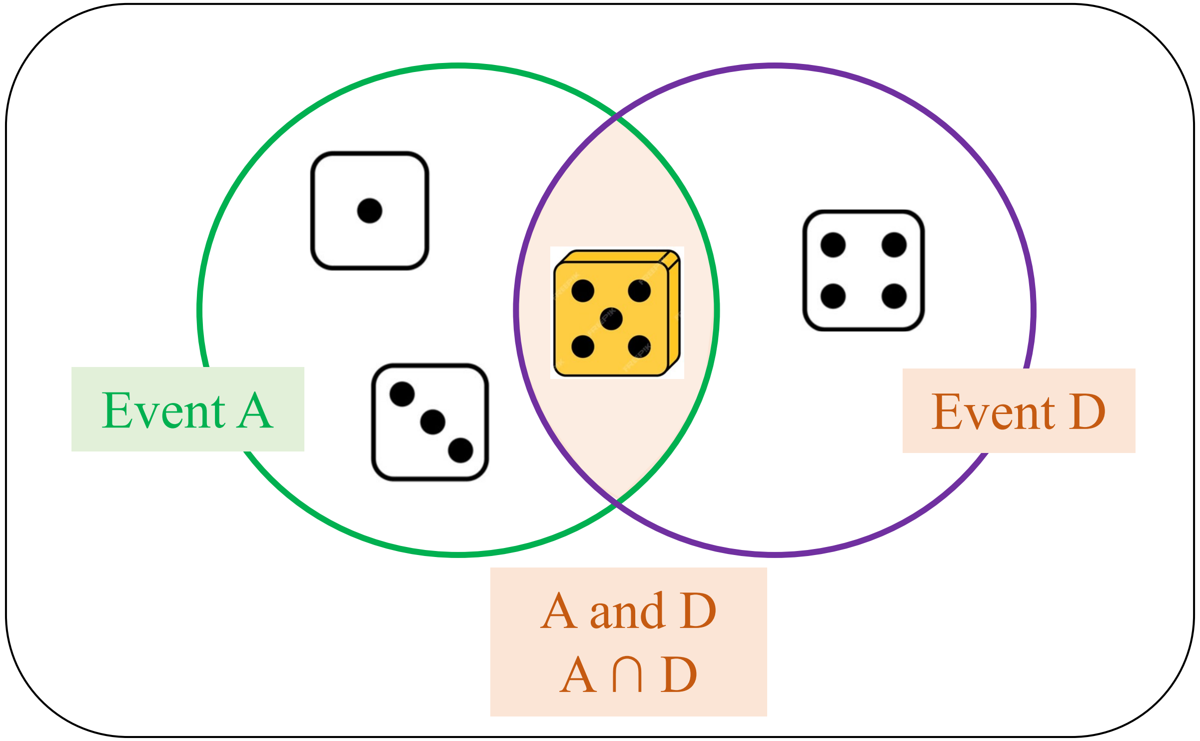
  => A = {1, 3, 5}

- Event D: "the number rolled is greater than 3"

  => D = {4, 5}

- $A \cap D = \{?\}$

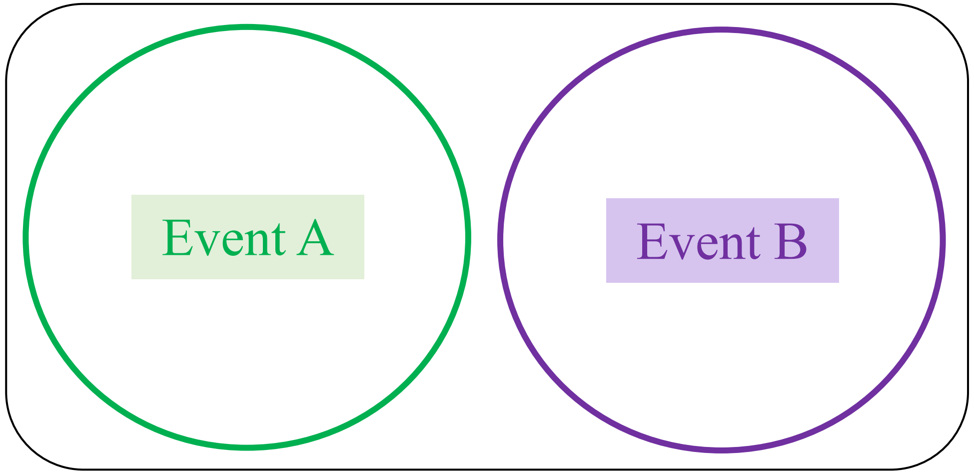




Operations with Events

❖ Mutually exclusive event

- ✓ Events A and B are mutually exclusive (cannot both occur at once) if they have no elements in common.
- ✓ For A and B to have no outcomes in common means precisely that it is impossible for both A and B to occur on a single trial of the random experiment.
- ✓ $A \cap B = \emptyset$



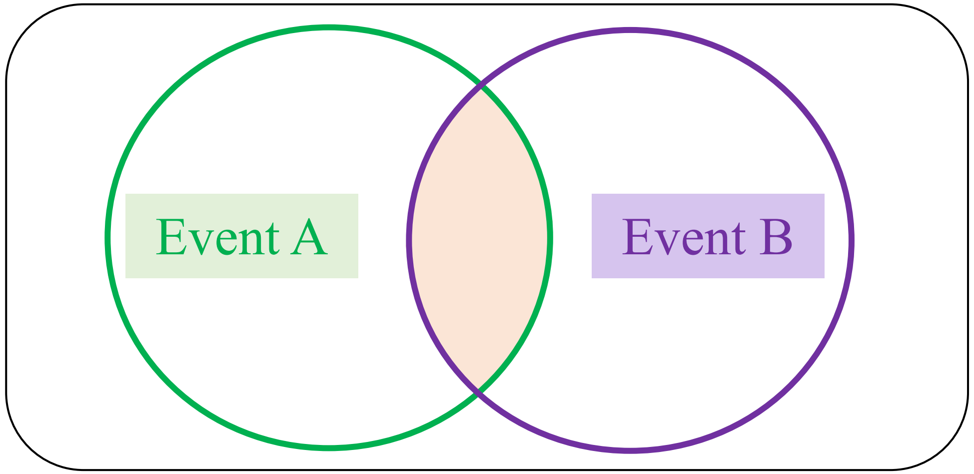
Mutually exclusive

Example

Event A: "the number rolled is even"
=> $A = \{2, 4, 6\}$

Event B: "the number rolled is odd"
=> $B = \{1, 3, 5\}$

$A \cap B = \emptyset$

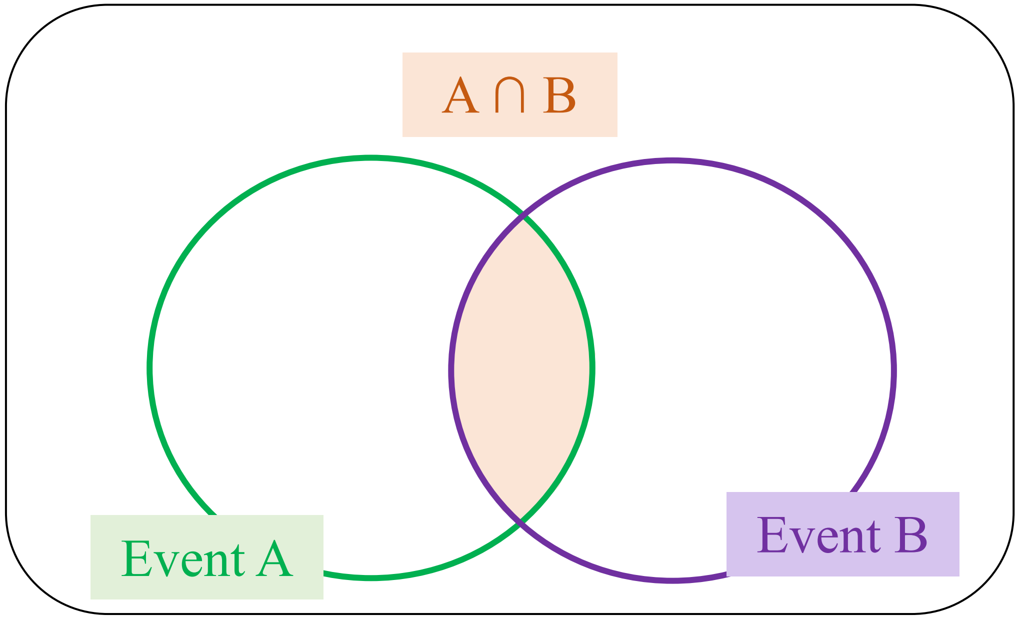


Non-mutually exclusive

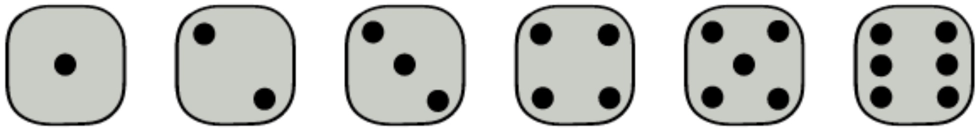
Operations with Events

❖ Union of events

- ✓ The union of events A and B, denoted A ∪ B
- ✓ The collection of all outcomes that are elements of one or the other of the sets A and B, or of both.



- ➢ In the experiment of rolling a single dice

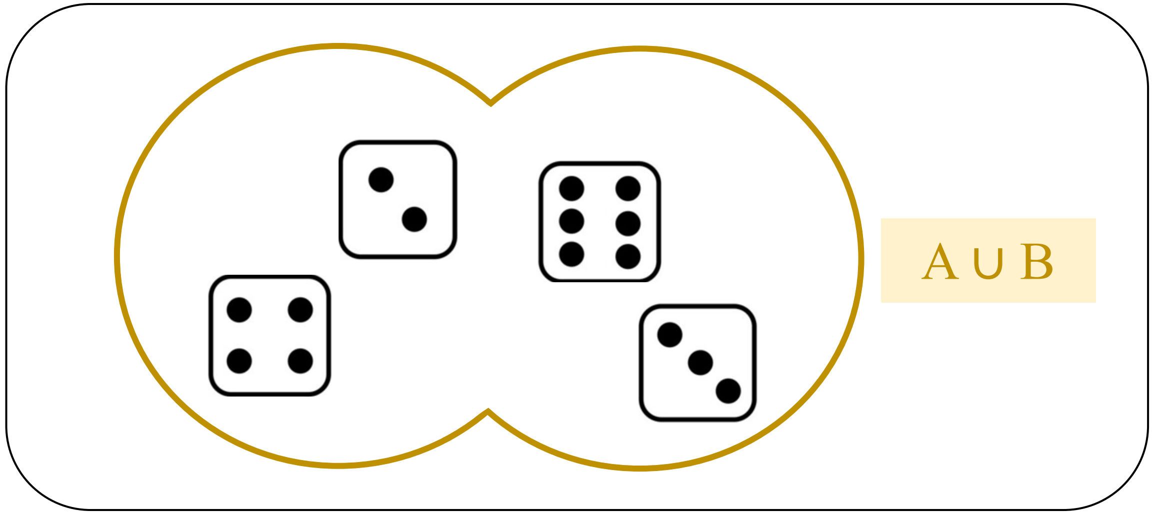


- ➢ Event A: "the number rolled is even"

  => A = {2, 4, 6}
- ➢ Event B: "the number rolled is divisible by 3"

  => B = {3, 6 }
- ➢ The union of A and B

  => A ∪ B = {2, 3, 4, 6}

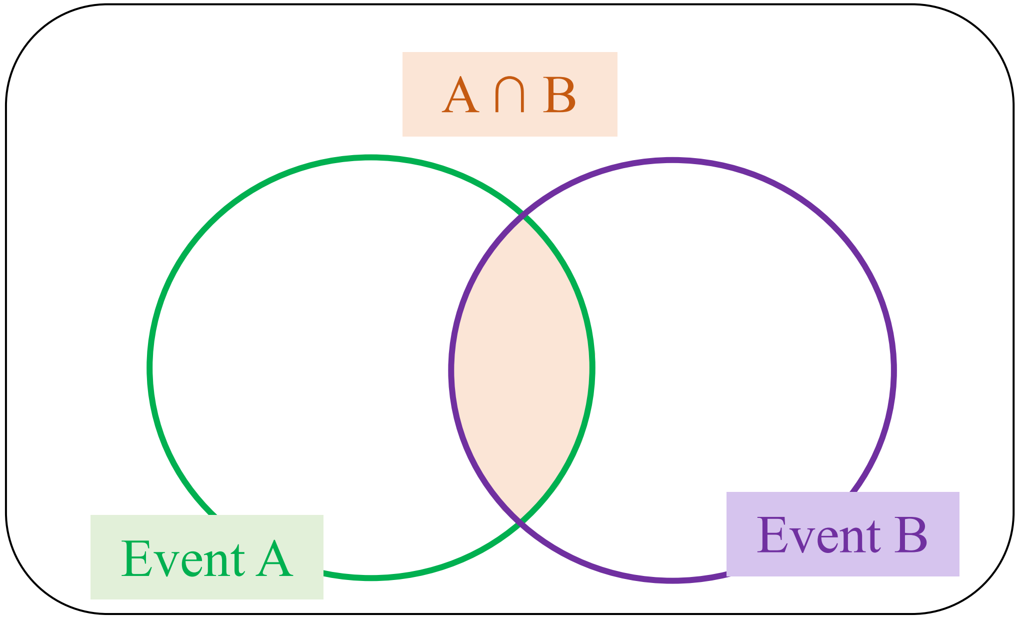


Operations with Events

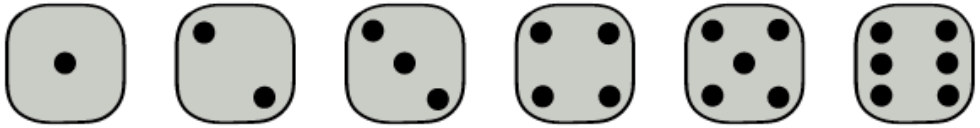
❖ Union of events

✓ The union of events A and B, denoted $A \cup B$

✓ The collection of all outcomes that are elements of one or the other of the sets A and B, or of both.



