Basic Statistics

Computational Social Intelligence - Lecture 02

Prof. Alessandro Vinciarelli
School of Computing Science &
Institute of Neuroscience and Psychology

http://www.dcs.gla.ac.uk/vincia
alessandro.Vinciarelli@glasgow.ac.uk

Texts (see Moodle)

This lecture is based on the following text (available on Moodle):

- Appendix A of F.Camastra and A.Vinciarelli, "Machine Learning for Audio, Image and Video Processing", Springer Verlag, 2008 (Pages 525 to 532 included).

Outline

- Definition of Probability
- Basic Laws of Probability
- Random Variables
- Conclusions

Outline

- Definition of Probability
- Basic Laws of Probability
- Random Variables
- Conclusions

Flipping Coins

- Flipping a coin is a simple statistical experiment and the outcome can be only Head (H) or Tail (T);
- Out of a sufficiently large number of attempts, it is possible to say that both H and T appear roughly half of the times;
- However, it is not possible to know what will be the outcome of the next attempt.

Number of times we observe outcome "Head" when flipping a coin

There can be small variations with respect to the expectations

$$\frac{n(H)}{n} \simeq \frac{1}{2}$$

Total number of times the coin is flipped

Fraction of the times the outcome "Head" is observed

Events are mutually exclusive when the occurrence of one implies the non-occurrence of the others (Head or Tail)

A set of K mutually exclusive events

$$A_1, A_2, \ldots, A_K$$

$$p(A_i) = \lim_{n \to \infty} \frac{n(A_i)}{n}$$

The probability of event "i"

The fraction of times the outcome is event "i" when n tends to infinity

The ratio is smaller or equal than 1 because n(A) cannot be larger than n

The value of p(A) is the value of the ratio when n tends to infinity

$$0 \leq \frac{n(A)}{n} \leq 1 \Rightarrow 0 \leq p(A) \leq 1$$

n(A) cannot be lower than 0, then the ratio cannot be lower than 0

The ratio is always smaller than 1, then p(A) is always smaller than 1

Outline

- Definition of Probability
- Basic Laws of Probability
- Random Variables
- Conclusions

The Sample Space

- The sample space is a set of L mutually exclusive events;
- An event A is the outcome of a statistical experiment that corresponds to one, several or all elements of the sample space;
- The main advantage of the sample space is that it allows one to explain the probability laws with the set theory.

