

# Basics of Probability

Carlo Cavicchia

[carlo.cavicchia@uniroma1.it](mailto:carlo.cavicchia@uniroma1.it) 

# Probability

We call a phenomenon **random** if we are **uncertain** about its outcome.

Probability allows us to deal with randomness, by **quantifying uncertainty** and measuring the chances of the possible outcomes.

Typically the randomness we have to deal with comes from the **sampling procedure**: when we observe data, their values depends on the units we randomly select.

# Examples of random phenomena

- > the moment it will first start raining
- > the result of a football match
- > tomorrow's price of a stock
- > the number of tweets Trump is going to write today
- > ...

# The basic ingredients

We call a phenomenon “random” if we are uncertain about its outcome. It is characterized by:

- › **Sample Space:** the set of all possible outcomes. Its elements are exhaustive (no possible is left out) and mutually exclusive (only one outcome can occur).
- › **Event:** a subset of the sample space corresponding to one or more possible outcomes
- › **Probability:** measure of how likely each element of the sample space is.

# The basic ingredients - an evergreen Example

**Random phenomenon:** throw of a die

- › **Sample Space:** all the possible outcomes

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

- › **Event:** “the die returns an even number”

$$E = \{2, 4, 6\}$$

- › **Probability:**

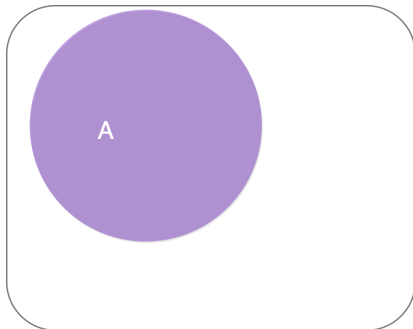
$$P(E) = 1/2$$

# Exercises

- › Two coins are tossed. Note: Each coin has two possible outcomes H (Heads) and T (Tails).
  1. Get the sample space.
  2. Find the probability that two heads are obtained.
  
- › A card is drawn at random from a deck of cards. Find the probability of getting a diamond.

# Recap of Set theory

- > **Complement** ( $A^c$  or  $\bar{A}$ ) everything that is not in  $A$ .

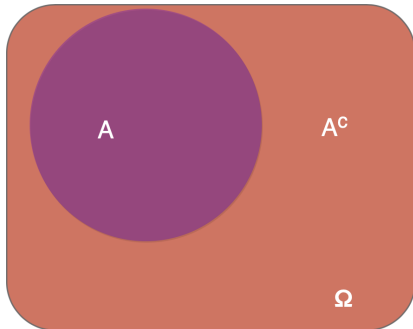


**Example**  $A =$  “the die returns an even number”

$\Rightarrow A^c =$  “the die returns an odd number”

## Recap of Set theory

- › **Complement** ( $A^c$  or  $\bar{A}$ ) everything that is not in  $A$ .

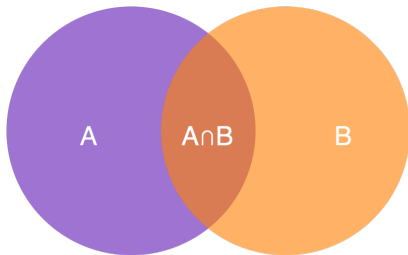


**Example**  $A$  = “the die returns an even number”,  $A^c$  = “the die returns an odd number”



# Recap of Set theory

- › **Intersection** ( $A \cap B$ ) given two events  $A, B$ , everything that is in *both*  $A$  and  $B$ .



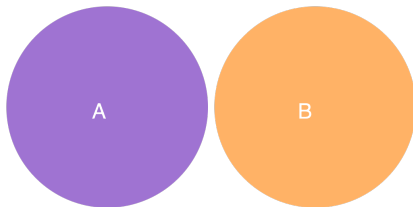
Tradition

**Example**  $A =$  “the die returns an even number”,  $B =$  “the die returns a number smaller than 5”

$$\Rightarrow A \cap B = \{2, 4\}$$

# Recap of Set theory

- > **Intersection** ( $A \cap B$ ) given two events  $A, B$ , everything that is in *both*  $A$  and  $B$ .