➢ In the experiment of rolling a single dice



➢ Event A: "the number rolled is odd"

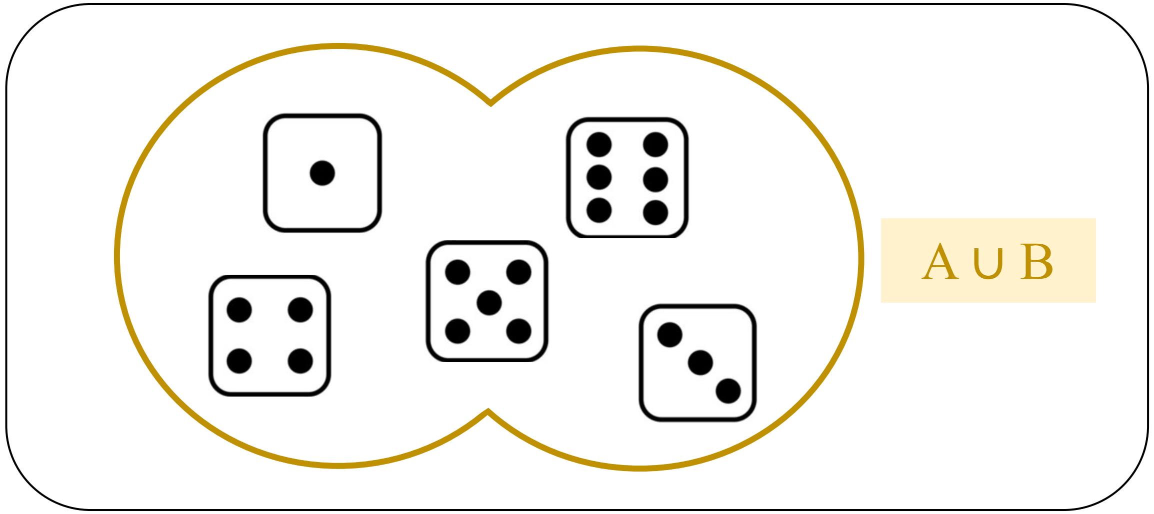
=> $A = \{1, 3, 5\}$

➢ Event B: "the number rolled is greater than 3"

=> $B = \{4, 5, 6\}$

➢ The union of A and B

=> $A \cup B = \{1, 3, 4, 5, 6\}$



Operations with Events

❖ Complements

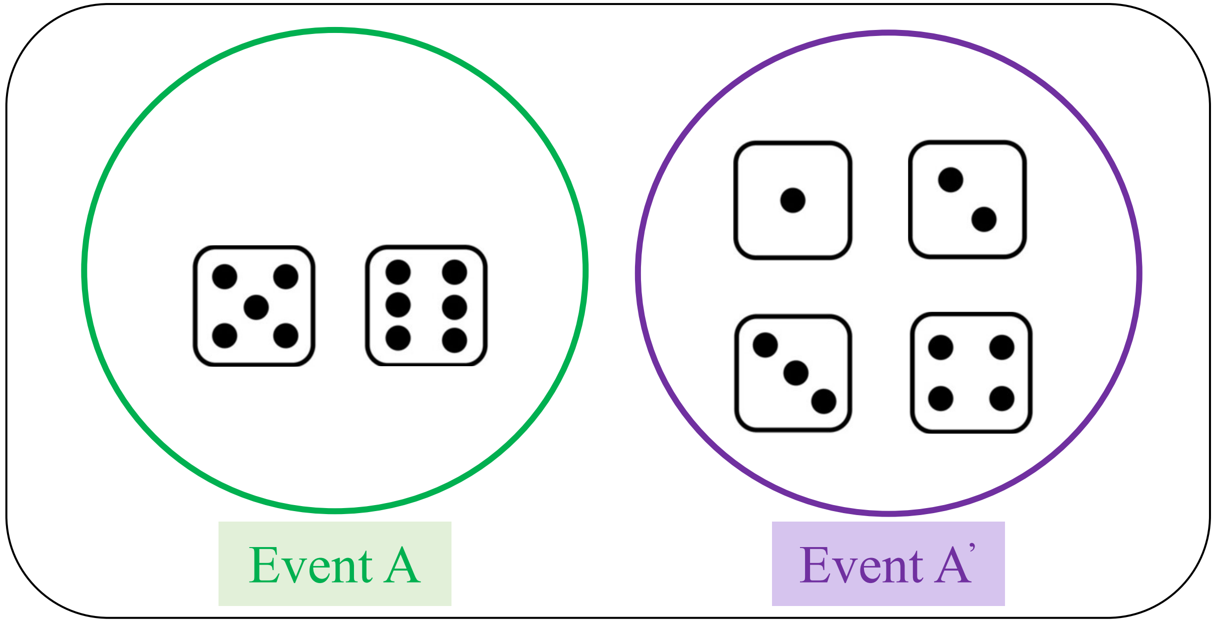
- ✓ The complement of an event A in a sample space Ω, denoted A' (or A^c or $\overline{A}$)
- ✓ The collection of all outcomes in Ω that are not elements of the set A
- ✓ $A' + A = \Omega$

Example:

A: "the number rolled is greater than 4."

=> A = {5, 6}

=> A' = {1, 2, 3, 4}



Rules of probability

❖ Probability for complements

For any event A

$$P(A^c) = 1 – P(A)$$

$$P(A) = 1 – P(A^c)$$

Example

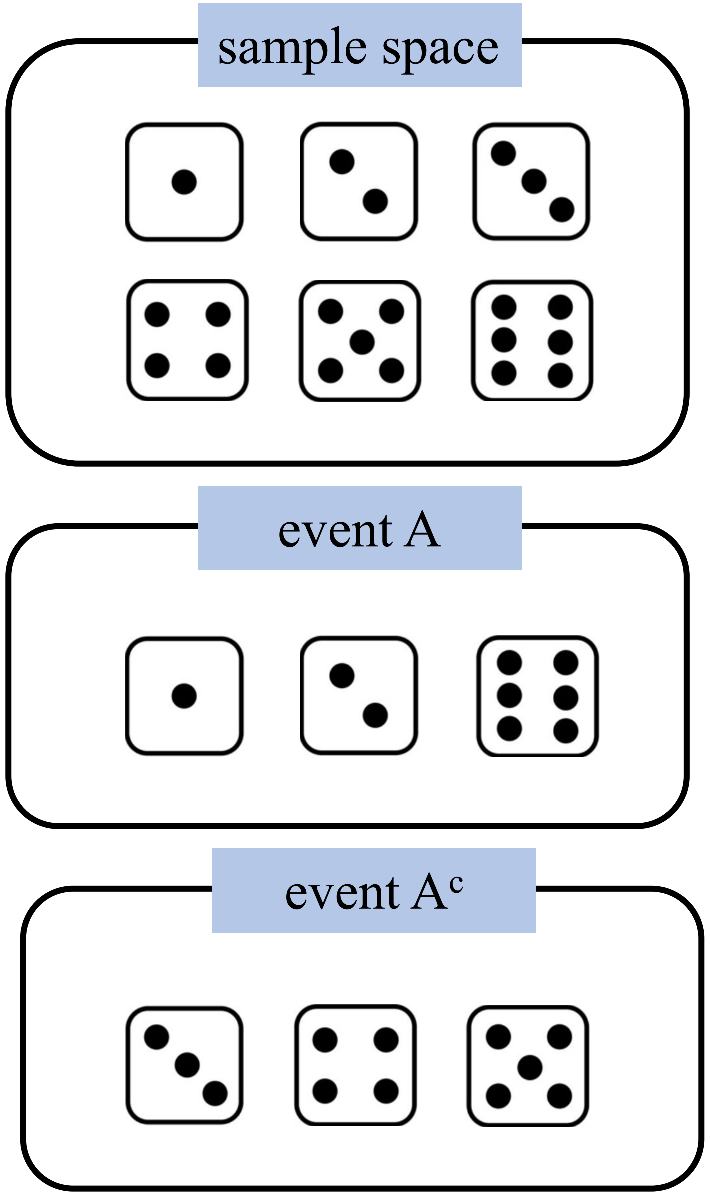
Find the probability that when we roll a dice, we get a number different from 1, 2, and 6?

Let's A: "Getting the number 1, 2, and 6" => A = {1, 2, 6}

"Getting a number different from 1, 2, and 6" = A^c

Since $P(A) = P(1) + P(2) + P(6) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2$

$P\ A^c) = 1 – P(A) = 1 – 1/2 = 1/2$



Operations with Events

❖ Quizzes

In the experiment of rolling a single dice

A: "the number rolled is even"

B: "the number rolled is greater than or equal 3"

C: "the number rolled is smaller than 2"

1. Event A' is

a) {} b) {1, 3, 5} c) {1, 3} d) {2, 4, 6} e) Khác

2. Event A.B is

a) {5, 7} b) {4, 6} c) {1, 3} d) {1, 3, 5} e) Khác

3. Event B + C is

a) {1, 2, 3, 4} b) {1, 2, 3, 5} c) {1, 2, 3, 6} d) {1, 2, 3} e) Khác

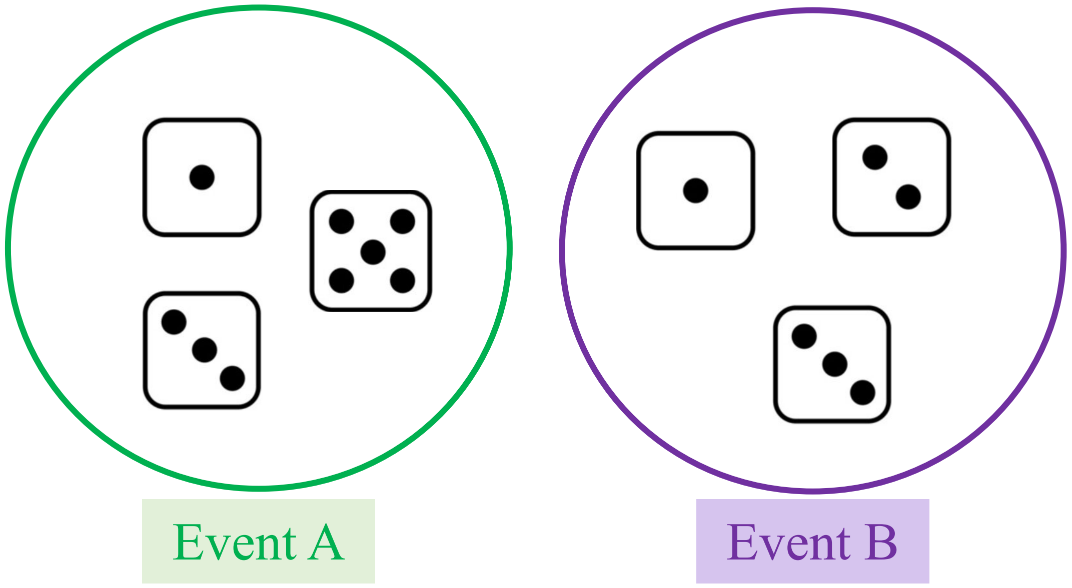
Operations with Events

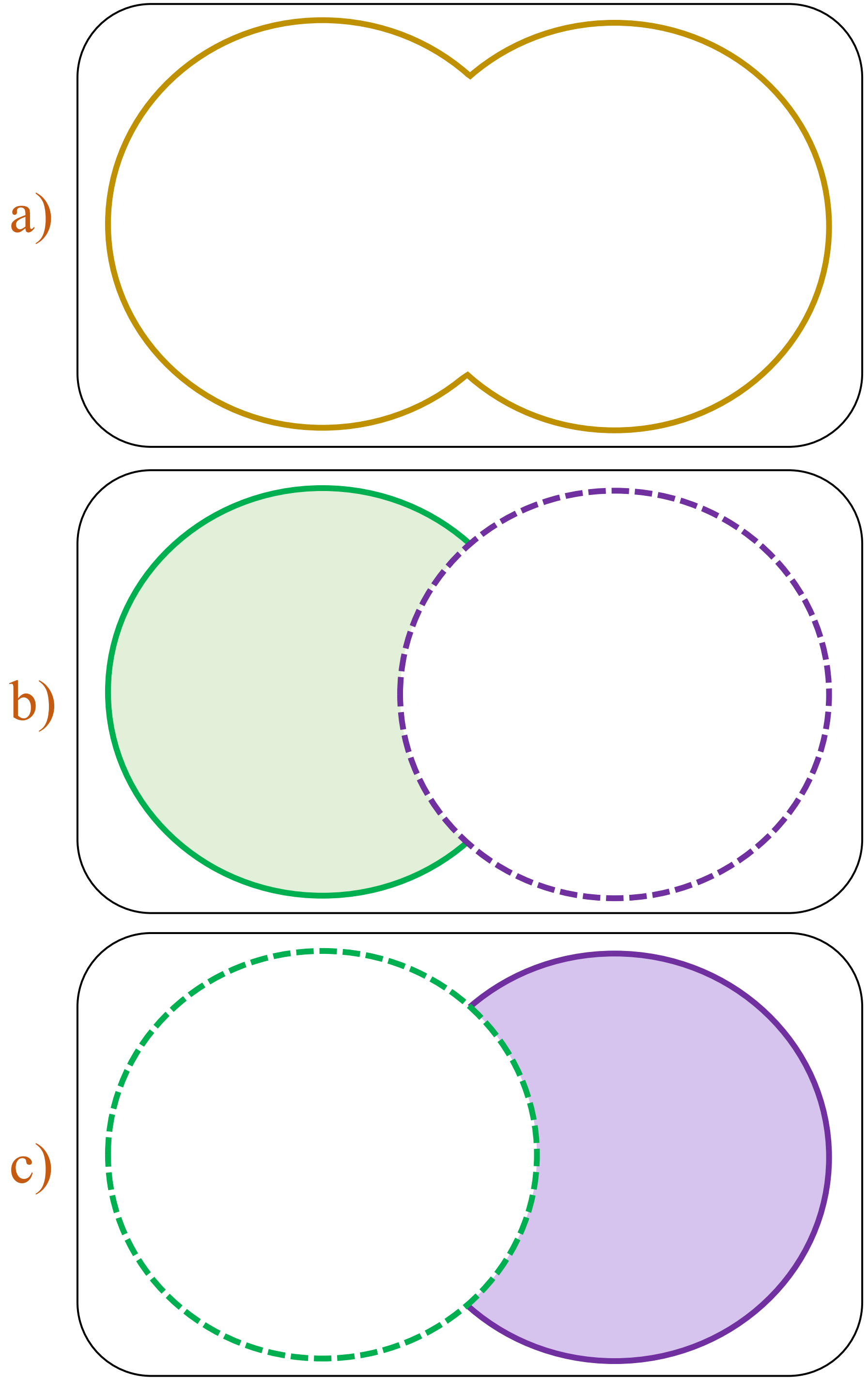
❖ **Quizzes**

In the experiment of rolling a single dice

A: “the number rolled is odd”