The sample space contains L mutually exclusive events

The sample space contains L mutually exclusive events

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_L\}$$

$$A = \{\omega_i, \omega_j, \ldots, \omega_k\} \subseteq \Omega$$

The event A corresponds to one, several or all the events of the sample space

The occurrence of A depends on the occurrence of the events of the sample space

The sample space contains the six faces of a dice

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$$

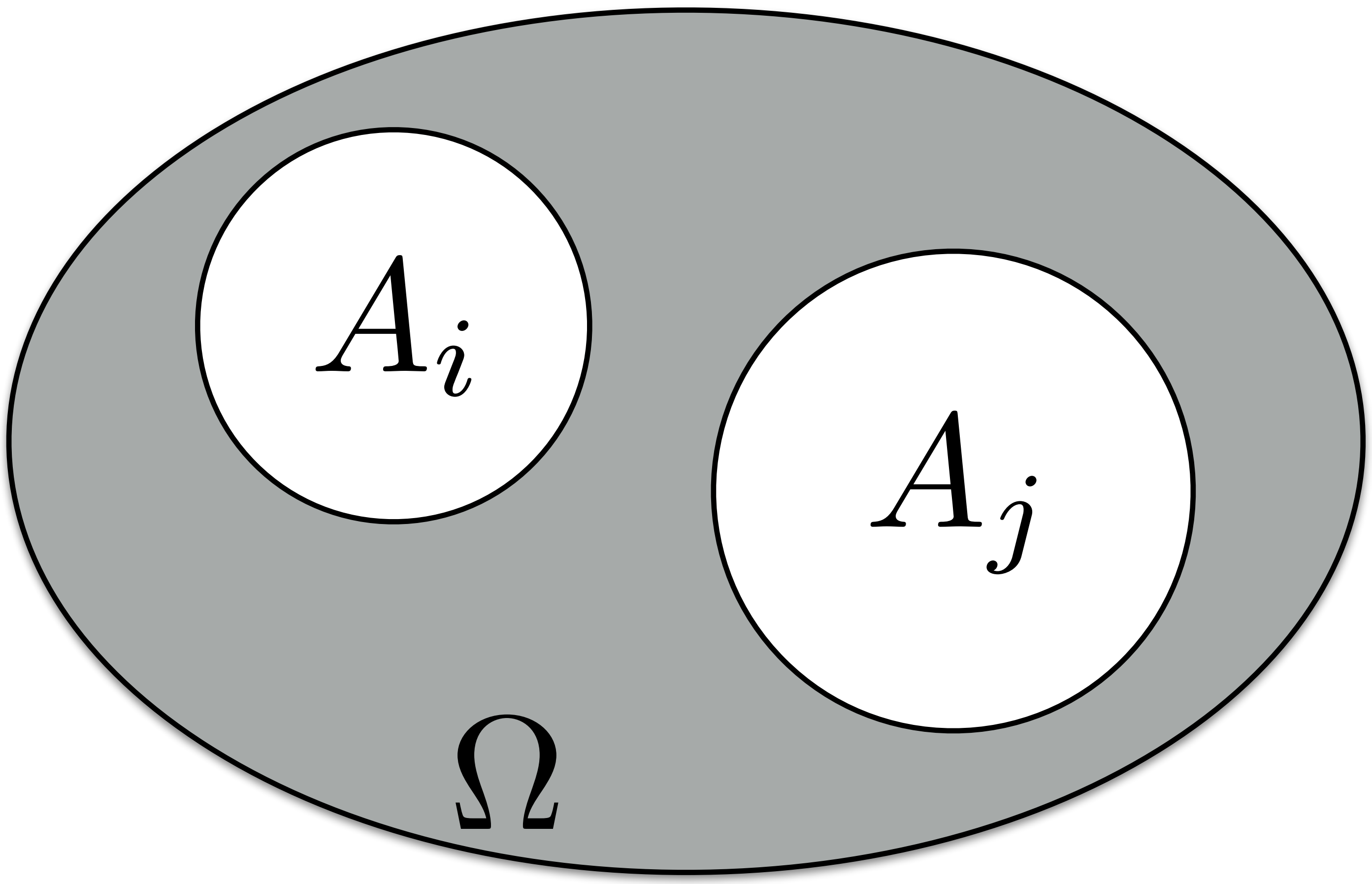
$$A = \{\omega_2, \omega_4, \omega_6\}$$