TwoDisjoint

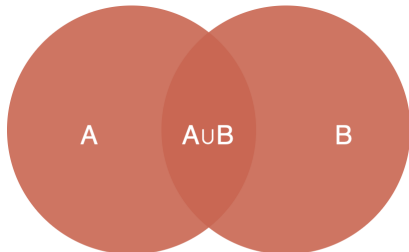
**Example**  $A =$  “the die returns an even number”,  $B =$  “the die returns a 5”

$\Rightarrow A \cap B = \emptyset$

$A$  and  $B$  are **disjoint**

# Recap of Set theory

- > **Union** ( $A \cup B$ ) given two events  $A, B$ , everything that is in *either*  $A, B$  or *in both*.



**Example**  $A =$  “the die returns an even number”,  $B =$  “the die returns a 5”

$$\Rightarrow A \cup B = \{2, 4, 5, 6\}$$

## Exercise :: challenging

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

- > Which is more probable?

Linda is a bank teller.

Linda is a bank teller and is active in the feminist movement.

# Some useful relationships

## Probability Axiomes...

- >  $0 \leq P(A) \leq 1$
- >  $P(\Omega) = 1$
- >  $P(\emptyset) = 0$

# Some useful relationships

## Probability Axiomes...

- ›  $0 \leq P(A) \leq 1$
- ›  $P(\Omega) = 1$
- ›  $P(\emptyset) = 0$

## ... and some trivial consequences

- ›  $P(A^c) = 1 - P(A)$
- ›  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

# Some useful relationships

## Probability Axiomes...

- ›  $0 \leq P(A) \leq 1$
- ›  $P(\Omega) = 1$
- ›  $P(\emptyset) = 0$

## ... and some trivial consequences

- ›  $P(A^c) = 1 - P(A)$
- ›  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

if  $A, B$  are disjoint then  $P(A \cup B) = P(A) + P(B)$

# Exercises

- › Which of the following is an impossible event?
  1. Choosing an odd number from 1 to 10.
  2. Getting an even number after rolling a single 6-sided die.
  3. Choosing a white marble from a jar of 25 green marbles.
  4. None of the above.
  
- › There are 4 parents, 3 students and 6 teachers in a room. If a person is selected at random, what is the probability that it is a teacher or a student?
  1.  $4/13$
  2.  $7/13$
  3.  $9/13$
  4. None of the above.



# How do we define probability?

something we don't really focus about

- › Classical approach: Assigning probabilities based on the assumption of equally likely outcomes.
- › Empirical approach: Assigning probabilities based on experimentation or historical data.
- › Subjective approach: Assigning probabilities based on the assignor's judgment.

Regardless of the approach we follow, **probability is a measure of uncertainty**, i.e. it quantifies how much we do not know, hence it strongly **depends on the information available** to us.

# Exercise

- › Get the probability of a Italian newborn is a girl.

# Exercise

- › Get the probability of a Italian newborn is a girl.

1/2... isn't it?

# Exercise

- › Get the probability of a Italian newborn is a girl.

1/2... isn't it?

But, if we know that the 51.3% of Italian people are women? What does it happen to that probability?

# Conditional Probability

Probability is a measure of uncertainty on the result of a random experiment, so **any additional information** on the outcome **affects it**.

Let  $A$  and  $B$  be two events, if we know that  $B$  happened, we can update the probability of  $A$  as follows:

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

Example: if we know that a die returned an *even* number, then the probability of observing a 3 is 0.

# Independence

absence of relation between events

If knowing an event  $B$  does not affect our probability evaluation of  $A$ , then we say that  $A$  and  $B$  are **independent**:

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

Combining this to the definition of conditional probability we can derive the **factorization criterion**, to assess if two events are independent:

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

**CAVEAT:** The fact that two events are independent **does not mean they are disjoint**, and actually this is almost never the case. In fact for  $P(A \cap B) = P(\emptyset) = 0 = P(A)P(B)$ , either  $A$  or  $B$  must have probability 0.

# Exercises

- › Three cards are chosen at random from a deck without replacement. What is the probability of choosing an eight, a seven and a six, in order?

$6/35152$

$1/2197$

$8/16575$

None of the above.

- › A jar contains 5 red, 3 green, 2 purple and 4 yellow marbles. A marble is chosen at random from the jar. After replacing it, a second marble is chosen. What is the probability of choosing a purple and then a red marble?

$5/98$

$1/2$

$3/98$

$2/49$

# Random variable

how to define it

Typically we are not interested in the single outcome itself or in the events but in a *function* of them.

A **random variable** is any function from the sample space to the real numbers.