B: “the number rolled is less than or equal 3”





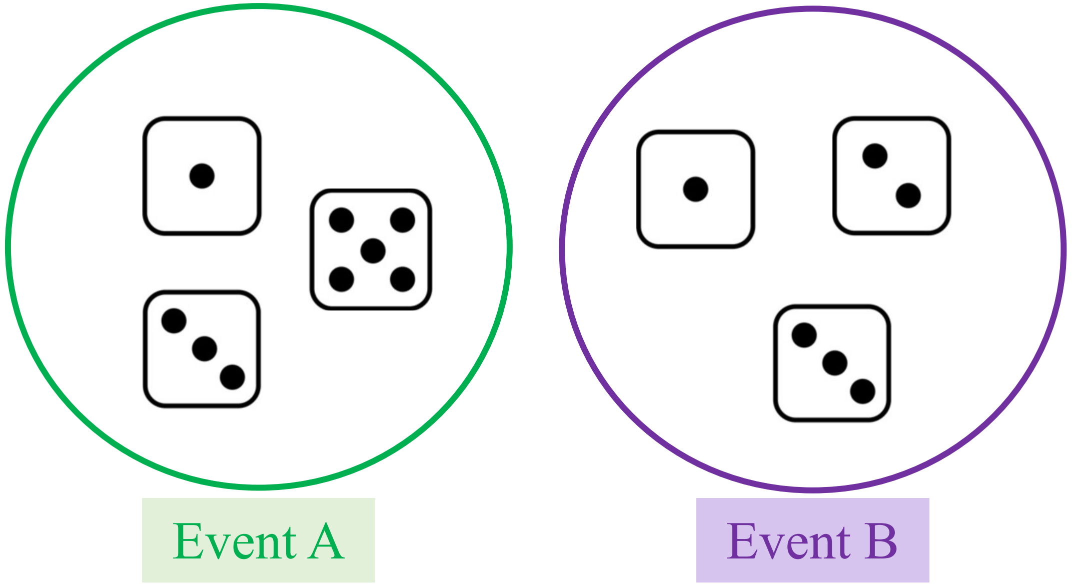
Operations with Events

❖ Quizzes

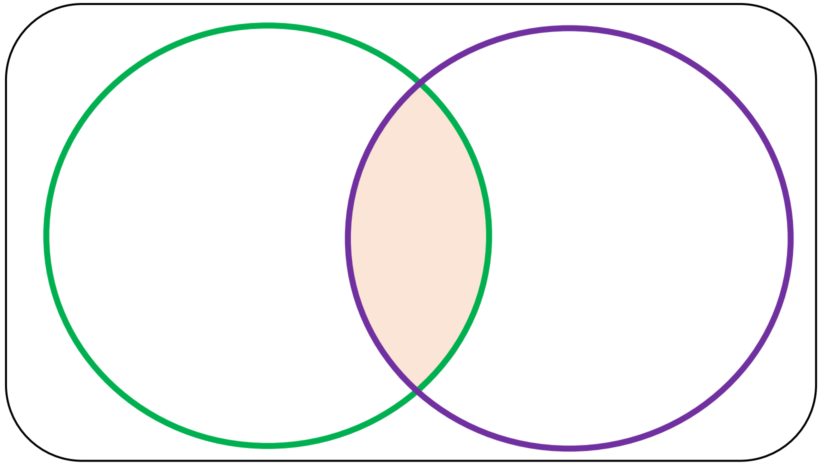
In the experiment of rolling a single dice

A: “the number rolled is odd”

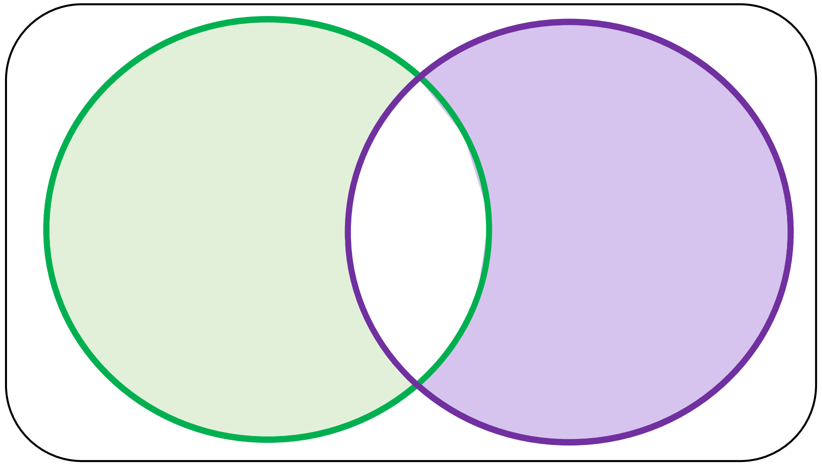
B: “the number rolled is less than or equal 3”



d)



e)



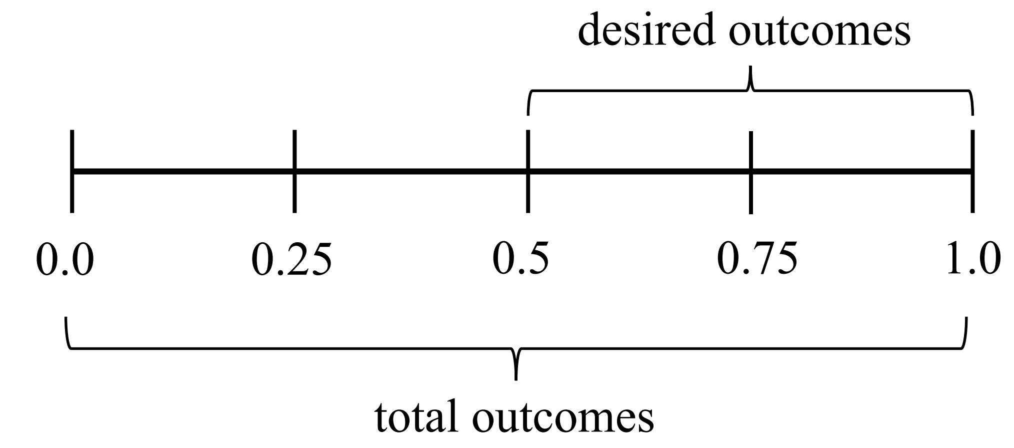
Outline

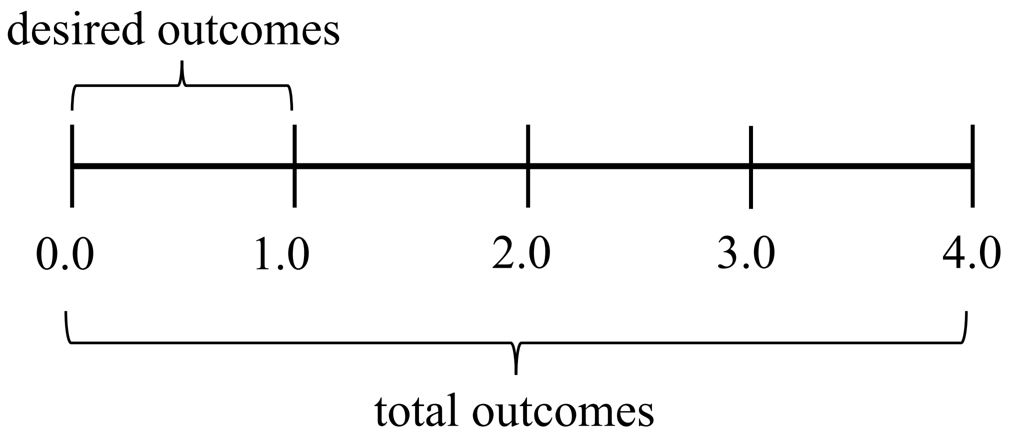
SECTION 1
Events

SECTION 2
Probability

SECTION 3
Bayes' Theorem

SECTION 4
Simple Classification





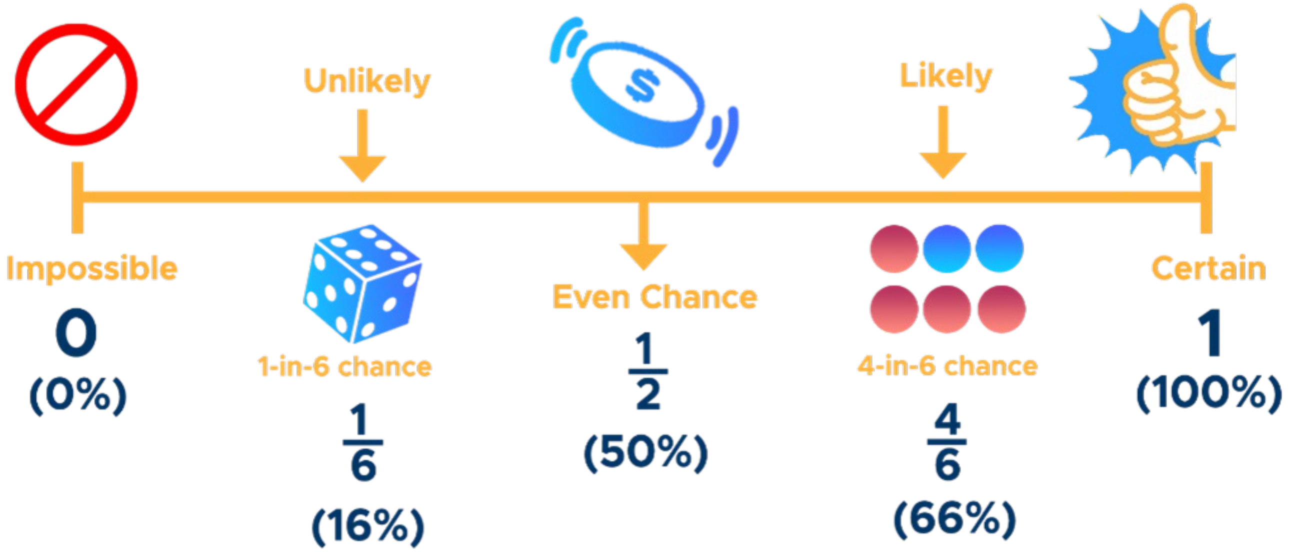
Probability

❖ Definition

- ✓ The probability of an event A is P(A)
  - ✓ A number between 0 and 1 that shows how likely the event is

- ✓ P(A) => 0: very unlikely that the event A occurs
- ✓ P(A) => 1: very likely to occur

- ✓ Some properties:
  - $0 \leq P(A) \leq 1$
  - $P(\Omega) = 1$
  - $P(\emptyset) = 0$



❖ Classical Probability

The theoretical probability of an event A is the number of ways the event can occur divided by the total number of possible outcomes:

$$P(A) = \frac{number\ of\ favorable\ outcomes}{total\ number\ of\ possible\ outcomes} = \frac{n_A}{n_\Omega}$$

Example 1

What is the probability of rolling a number is odd on a regular dice?

- There are 6 faces on a fair dice, numbered 1 to 6

=> $n_\Omega = 6$

- A : "odd number" => A = {1, 3, 5} => $n_A = 3$

=> P(A) = 3/6 = 0.5

Probability

❖ Classical Probability

The theoretical probability of an event A is the number of ways the event can occur divided by the total number of possible outcomes:

$$P(A) = \frac{number\ of\ favorable\ outcomes}{total\ number\ of\ possible\ outcomes} = \frac{n_A}{n_\Omega}$$

Example 2

What is the probability of rolling a number is greater than 3?