The event A corresponds to the occurrence of an even number

The event A is a subset of the sample space

Mutually Exclusive Events

No shared element ω_k between the events A_i and A_j

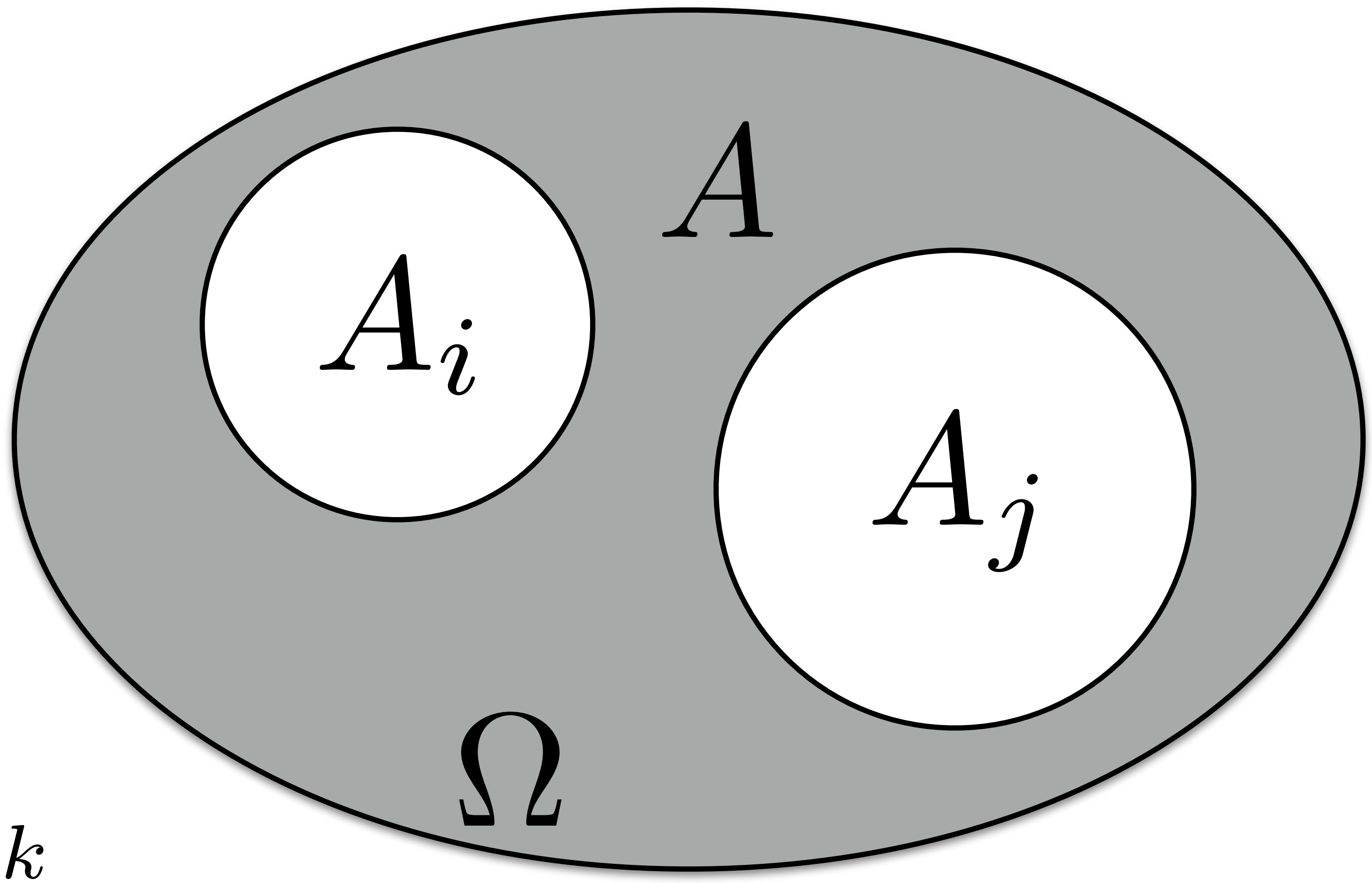


$$A_i \cap A_j = \emptyset$$

A_i and A_j are said to be mutually exclusive

Union of Mutually Exclusive Events