Examples:

- › toss a coin three times and **count** the number of tails
- › roll two dice and **sum** the values of the faces

**NB** A random variable is a *number*: we can do all sorts of operations with it!



# Random variables

how to characterize it

- ›  $X$  *random variable*: the random function (before it is observed!)
- ›  $x$  *realization of the random variable*: the number we get after we observe the result of the random experiment
  
- ›  $\mathcal{X}$  *support of the random variable*: all the possible values assumed by  $X$

Example:

- › toss a three coins.  $X$  is the number of tails  
 $\mathcal{X} = \{0, 1, 2, 3\}$

Probability statement on a random variable can be derived from the probability on the basic events!

# Distribution of a random variable

an example of how to derive it

- › Toss a coin three times.  $X$  is the random variable representing the *number of Tails*

$\omega$	$P(\omega)$	$x$
HHH	$1/8$	0
THH	$1/8$	1
HTH	$1/8$	1
HHT	$1/8$	1
TTH	$1/8$	2
THT	$1/8$	2
HTT	$1/8$	2
TTT	$1/8$	3

$x$	$p_x = P(X = x)$
0	$1/8 \times 1 = 1/8$
1	$1/8 \times 3 = 3/8$
2	$1/8 \times 3 = 3/8$
3	$1/8 \times 1 = 1/8$

The distribution of a random variable  $p_x$  is just a convenient way of summarizing single outcomes probabilities.

# Exercise

› Two dice are rolled:

1. Construct the sample space. How many outcomes are there?
2. Find the probability of rolling a sum of 7.
3. Find the probability of getting a total of at least 10.
4. Find the probability of getting an odd number as the sum.

# Distribution of a Discrete Random Variable:

When  $\mathcal{X}$  is countable, the random variable  $X$  is said to be **discrete**, and it is characterized by:

- › **Probability mass distribution**

$$p_x = P(X = x) \quad \forall x \in \mathcal{X}$$

- › **Cumulative distribution function**

$$F_X(x) = P(X \leq x) = \sum_{y \leq x} P(X = y) = \sum_{y \leq x} p_y$$

Examples:

- › What is the probability of **exactly** 1 head?  $P(X = 1) = p_1 = 3/8$
- › What is the probability of **at most** two heads?

$$P(X \leq 2) = F_X(2) = p_0 + p_1 + p_2 = 7/8$$

## Remark

an ode to recycling

**Remember:** statements such as  $X = 1$  or  $X \leq 2$  are **events**, we can use *intersection, union, complement* and all the operations we have seen before!

Examples:

- › What is the probability of **note getting** 1 head?

$$P(X \neq 1) = P((X = 1)^c) = 1 - P(X = 1) = 5/8$$

- › What is the probability of **at least** 2 heads?

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - F_X(1) = 1 - (p_0 + p_1) = 4/8$$

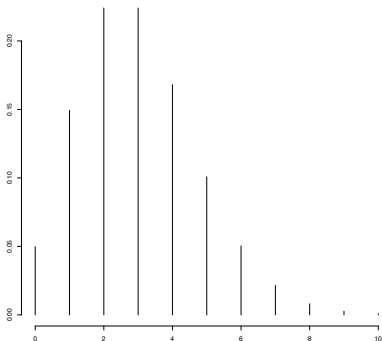
- › What is the probability of **0 or** 2 heads? (*disjoint events!*)

$$P(X = 2 \cup X = 0) = P(X = 2) + P(X = 0) = p_0 + p_2 = 4/8$$

# Properties

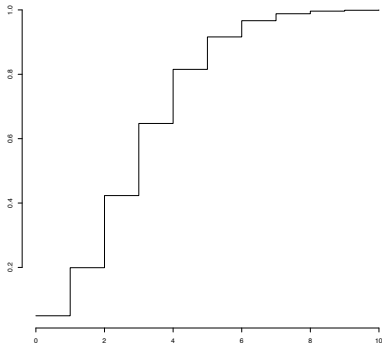
## Probability mass distribution

- >  $p_x \geq 0$
- >  $p_x \leq 1$
- >  $\sum_x p_x = 1$



## Cumulative distribution function

- >  $0 \leq F(x) \leq 1$
- >  $F$  is non-decreasing
- >  $F$  is right continuous



# Exercise

- › Let  $X$  be a discrete random variable with the following probability distribution

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1/5 & \text{if } 1 \leq x < 4 \\ 3/4 & \text{if } 4 \leq x < 6 \\ 1 & \text{if } x \geq 6 \end{cases}$$

- › Calculate the probability function.
- › Calculate the following probabilities
- $P(X=6)$
  - $P(X=5)$
  - $P(2 < X < 5.5)$
  - $P(0 \leq X < 4)$

# Distribution of a Continuous Random Variable:

When  $\mathcal{X}$  is **not** countable, the random variable  $X$  is said to be **continuous**.