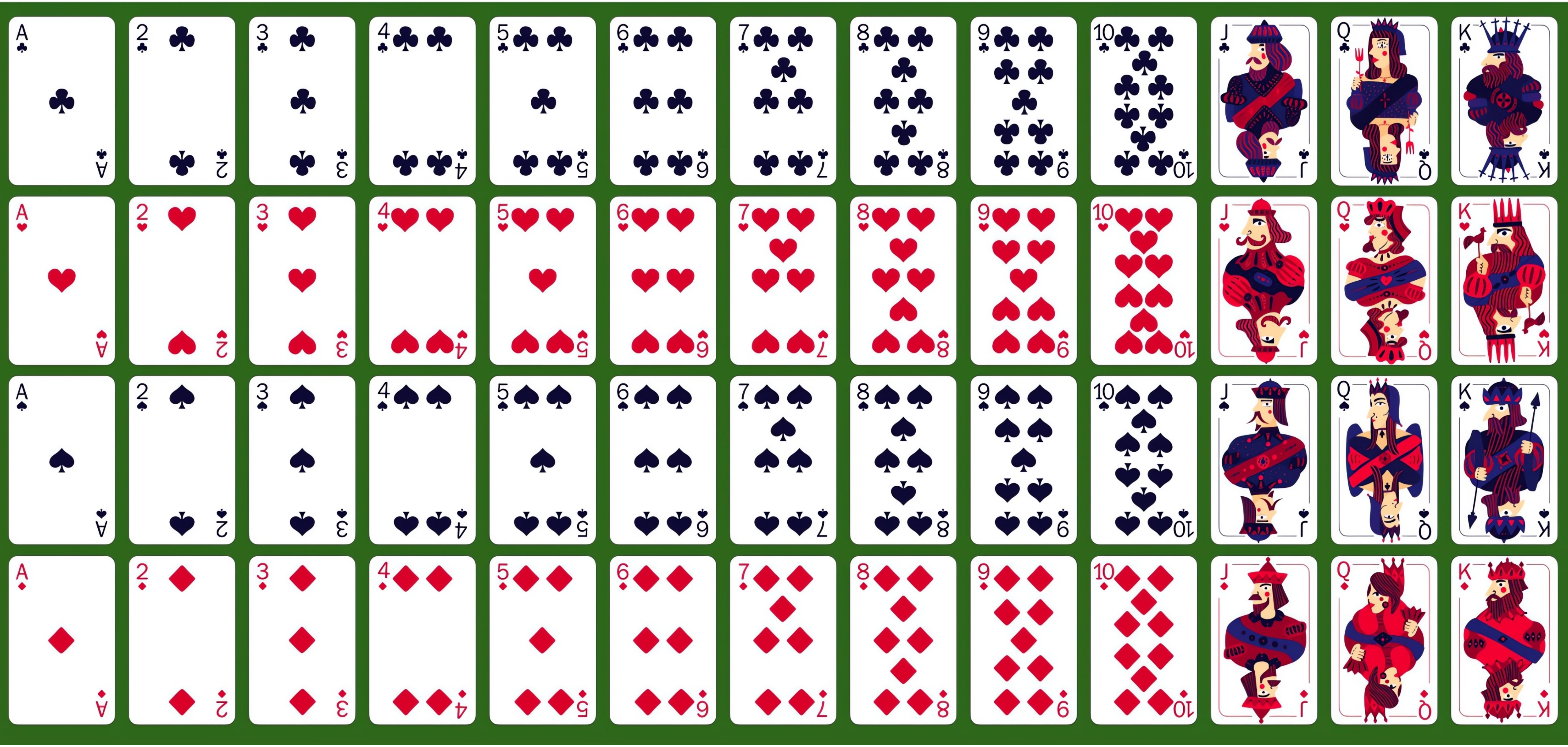
- There are 6 faces on a fair dice, numbered 1 to 6

=> $n_\Omega = 6$

- A : "a number is greater than 3" => A = {4, 5} => $n_A = 2$

=> P(A) = 2/6 = 1/3

Probability



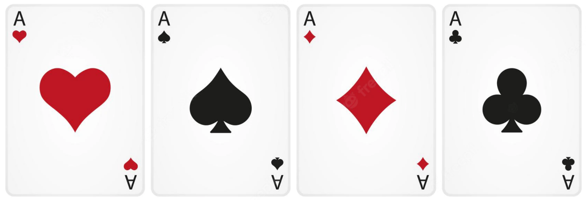
Probability

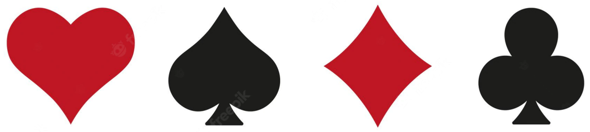
❖ Classical Probability

Example 3
Drawing a card from a well-shuffled deck. Find the probability of some events

Drawing a king

Drawing a black card





Drawing a king

- A: "Drawing a king from a deck of cards"
- There are 52 cards in a deck of cards
  => $n_\Omega = 52$
- There are 4 kings in a deck
  => $n_A = 4$
  => P(A) = 4/52 = 1/13

Drawing a black card

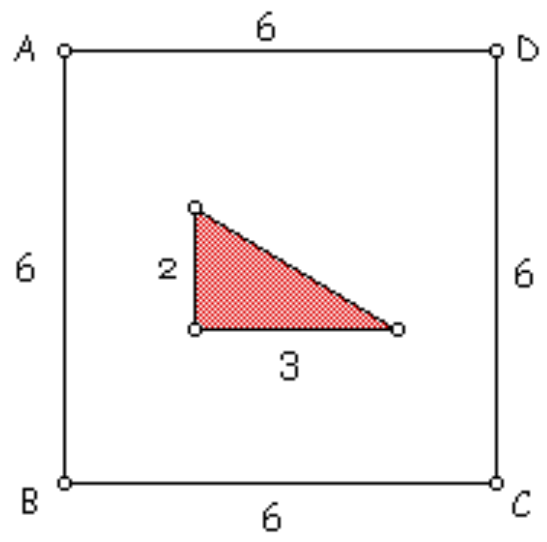
- A: "Drawing a black card from a deck of cards"
- There are 52 cards in a deck of cards
  => $n_\Omega = 52$
- There are 26 black cards in a deck
  => $n_A = 26$
  => P(A) = 26/52 = 1/2

Probability

❖ Geometric Probability

When a variable is continuous, classical probability becomes impossible to "count" the outcomes.

$$P(A) = \frac{measure\ of\ domain\ A}{measure\ of\ domain\ \Omega}$$

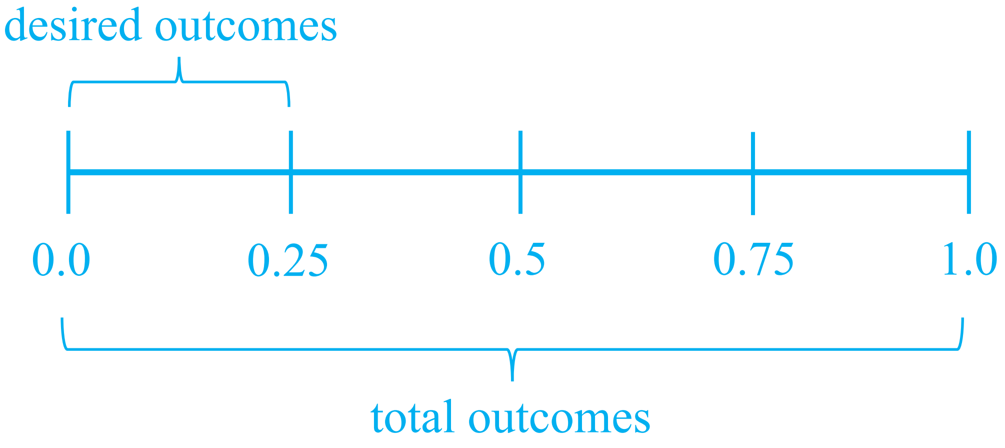


$P(A) = ?$

X is a random real number between 0 and 1.

$$P(A) = \frac{length\ of\ segment\ where\ 0 < X < 0.25}{length\ of\ segment\ where\ 0 < X < 1}$$

$$= \frac{0.25}{1} = 0.25$$

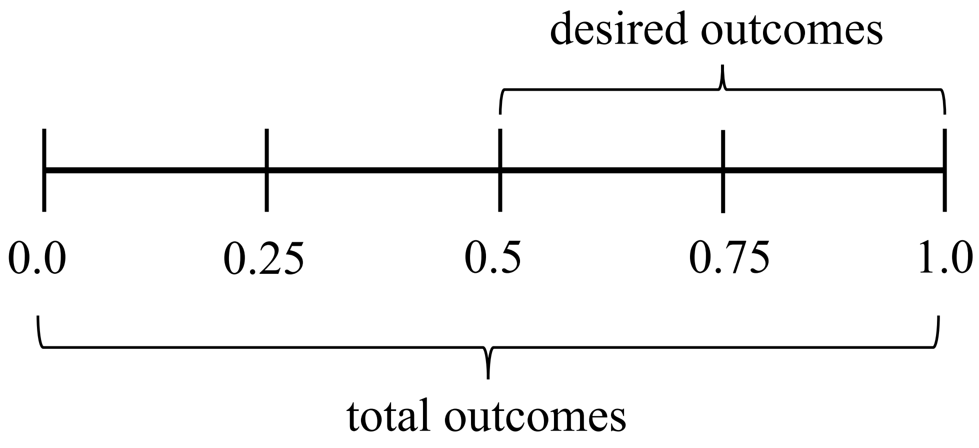


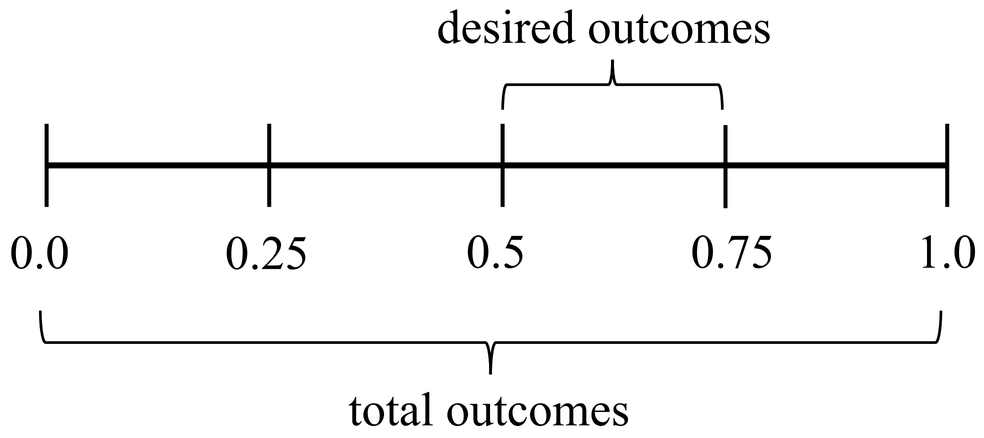
Probability

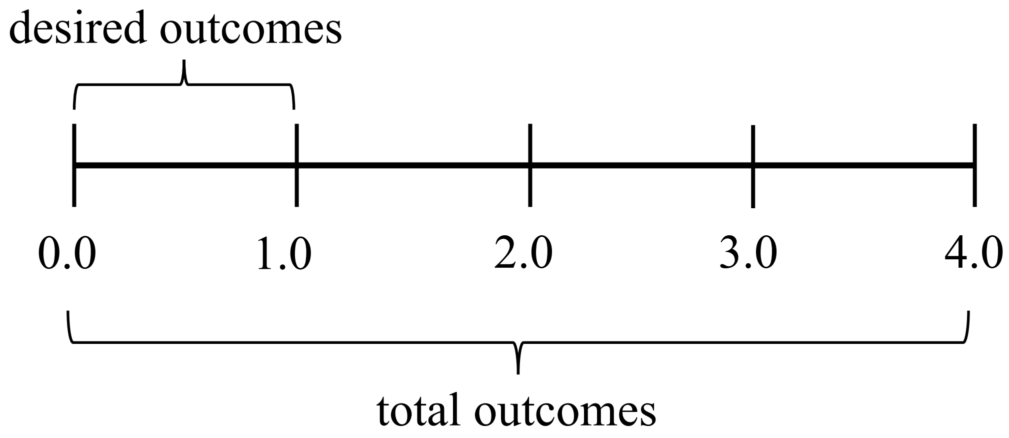
❖ Geometric Probability

When a variable is continuous, classical probability becomes impossible to "count" the outcomes.

$$P(A) = \frac{\textit{measure of domain } A}{\textit{measure of domain } \Omega}$$







Probability

❖ Simulation of coin tossing

Count #heads

Count #tails

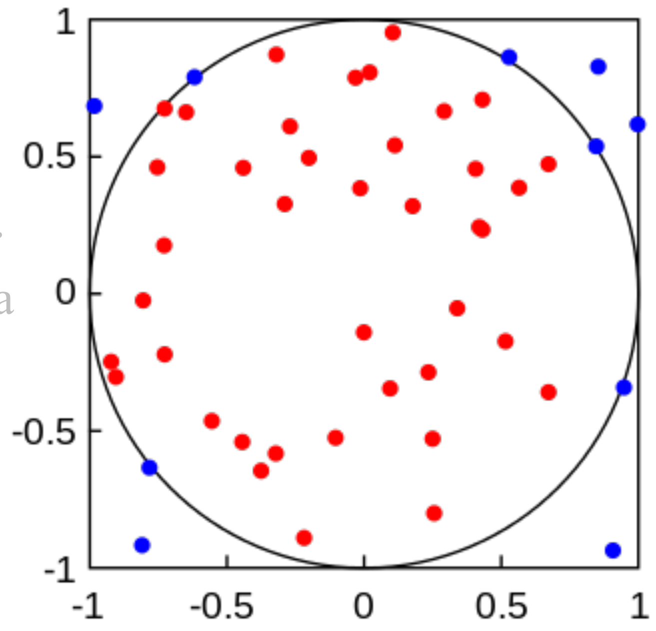
Check if the two numbers are similar

```python
# aivietnam.ai
import random

# Tổng số lần búng đồng xu
total_flips = 0

# số lần mặt sau xuất hiện
num_tails = 0

# số lần mặt trước xuất hiện
num_heads = 0

for _ in range(1000):
    # sinh số ngẫu nhiên nằm trong khoảng [0,1)
    n = random.random()
    if n < 0.5:
        num_tails = num_tails + 1
    else:
        num_heads = num_heads + 1

    # code ở vị trí này không thuộc khối else
    total_flips = total_flips + 1
```

Probability

❖ PI estimation

hình từ
wikipedia



N_s is #random samples within the square generated according to uniform distribution

N_c is #random samples within the circle generated according to uniform distribution

circle radius $r = 1$

circle_area $A_c = \pi r^2$

square side $s = 2$

square_area $A_s = s^2$

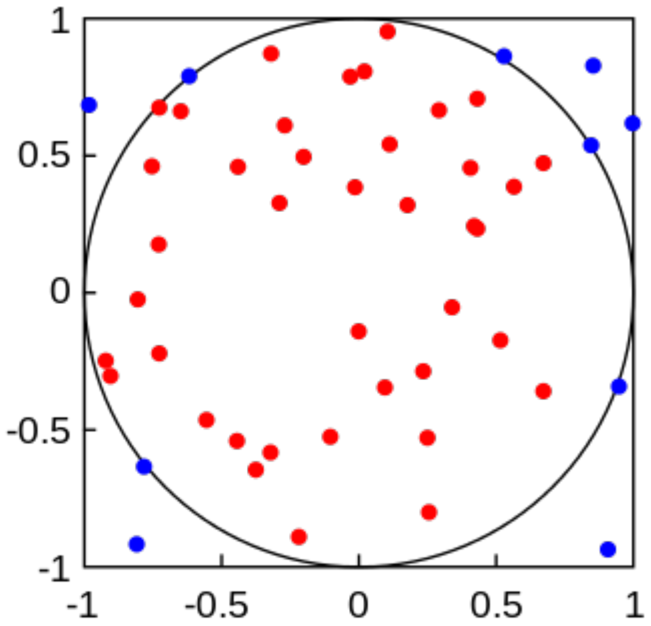
$$\frac{A_s}{A_c} \approx \frac{N_s}{N_c}$$

$$\frac{s^2}{\pi r^2} \approx \frac{N_s}{N_c}$$

$$\pi \approx \frac{s^2 N_c}{N_s}$$

Probability

❖ PI estimation



(hình từ wikipedia)

$$\pi \approx \frac{s^2 N_c}{N_s}$$

```python
# aivietnam.ai
import random
import math


# Tổng số điểm p được sinh ra
N = 100000

# số điểm thuộc tình tròn
N_T = 0

# Sinh ra N điểm ngẫu nhiên
for i in range(N):
    # sinh ra x, y thuộc [-1, 1].
    x = random.random()*2 - 1
    y = random.random()*2 - 1

    x2 = x**2
    y2 = y**2

    # kiểm tra p có nằm trong đường tròn
    if math.sqrt(x2 + y2) <= 1.0:
        N_T = N_T + 1

# tính PI
pi = (N_T / N) * 4
print(pi)
```