$$A = A_i \cup A_j$$



A includes all elements ω_k that belong to A_i or A_j

A is the <u>union</u> of A_i and A_j

Addition Law (I)

The probability of the union of two mutually exclusive events

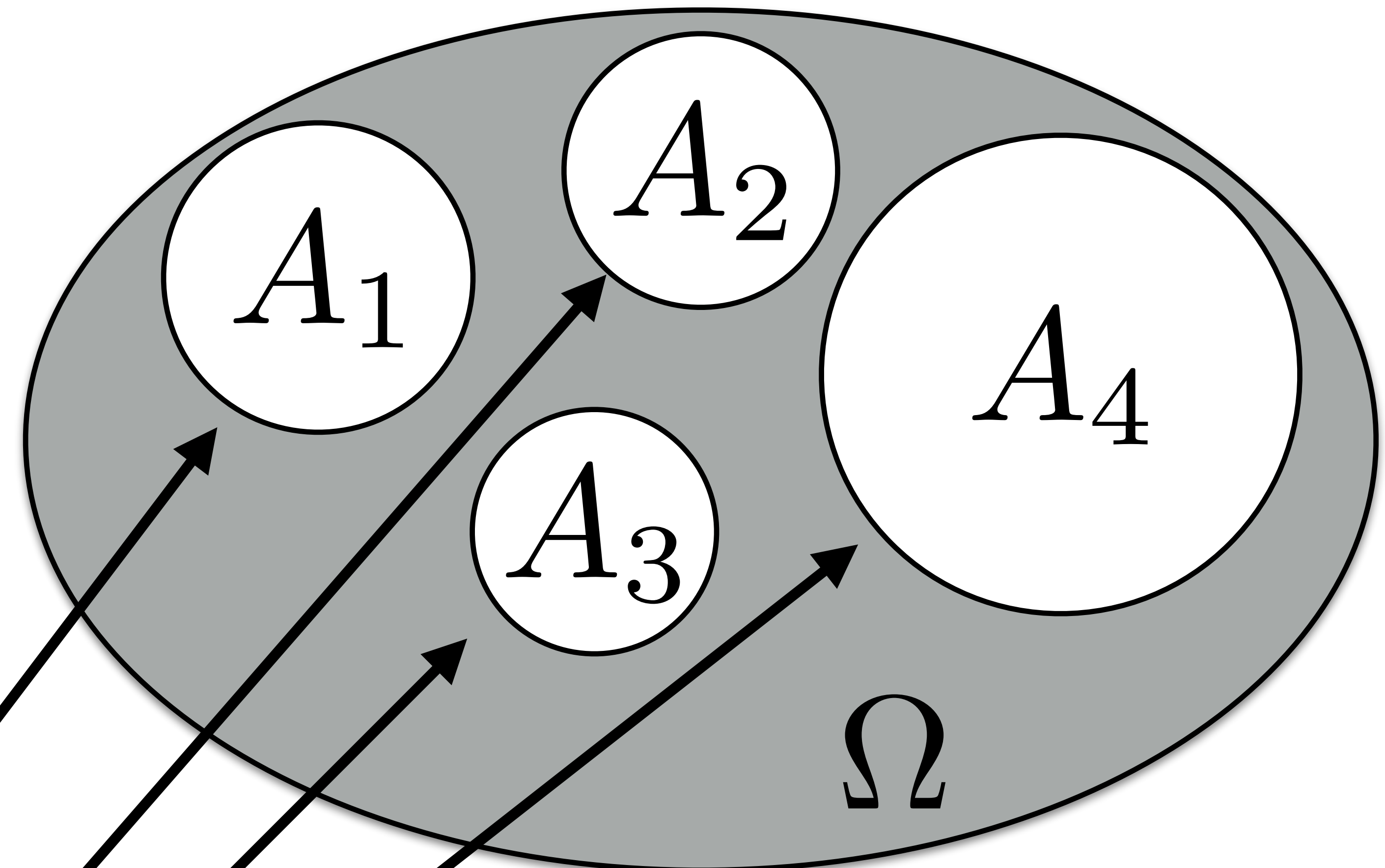
$$p(A) = \lim_{n \to \infty} \left[\frac{n(A_i)}{n} + \frac{n(A_j)}{n} \right] =$$