If  $\mathcal{X}$  is not countable, it is not possible to put mass on any value  $x \in \mathcal{X}$ , meaning that

$$P(X = x) = 0 \quad \forall x \in \mathcal{X}$$

› **Cumulative distribution function**

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt \quad \forall x \in \mathcal{X}$$

› **Probability density distribution**

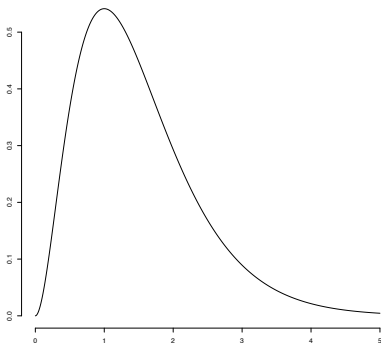
$$f_X(x) = \frac{dF_X(x)}{dx} \quad \forall x \in \mathcal{X}$$



# Properties

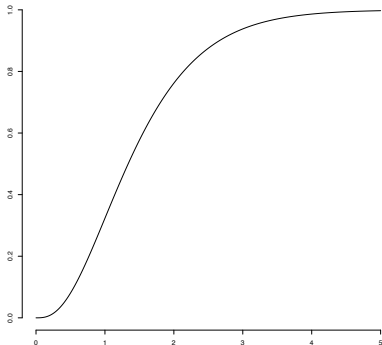
## Probability density function

- >  $f_X(x) \geq 0$
- >  $f_X(x)$  **needs not** be  $\leq 1$
- >  $\int_{-\infty}^{\infty} f_X(x) dx = 1$



## Cumulative distribution function

- >  $0 \leq F(x) \leq 1$
- >  $F$  is *non-decreasing*
- >  $F$  is *right continuous*



# Exercise

- › Let  $X$  be a continuous random variable with the following probability distribution

$$f(x) = \begin{cases} cx^2(1-x) & \text{if } 0 < x \leq 1 \\ 0 & \text{if otherwise} \end{cases}$$

- › Determine  $c$  so that  $f(x)$  is a valid pdf.

# Comparison

- ›  $X$  discrete rv with pmf  $p_x$
- ›  $P(X \in A) = \sum_{x \in A} p_x$

if  $A = \{x_1, \dots, x_k\}$  then

$$P(X \in A) = \sum_{i=1}^k p_{x_i}$$

- ›  $X$  continuous rv with pdf  $f_X(x)$
- ›  $P(X \in A) = \int_A f_X(x) dx$

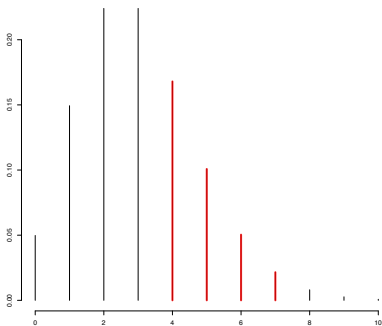
if  $A = [a, b]$  then

$$\begin{aligned} P(X \in A) &= \int_a^b f_X(x) dx \\ &= F_X(b) - F_X(a) \end{aligned}$$

# Comparison

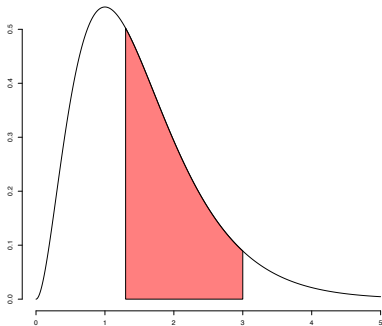
$$A = \{x_1, \dots, x_k\}$$

$$P(X \in A) = \sum_{i=1}^k p_{x_i}$$



$$A = [a, b]$$

$$P(X \in A) = F_X(b) - F_X(a)$$



# Summaries

The distribution of a random variables provides fully characterize it, but it may not be “immediate” to gain insights from it.

Once more we need to summarize the information contained in the distribution.

## Candidates:

- › *Mode*: the value that is “more likely”, i.e. the value that maximises the density
- › *Median*: the value that “splits in half” the distribution, i.e.  $m$  s.t.

$$P(X \leq m) = P(X > m) = 0.5$$

# Expected Value

king of all summaries

The **Mean** or **Expected Value** is the “average” of the elements in the support of  $X$ , weighted by the probability of each outcome.