Probability

❖ Empirical probability (experimental probability)

Estimating probabilities from experience and observation

Using the number of occurrences of a given outcome within a sample set as a basis for determining the probability

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail
Yes	Yes	Yes	???

$$P(A) = \frac{\#\text{ of times occurred}}{\text{total }\#\text{ of times experiment performed}}$$

The Result column is a special one.
In ML, it is called Target or Label

$$\text{P(Result = "Pass")} = \frac{3}{6} = 1/2$$

$$\text{P(Result = "Fail")} = \frac{3}{6} = 1/2$$

Rules of probability

❖ The additive rule

Mutually exclusive events

$$P(A+B) = P(A) + P(B)$$

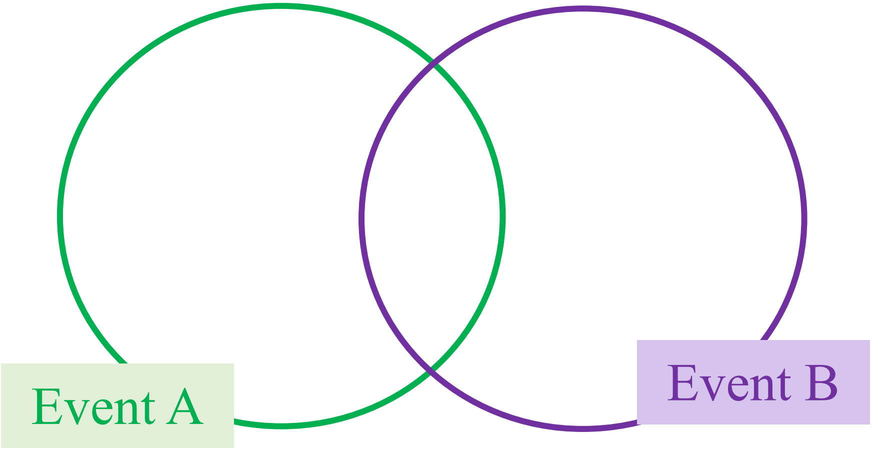
$$P(A \text{ or } B) = P(A) + P(B)$$

where A and B are mutually exclusive

In general

$$P(A+B) = P(A) + P(B) - P(AB)$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$



Example

Rolling a fair dice. What is the probability of $A = \{3, 4\}$?

- The dice is fair => all six possible outcomes are equally likely

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\})$$

- The events $\{1\},\ldots,\{6\}$ are disjoint

$$1 = P(S) = P(\{1\}) + P(\{2\}) + \ldots + P(\{6\}) = 6P(\{1\})$$

$$\Rightarrow P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = 1/6$$

- Since $\{3\}$ and $\{4\}$ are disjoint

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

Rules of probability

❖ **Example**

Suppose we have the following information:

1. There is a 70 percent that ad visits Ha Noi.
2. There is a 60 percent that ad visits Ho Chi Minh.
3. There is a 40 percent that ad visits 2 cities: Ha Noi and Ho Chi Minh

Find the probability that ad visits Ha Noi or Ho Chi Minh?

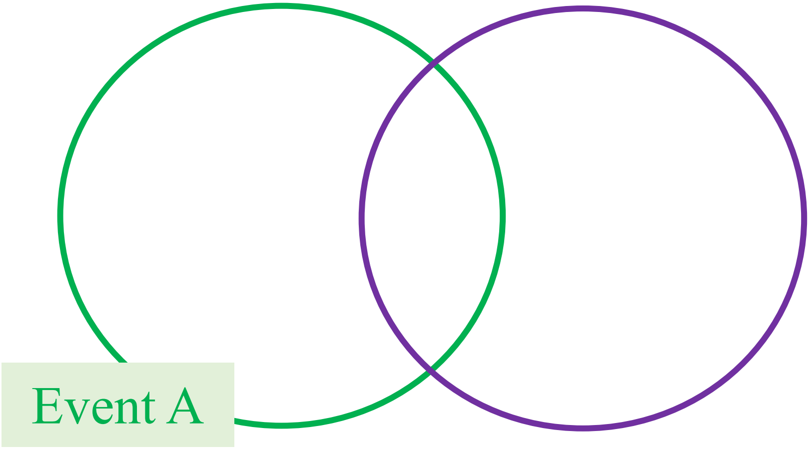
Let's define

A: "Ad visits Ha Noi" => $P(A) = 0.7$

B: "Ad visits Ho Chi Minh" => $P(B) = 0.6$

$P(A \text{ and } B) = 0.4$

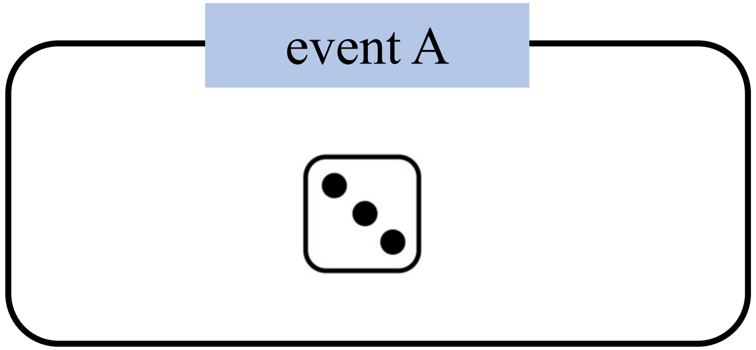
$\Rightarrow P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.7 + 0.6 - 0.4 = 0.9$

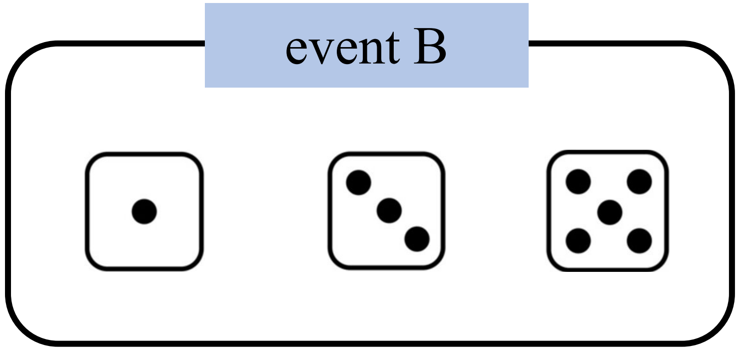


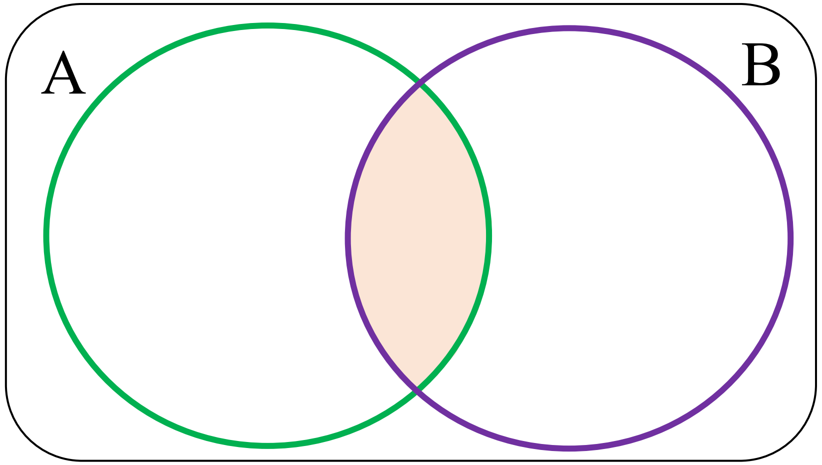
Conditional Probability

❖ Definition

Find the probability that the number rolled is a three, given that it is odd







event B → event A

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

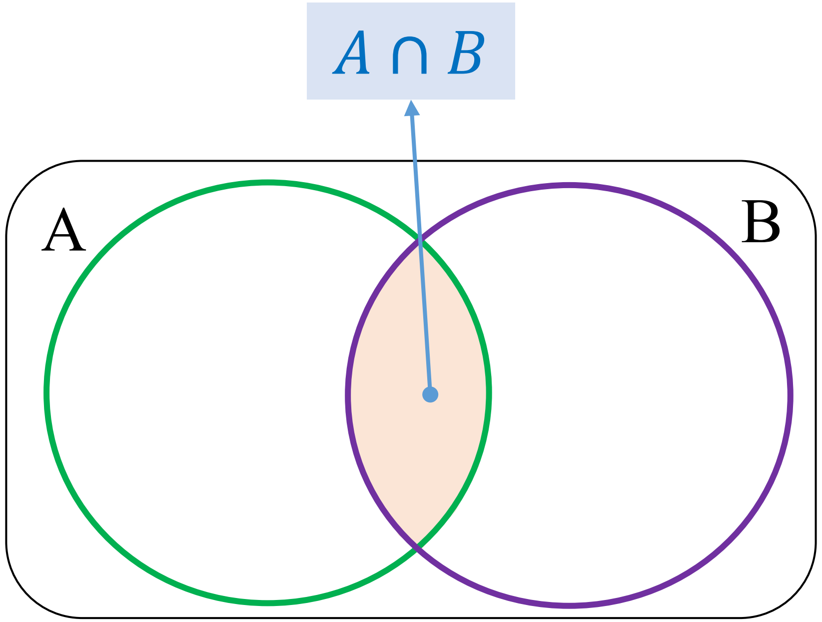
Conditional Probability

❖ Definition

Conditional Probability Formula

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

Probability that A occurs given that B has already occurred



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

A fair die is rolled

a) Find the probability that the number rolled is a three, given that it is odd.
b) Find the probability that the number rolled is odd, given that it is a three.

Conditional Probability

❖ Example

A fair dice is rolled

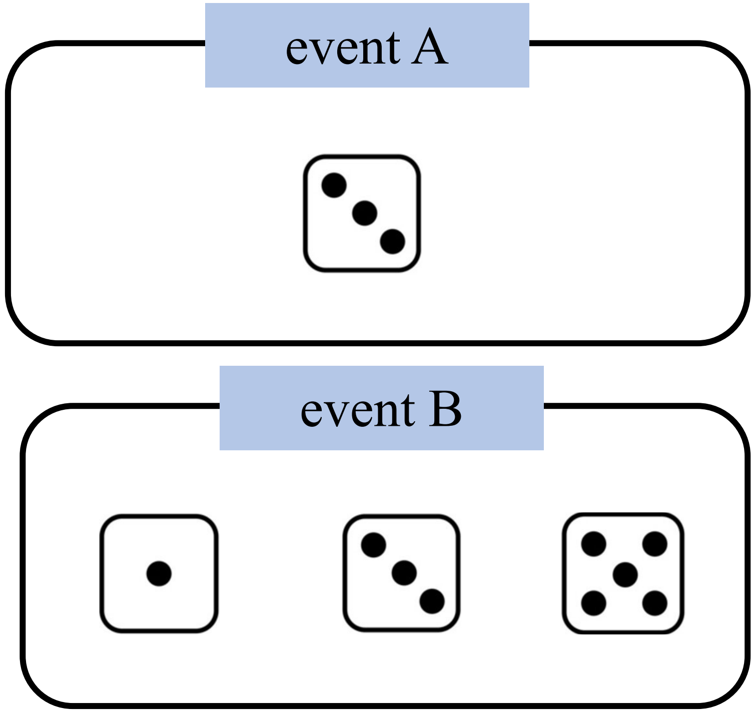
- Sample space S = {1, 2, 3, 4, 5, 6}, consisting of 6 equally likely outcomes
- A: “a three is rolled”
  => A = {3} => P(A) = 1/6
- B: “an odd number is rolled”
  => B = {1, 3, 5} => P(B) = 3/6 = 1/2
  => A and B = {3} => P(A and B) = 1/6

a) Find the probability that the number rolled is a three, given that it is odd.
P(A|B) = P(A and B)/P(B) = (1/6)/(1/2) = 1/3

b) Find the probability that the number rolled is odd, given that it is a three.
P(B|A) = P(B and A)/P(A) = P(A and B)/P(A) = (1/6)/(1/6) = 1

Conditional probability

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



Multiplication Rule

Multiplication rule:

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

General:

$$P(A_1A_2\ldots A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$

Example

There are 10 identical keys, two of which can open a door.

We randomly pick and try each key, discarding it if it is unusable.

What is the probability that the door is opened on the 3rd attempts?

Multiplication Rule

$$P(A_1A_2\ldots A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$

Example

Let's A_i as the event i^{th} chosen the right key, for i =1, 2, 3

=> Compute $P(\bar{A}_1\bar{A}_2A_3) = P(A_3|\bar{A}_1\bar{A}_2)P(\bar{A}_2|\bar{A}_1)P(\bar{A}_1)$

$P(A_1) = 2/10$ => $P(\bar{A}_1) = 8/10$

Given that the first chosen key was wrong, the second key will be chosen from 7 wrong keys and 2 right keys, thus: $P(A_2|\bar{A}_1) = 2/9$ => $P(\bar{A}_2|\bar{A}_1) = 7/9$

Given that the first and second chosen keys were wrong, the third key will be chosen from 6 wrong keys and 2 right keys, thus: $P(A_3|\bar{A}_1\bar{A}_2) = 2/8$

=> $P(\bar{A}_1\bar{A}_2A_3) = 8/10 * 7/9 * 2/8 = 0.155$

Multiplication Rule

Multiplication rule:

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

General:

$$P(A_1A_2\ldots A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$

Example

There are 10 identical keys, two of which can open a door.