$$= p(A_i) + p(A_j)$$

The sum of the probabilities of the two mutually exclusive events

Union of Mutually Exclusive Events

$$A = \bigcup_{n=1}^{N} A_n$$



An event A can be thought of as the union of N mutually exclusive events

Addition Law (II)

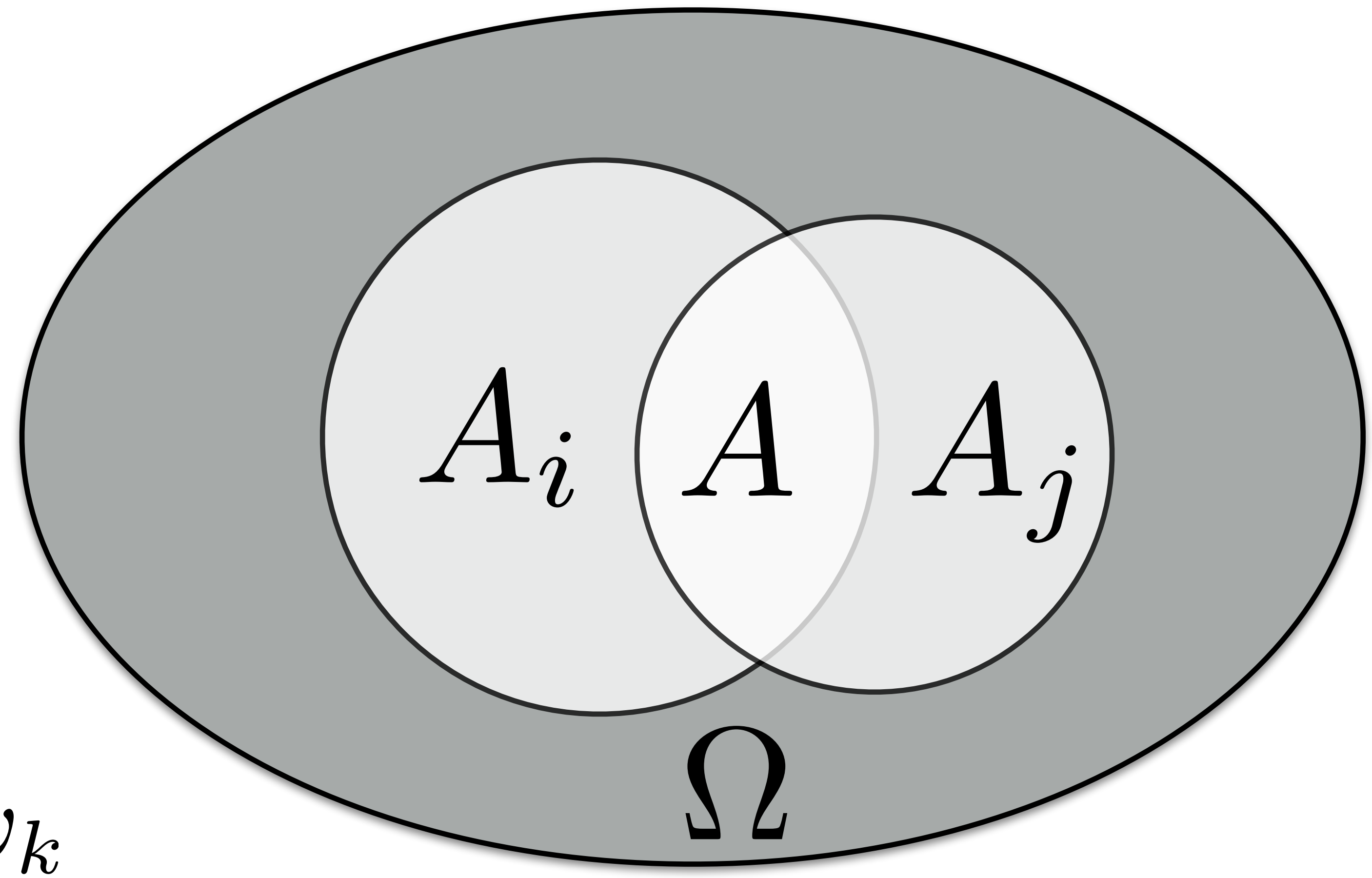
A is the union of N mutually exclusive events