We randomly pick and try each key, discarding it if it is unusable.

What is the probability that the door is opened within at most three attempts?

Multiplication Rule

$$P(A_1A_2\ldots A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$

Example

Let's A_i as the event i^{th} chosen the right key, for i =1, 2, 3

=> Compute $P(A_1) + P(\bar{A}_1 A_2) + P(\bar{A}_1 \bar{A}_2 A_3)$

$$\ldots = \frac{2}{10} + \frac{16}{90} + \frac{14}{90} = \frac{8}{15}$$

$P(A_1) = 2/10$ => $P(\bar{A}_1) = 8/10$

Given that the first chosen key was wrong, the second key will be chosen from 7 wrong keys and 2 right keys, thus: $P(A_2|\bar{A}_1) = 2/9$ => $P(\bar{A}_2|\bar{A}_1) = 7/9$

Given that the first and second chosen keys were wrong, the third key will be chosen from 6 wrong keys and 2 right keys, thus: $P(A_3|\bar{A}_1\bar{A}_2) = 2/8$

=> $P(\bar{A}_1\bar{A}_2A_3) = 8/10 * 7/9 * 2/8 = 0.155$

Independent events

- Events A and B are independent if:

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

- If A and B are not independent, they are dependent.

Example

A single fair dice is rolled. Let A={4} and B={2, 4, 6}.

Are A and B independent?

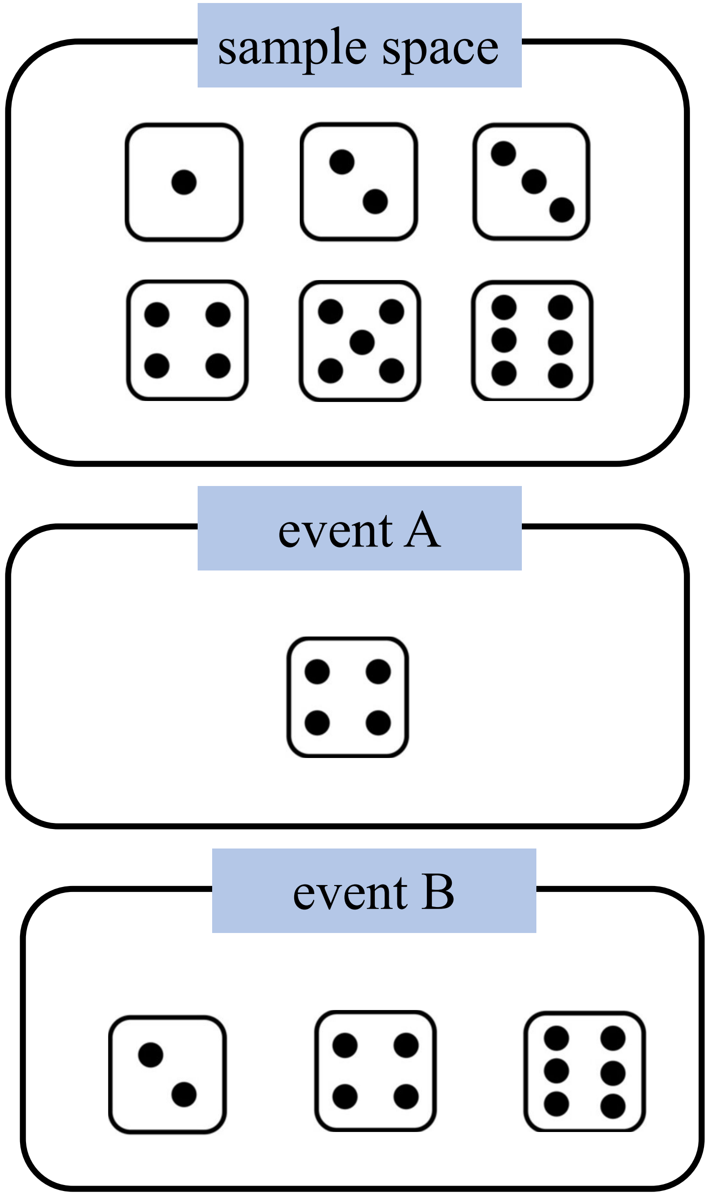
Compute: $P(A) = 1/6$

$P(B) = 1/2$

$P(A \text{ and } B) = 1/6$

Since $P(A)P(B) = (1/6)*(1/2) = 1/12 \neq P(A \text{ and } B) = 1/6$

=> Events A and B: not independent



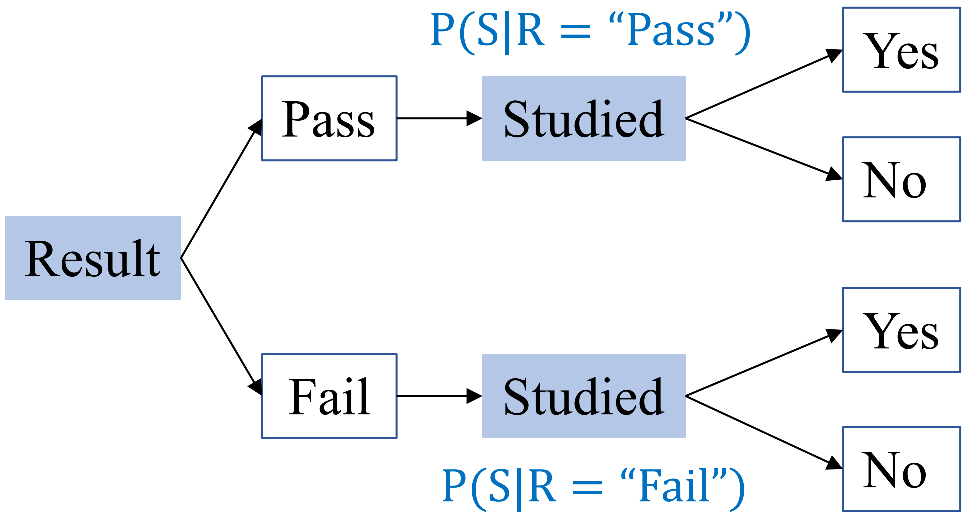
Outline

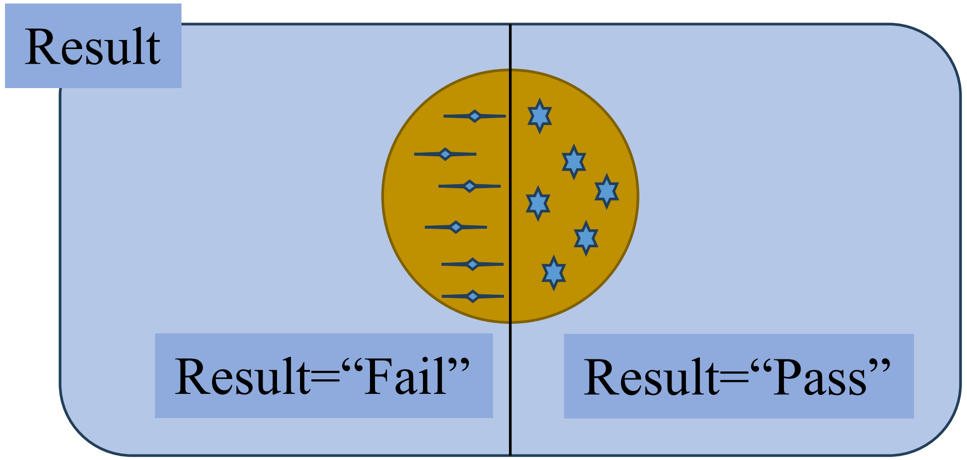
SECTION 1
Events

SECTION 2
Probability

SECTION 3
Bayes' Theorem

SECTION 4
Simple Classification



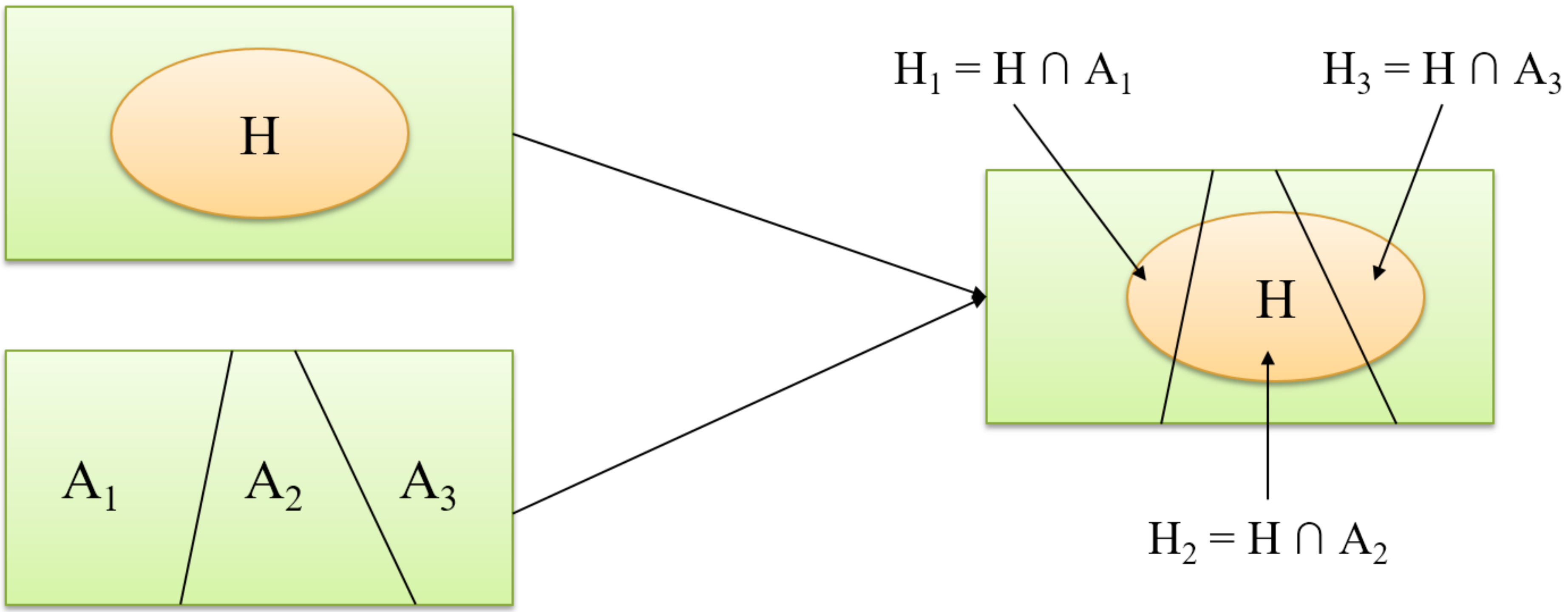


Total Probability Theorem

Let A_1, A_2, …, A_n – complete system of events. Consider any event H such that H occurs only when one of the events $A_1, A_2, \ldots, A_n$ occurred

$$P(H) = P(H_1) + P(H_2) + P(H_3)$$

$$= P(A_1).P(H|A_1) + P(A_2).P(H|A_2) + P(A_3).P(H|A_3)$$



Total Probability Theorem

In general: $P(H) = \sum_{i=1}^{n} P(A_i).P(H|A_i)$

Example

I have three bags that each contain 100 marbles:

- Bag 1 has 75 red and 25 blue marbles
- Bag 2 has 60 red and 40 blue marbles
- Bag 3 has 45 red and 55 blue marbles.

I choose one of the bags at random and then pick a marble from the chosen bag, also at random. What is the probability that the chosen marble is red?

Total Probability Theorem

In general: $P(H) = \sum_{i=1}^{n} P(A_i).P(H|A_i)$

Example

H: “the chosen marble is red”

A_i : the event that I choose Bag I

=> $P(H|A_1) = 0.75$; $P(H|A_2) = 0.6$; $P(H|A_3) = 0.45$

Each bag contain 100 marbles and because their union is the entire sample space

$P(A_1 \cup A_2 \cup A_3) = 1$

The probability that the chosen marble is red:

$$P(H) = P(A_1).P(H|A_1) + P(A_2).P(H|A_2) + P(A_3).P(H|A_3)$$

$$= 1/3*0.75 + 1/3*0.60 + 1/3*0.45 = 0.60$$

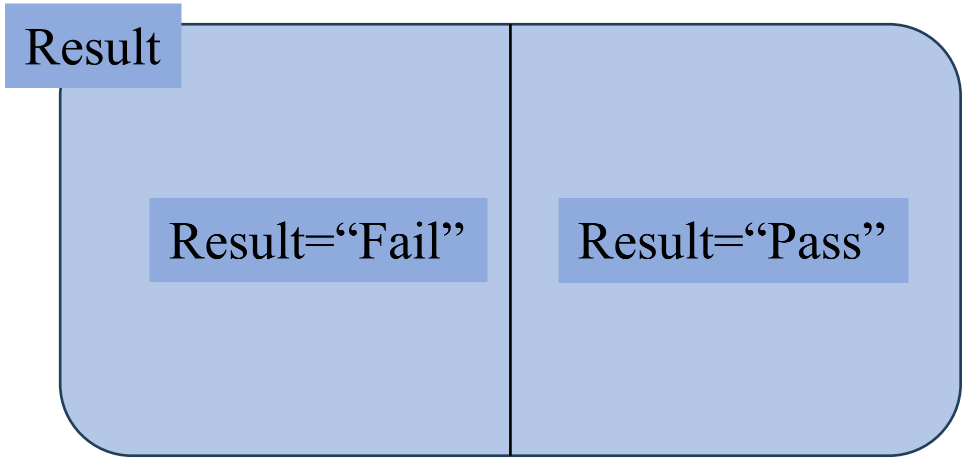
Total Probability Theorem

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail
Yes	Yes	Yes	???

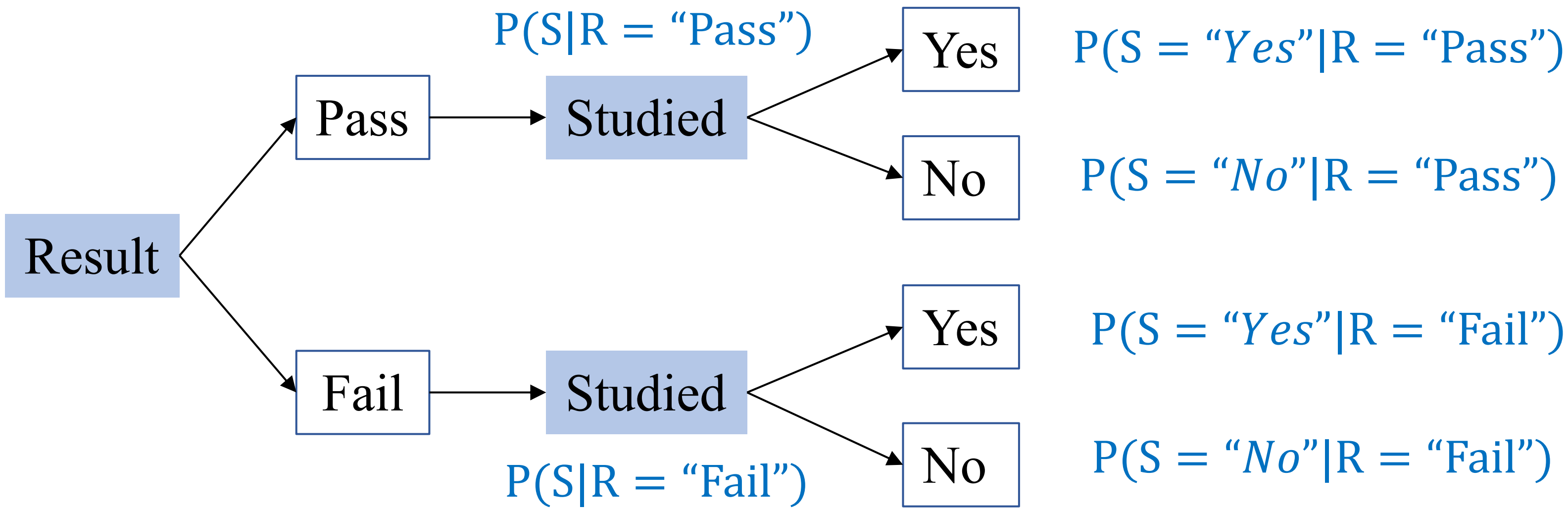
Three features: Confident, Studied, and Sick

Two classes: Fail and Pass

Compute P(S|R)?



Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail
Yes	???

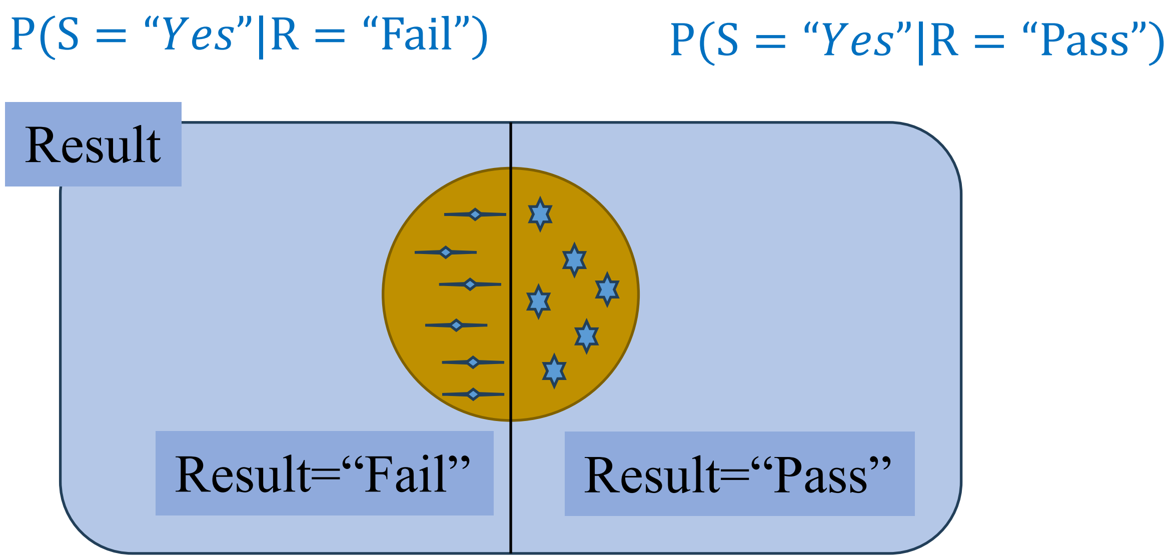


Total Probability Theorem

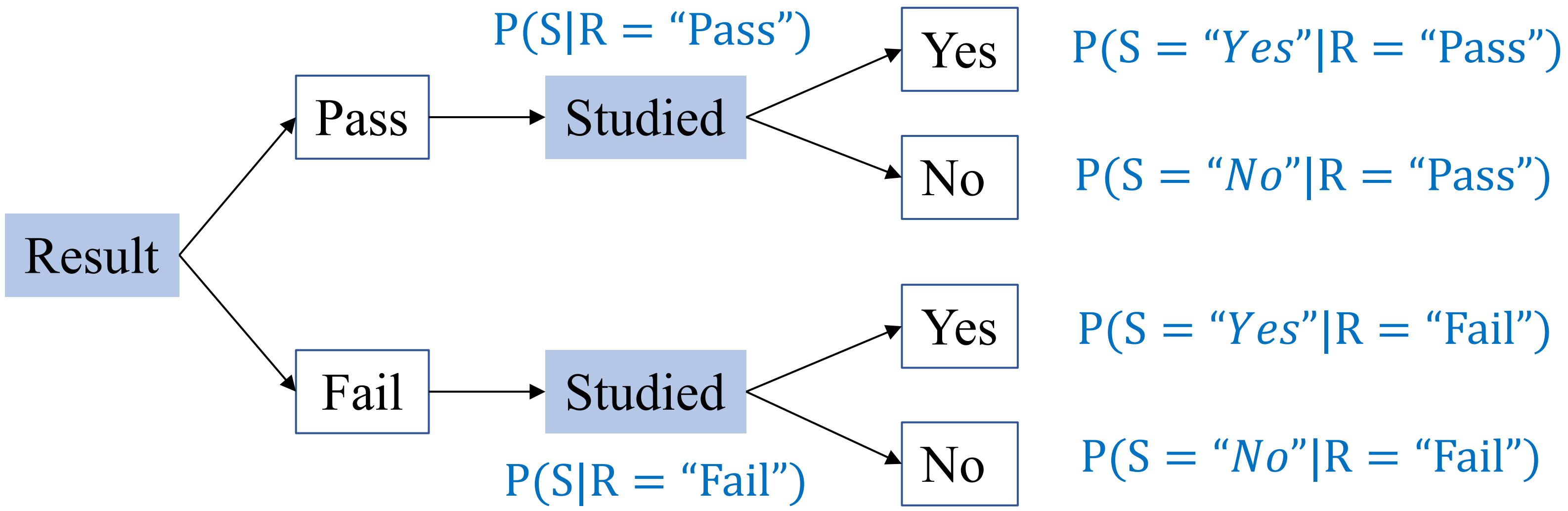
Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	No	Pass
No	No	Yes	Fail
Yes	Yes	Yes	???

Three features: Confident, Studied, and Sick

Two classes: Fail and Pass



Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail
Yes	???



Bayes' Rule

For any two events A and B, where $P(A) \neq 0$:

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

PRIOR
The probability of "B" being True. This is the knowledge

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

POSTERIOR
The probability of "B" being True. Given "A" True

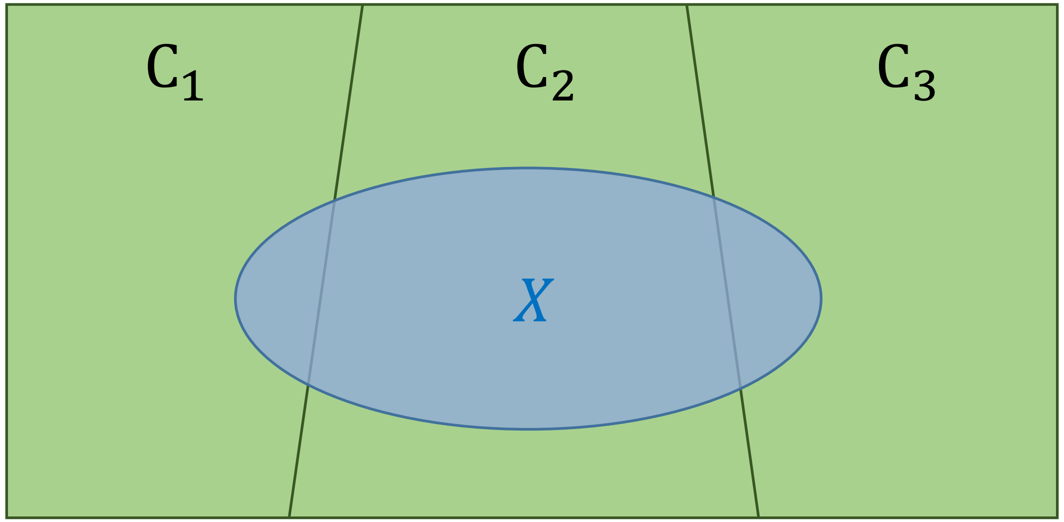
MARGINALIZATION
The probability of "A" being True.

Bayes' Rule

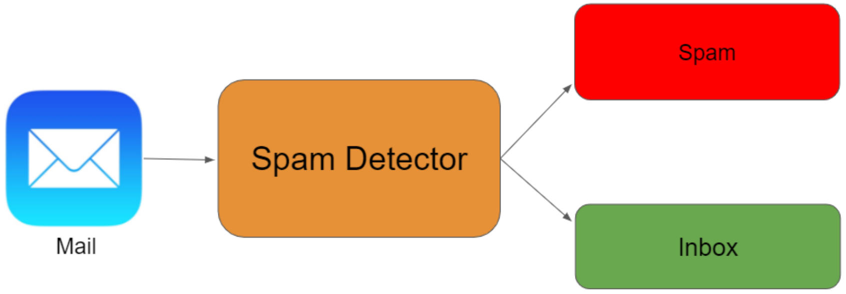
For any two events A and B, where $P(A) \neq 0$: $P(B = B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}$

If $C_1, C_2, .. C_n$ are complete system of events, and X is any event with $P(X) \neq 0$

$$P(C_i|X) = \frac{P(C_i)P(X|C_i)}{P(X)} = \frac{P(C_i)P(X|C_i)}{\sum_{j=1}^{n} P(C_j)P(X|C_j)}, i = 1, 2, \dots, n$$



Bayes' Rule



Example: Detect Spam E-Mail (Simple NLP problem)

Assume that the word 'offer' occurs in 80% of the spam messages in my account. Also, let's assume 'offer' occurs in 10% of my desired e-mails. If 30% of the received e-mails are considered as a spam. I will receive a new message which contains 'offer', what is the probability that it is spam?

Assume that I received 100 e-mails

100 e-mails

100*0.3=30 → Spam

100*0.7=70 → Desired

Spam:

- Contains 'offer' 30*0.8=24
- NOT contains 'offer' 30*0.2=6

Desired:

- Contains 'offer' 70*0.1=7
- NOT contains 'offer' 70*0.9=63

Bayes' Rule

Example: Detect Spam E-Mail (Simple NLP problem)

Let C_1: "Spam" and C_2: "Not spam"

=> C_1, C_2 : complete system of events

X: "contains the word 'offer'"

If a new message which contains 'offer', the probability that it is spam is:

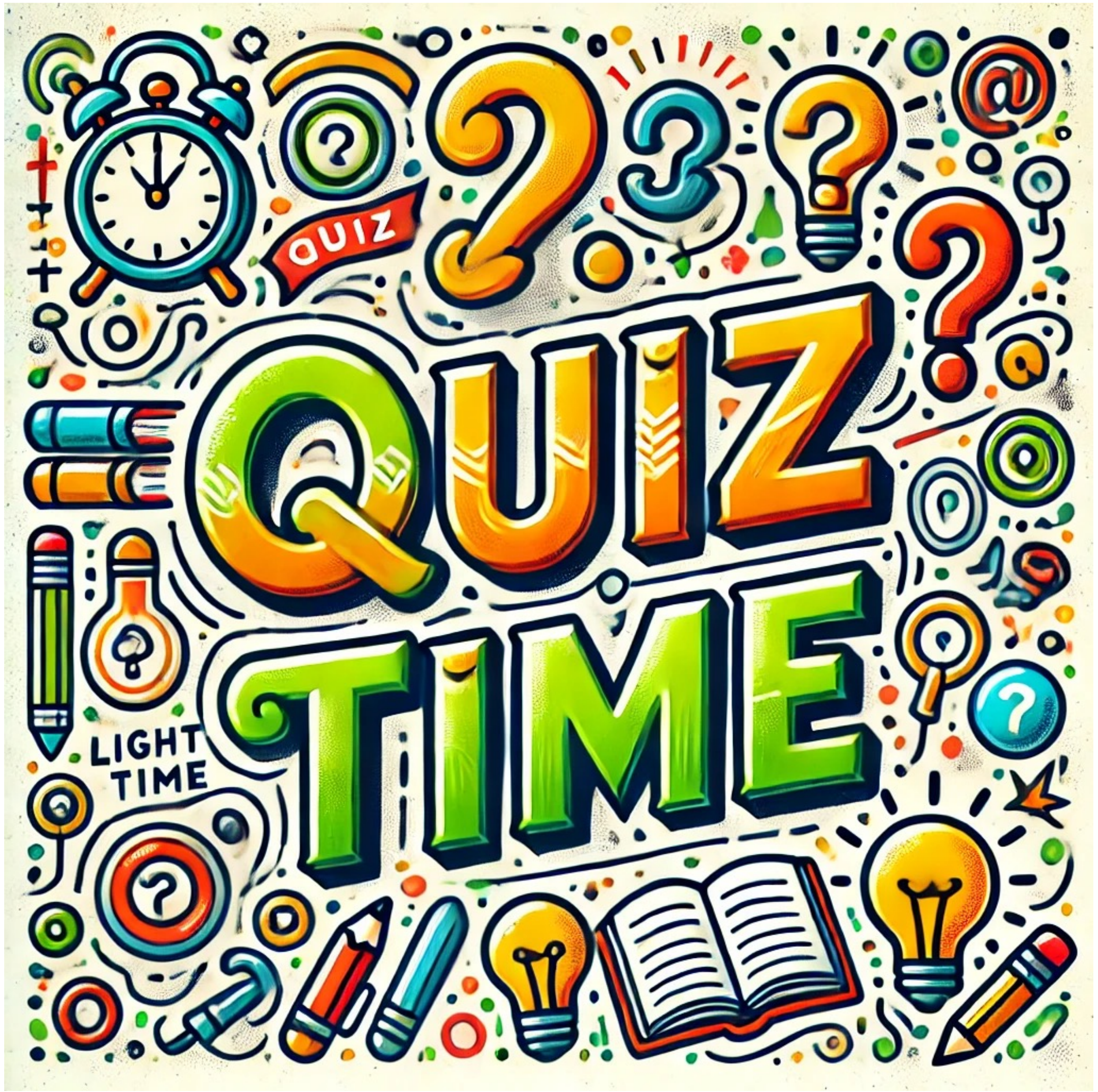
$$P(C_1|X) = \frac{P(C_1)P(X|C_1)}{P(X)}$$