$$p(A) = p\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} p(A_n)$$

p(A) is the sum of the probabilities of the mutually exclusive events

Intersection of Multiple Events

$$A = A_i \cap A_j$$



A includes all elements ω_k that belong to both A_i and A_j

A is the intersection of A_i and A_j

Conditional Probability

The probability of A given B

Probability of the intersection of A and B

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

Probability of B

Product Law

The probability of the intersection

$$p(A \cap B) = p(A, B) = p(A|B)p(B)$$

The joint probability of A and B

The product comes from the definition of the conditional probability

The intersection is a subset of the two sets

$$A \cap B \subseteq B \Rightarrow 0 \leq p(A|B) \leq 1$$

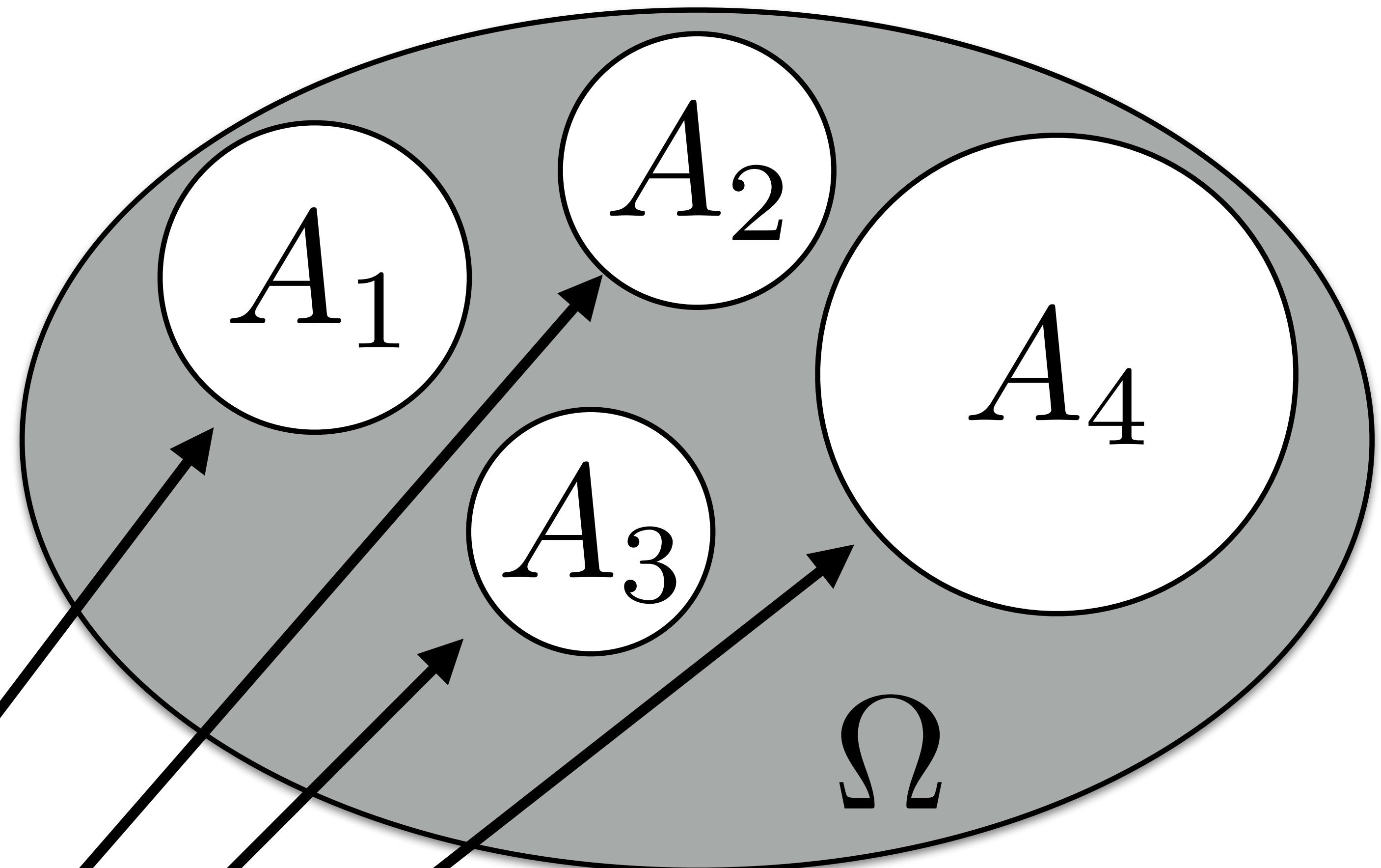
$$B \subset A \Rightarrow p(A|B) = 1$$

$$A \cap B = \emptyset \Rightarrow p(A|B) = 0$$

The intersection is empty

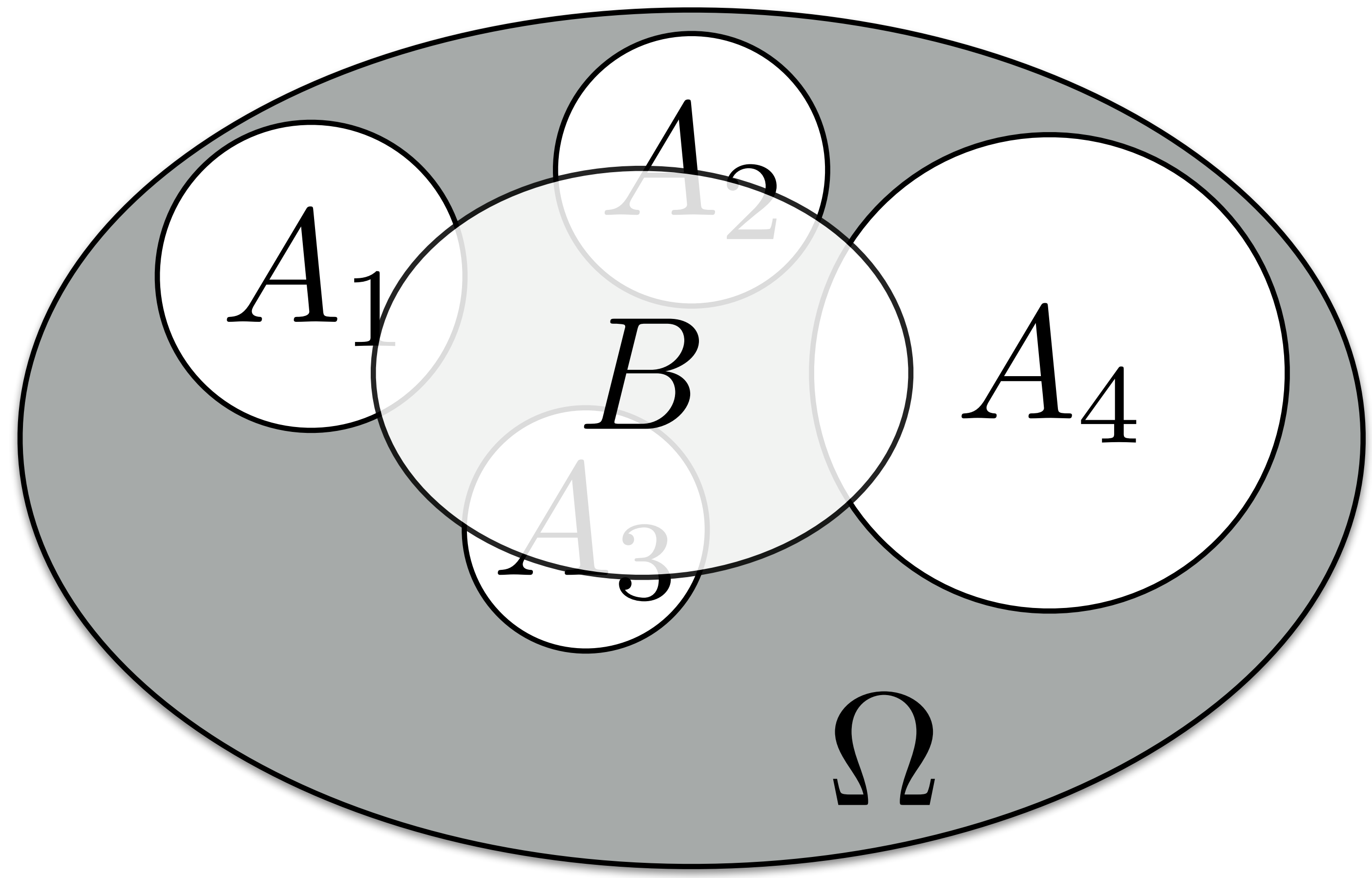
Union of Mutually Exclusive Events

$$A = \bigcup_{n=1}^{N} A_n$$



An event A can be thought of as the union of N mutually exclusive events

Union of Mutually Exclusive Events



The intersection of A and B is the union of the intersections

$$A \cap B = \bigcup_{n=1}^{N} A_n \cap B$$

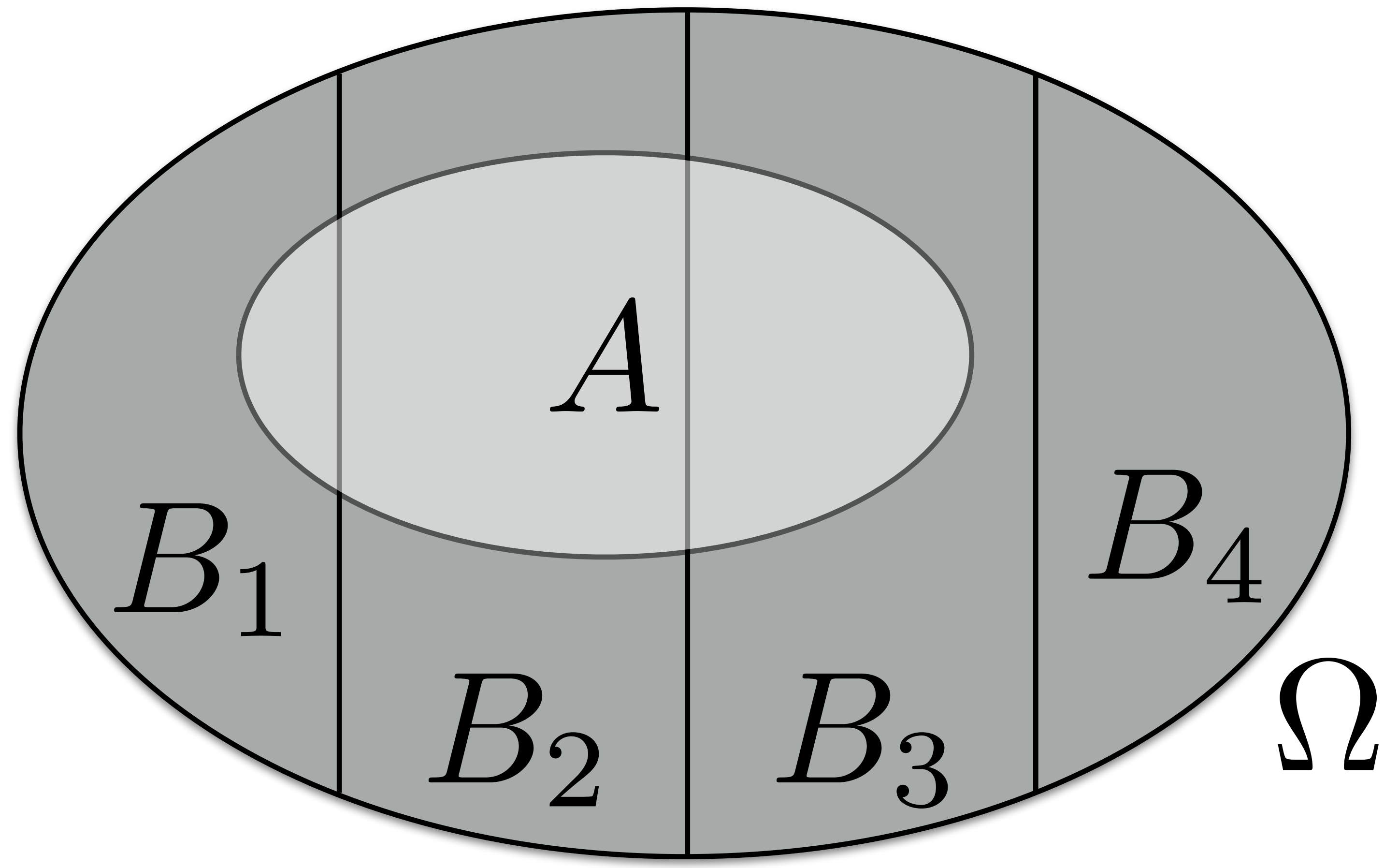
Addition Law for Conditional Probability (I)