$P(C_1) = 0.3$; $P(C_2) = 1 - P(C_1) = 0.7$

$P(X|C_1) = 0.8$; $P(X|C_2) = 0.1$

$P(X) = P(C_1)P(X|C_1) + P(C_2)P(X|C_2) = 0.3*0.8 + 0.7*0.1 = 0.31$

=> $P(C_1|X) = (0.8*0.3)/(0.31) = 0.774$


QUIZ
QUIZ TIME
LIGHT TIME

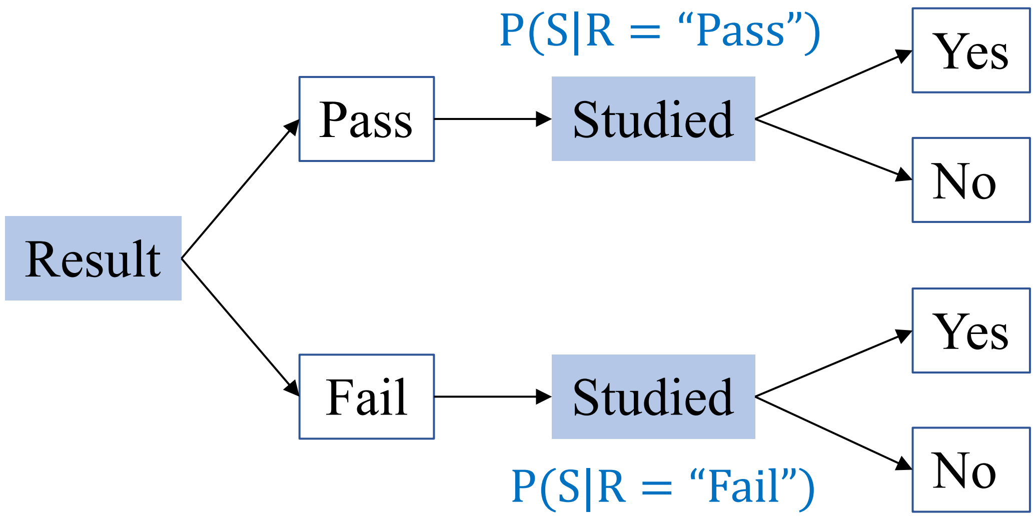
Outline

SECTION 1
Events

SECTION 2
Probability

SECTION 3
Bayes' Theorem

SECTION 4
Simple Classification



Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail
Yes	???

Simple Classification

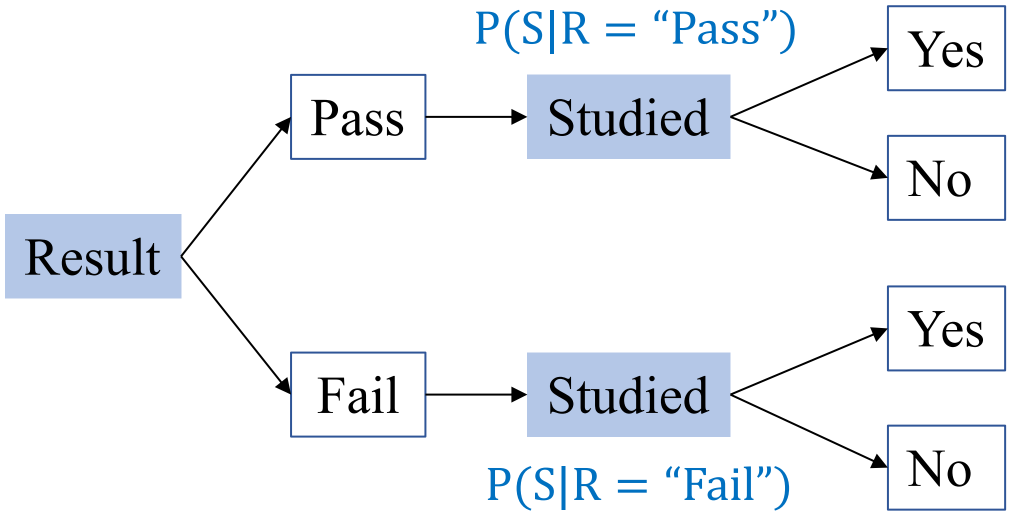
❖ Example: One feature

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail
Yes	???

One feature: Studies

Two classes: Fail and Pass

$$P(C_i|X) = \frac{P(C_i)P(X|C_i)}{P(X)}$$



Let C and X are random variables

$$p(C = c|X = x) = \frac{p(X = x|C = c) * p(C = c)}{p(X = x)}$$

Simple Classification

❖ Example: One feature

$$p(C = c|X = x) = \frac{p(X = x|C = c) * p(C = c)}{p(X = x)}$$

$p(C = c_1|X = x) =?$ $\quad$ $p(C = c_2|X = x) =?$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail

$$\boxed{p(res = pass) = \frac{3}{6} = \frac{1}{2}}$$

$$\boxed{p(res = fail) = \frac{3}{6} = \frac{1}{2}}$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail

$$p(res = pass) = \frac{3}{6} = 0.5$$

$$p(res = fail) = \frac{3}{6} = 0.5$$

$$p(stud = yes|res = pass) = \frac{2}{3}$$

$$p(stud = yes|res = fail) = \frac{1}{3}$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail

$$p(res = pass) = \frac{3}{6} = \frac{1}{2}$$

$$p(res = fail) = \frac{3}{6} = \frac{1}{2}$$

$$p(stud = yes|res = pass) = \frac{2}{3}$$

$$p(stud = yes|res = fail) = \frac{1}{3}$$

$$p(stud = yes)$$

$$= p(stud = yes|res = pass) * p(res = pass) + p(stud = yes|res = fail) * p(res = fail)$$

$$= \frac{2}{3} * \frac{1}{2} + \frac{1}{3} * \frac{1}{2} = \frac{1}{2}$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

Studied	Result
No	Fail
No	Pass
Yes	Fail
Yes	Pass
Yes	Pass
No	Fail
Yes	???

$$p(res = pass) = \frac{3}{6} = \frac{1}{2}$$

$$p(res = fail) = \frac{3}{6} = \frac{1}{2}$$

$$p(stud = yes) = \frac{1}{2}$$

$$p(stud = yes|res = pass) = \frac{2}{3}$$

$$p(stud = yes|res = fail) = \frac{1}{3}$$

$$p(res = pass \mid stud = yes) = \frac{p(stud = yes \mid res = pass) * p(res = pass)}{p(stud = yes)}$$

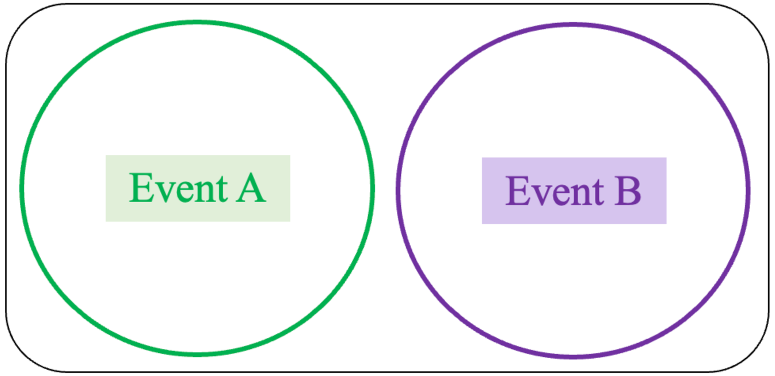
$$= \frac{2}{3} * \frac{1}{2} * \frac{2}{1} = \frac{2}{3}$$

$$p(res = fail \mid stud = yes) = \frac{p(stud = yes \mid res = fail) * p(res = fail)}{p(stud = yes)}$$

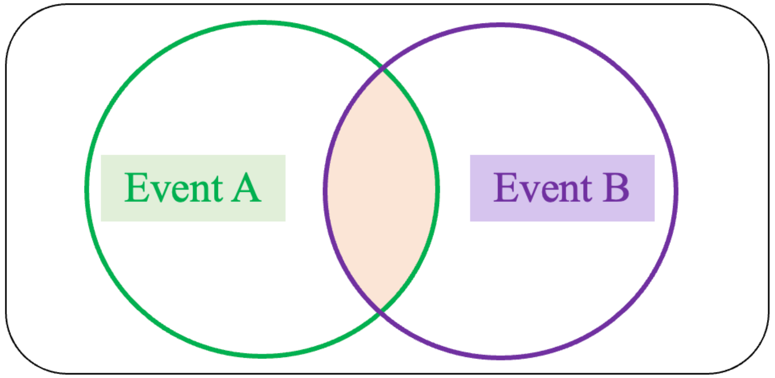
$$= \frac{1}{3} * \frac{1}{2} * \frac{2}{1} = \frac{1}{3}$$

Summary

Events

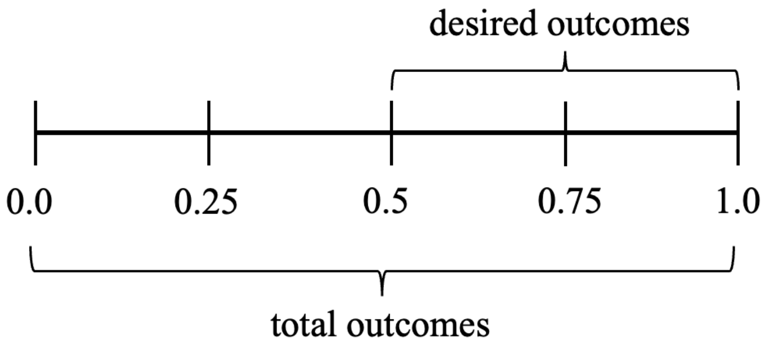


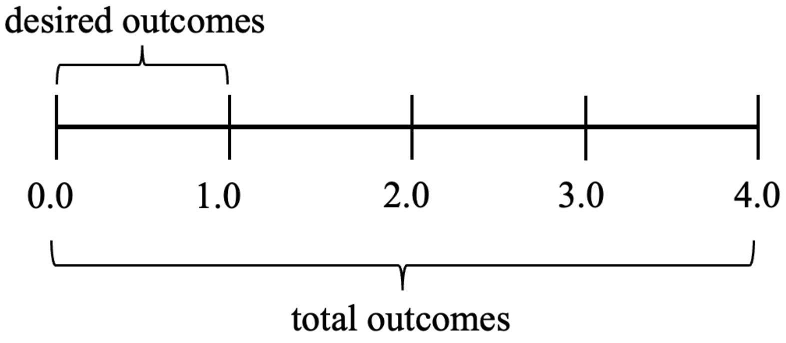
Mutually exclusive



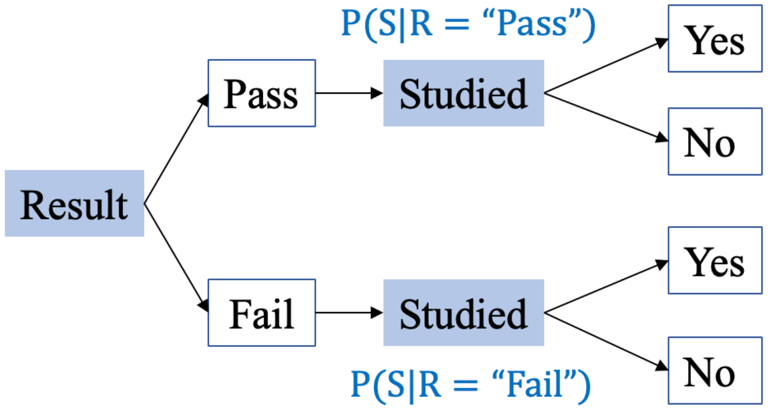
Non-mutually exclusive

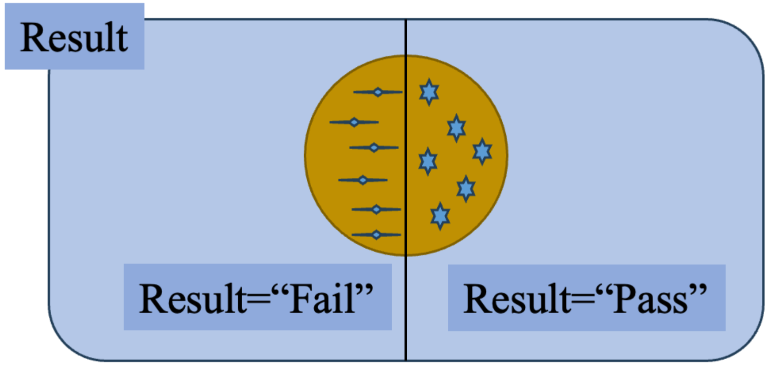
Probability





Bayes' Theorem





What's next?

- Continuous Radom Variables
- Gaussian Distribution
- Naïve Bayes Classification