$$p(A|B) = \sum_{n=1}^{N} \frac{p(A_n \cap B)}{p(B)} =$$

$$= \sum_{n=1}^{N} p(A_n|B)$$

Union of Mutually Exclusive Events

$$\Omega = \bigcup_{n=1}^{N} B_n$$



The sample space is the union of N mutually exclusive events

Addition Law for Conditional Probability (II)

$$p(A) = \sum_{n=1}^{N} p(A \cap B_n) =$$

$$= \sum_{n=1}^{N} p(A|B_n)p(B_n)$$

When there are two dices, the sample spaces include the same events, but they are different

$$\Omega^{(1)} = \{\omega_1^{(1)}, \ldots, \omega_6^{(1)}\}$$

$$\Omega^{(2)} = \{\omega_1^{(2)}, \ldots, \omega_6^{(2)}\}$$

$$A_1 = \{\omega_i^{(1)}, \ldots, \omega_j^{(1)}\}$$

$$A_2 = \{\omega_m^{(2)}, \ldots, \omega_n^{(2)}\}$$

The events build upon different sample spaces

The definition of probability

$$p(A_1, A_2) = \lim_{n\to\infty} \frac{n(A_1, A_2)}{n} \simeq$$

$$\simeq \lim_{n\to\infty} \frac{n(A_1, A_2)}{n(A_2)} \simeq p(A_1)$$

The number of times A2 occurs tends to infinite when n does

<u>If the occurrence of A2 does not influence the occurrence of A1</u>

Statistical Independence

$$p(A_1, A_2) = \lim_{n \to \infty} \frac{n(A_1, A_2)}{n} \simeq$$

$$\simeq \lim_{n \to \infty} \frac{n(A_1, A_2)}{n(A_2)} \frac{n(A_2)}{n} \simeq p(A_1)p(A_2)$$

Multiply and divide by n(A2)

<u>Only if the occurrence of A2 does not influence the occurrence of A1 and vice versa</u>

Recap

- The probability is an estimate of how frequently an event occurs as an outcome of a statistical experiment;
- Set theory allows one to demonstrate the basic laws of probability (addition and product);
- Statistical independence allows one to write the joint probability of several events as the product of the individual event probabilities.

Outline

- Definition of Probability
- Basic Laws of Probability
- Random Variables
- Conclusions

A variable is said to be random when it is related to a statistical experiment

The value of a random variable depends on the outcome of the statistical experiment

$$\xi = \xi(\omega)$$

In the case of a dice, the outcome is the face and the value is the number “written” on it

A variable is discrete when its value belongs to a finite set X

If a value is not in X, its probability is null

$$\xi \in X = \{x_1, x_2, \ldots, x_T\}$$

$$P(\xi = x_j) = p(x_j)$$

It is possible to estimate the probability that the variable takes a certain value in X

The probability distribution is a function that associates value and probability

The sum goes over all values of the discrete random variable

This is a constraint that must be respect for p to be a distribution

$$\sum_{k=1}^{T} p(x_k) = 1$$

The probability of a
variable being lower
than a certain value

The Addition Law allows
the following estimate

$$P(\xi \leq x_j) = \sum_{x_k \leq x_j} p(x_k)$$

$$F(x_j) = P(\xi \leq x_j)$$

The cumulative
probability function

The value of the
cumulative probability
function is a probability

A variable is continuous when its value belongs to a continuous interval

$$\xi \in [a, b]$$

$$P(x_1 \leq \xi \leq x_2) = \int_{x_1}^{x_2} p(x)dx$$

It is possible to estimate the probability of the value falling between two extremes

The function p(x) is called probability density function

The integral goes over all values of the continuous random variable

This is a constraint that must be respect for p to be a probability density function

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

The probability of a variable being lower than a certain value

The probability density function allows the following estimate

$$F(x_1) = \int_{-\infty}^{x_1} p(x)dx$$

$$F(x_1) = P(\xi \leq x_1)$$

The cumulative probability function

The value of the cumulative probability function is a probability

Recap

- The value of the random variables depends on an underlying event;
- When a random variable is discrete, it is possible to define a probability distribution;
- When a random variable is continuous, it is necessary to adopt a probability density function.

Outline

- Definition of Probability
- Basic Laws of Probability
- Random Variables
- Conclusions

Conclusions

- The probability is an estimate of how frequently an event takes place over a sufficient number of attempts;
- Discrete variables have a distribution, continuous variables have a probability density function.

Thank You!