The expected value gives a rough idea of what to expect for the **average** of the observed values in a **large repetition** of the random experiment (*not what we'll observe in a single observation!*)

$X$  discrete r.v. with p.m.f.  $p_x$

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} xp_x$$

$X$  continuous r.v. with p.d.f.  $f_X(x)$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf_X(x)dx$$

**WATCH OUT** The expected value *may not exist!*

# Properties

›  $\mathbb{E}[c] = c$  for any constant  $c$

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$$

›  $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

›  $\mathbb{E}[X - \mathbb{E}[X]] = 0$

›  $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 0$

**The Law of the Lazy Statistician** Given a continuous (respectively discrete) random variable  $X$  whose expectation exists, and a function  $g$ , then

$$\mathbb{E}[g(X)] = \int g(x)f_X(x)dx \quad \left( \mathbb{E}[g(X)] = \sum_x g(x)p_x \right)$$

# Measuring Variability

The expected value gives an idea about the **center** of the distribution, but does not account for the dispersion of the values

Example:

- › Given two investment strategies with the same expected payout, we would like to choose the one with less variability

(Bad) Candidates:

- › average deviation from the mean  $\mathbb{E}[X - \mathbb{E}[X]]$  (**not informative**)
- › average absolute deviation from the mean  $\mathbb{E}|X - \mathbb{E}[X]|$  (**computationally challenging**)



# Variance

queen of all summaries

The **variance** of a random variable  $X$

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

tells us of **how much** the variable oscillates around the mean.

$X$  discrete r.v. with p.m.f.  $p_x$

$$\mathbb{V}[X] = \sum_{x \in \mathcal{X}} (x - \mathbb{E}[X])^2 p_x$$

$X$  continuous r.v. with p.d.f.  $f_X(x)$

$$\mathbb{V}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x) dx$$

# Properties

- > the variance is always **non-negative**,  $\mathbb{V}[X] \geq 0$  and is 0 only when  $X$  is constant
- > the square root of the variance  $\text{sd}(X) = \sqrt{\mathbb{V}[X]}$  is called **standard deviation**. Roughly,  $\square$  describes how far values of the random variable fall, on the average, from the expected value of the distribution
- > the variance is *insensitive to the location* of the distribution but **depends only on its scale**

$$\mathbb{V}[aX + b] = a^2\mathbb{V}[X]$$

- > a **computation-friendlier** alternative definition of the variance is:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

# Exercise

- › Let  $X$  be a discrete random variable with the following probability distribution

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1/5 & \text{if } 1 \leq x < 4 \\ 3/4 & \text{if } 4 \leq x < 6 \\ 1 & \text{if } x \geq 6 \end{cases}$$

- › Calculate mean and variance

# Exercise

- › Let  $X$  be a continuous random variable with the following probability distribution

$$f(x) = \begin{cases} cx^2(1-x) & \text{if } 0 < x \leq 1 \\ 0 & \text{if otherwise} \end{cases}$$

- › Calculate mean and variance

# Covariance

If we have 2 random variables, the **covariance** gives us a measure of association between them.

$$\begin{aligned}\mathbb{C}ov(X, Y) &= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \\ &= \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y\end{aligned}$$

- › The sign of  $\mathbb{C}ov(X, Y)$  informs on the nature of the association
- › The higher  $|\mathbb{C}ov(X, Y)|$ , the stronger the association

**Remark**  $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\mathbb{C}ov(X, Y)$

# Independence of Random Variables

Two random variables  $X, Y$  are independent if

$$\begin{aligned}F_{X,Y}(x, y) &= P(X \leq x \cap Y \leq y) \\ &= P(X \leq x)P(Y \leq y) \\ &= F_X(x)F_Y(y) \quad \forall x, y \in \mathbb{R}\end{aligned}$$

Intuitively if  $X$  and  $Y$  are independent, the value of one does not affect the value of the other.

**Remark:** if  $X_1, \dots, X_n$  are independent then

- $\triangleright p_{x_1, \dots, x_n} = p_{x_1} \cdot \dots \cdot p_{x_n}$
- $\triangleright f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$

# Independence of Random Variables

## Factorization Criterion

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \forall x,y \in \mathbb{R}$$

If  $X$  and  $Y$  are independent then  $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$

As a consequence

$$\mathbb{C}ov(X,Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y = 0$$

**WATCH OUT:** the converse is not true! If  $\mathbb{C}ov(X,Y) = 0$ , the two random variables may still be associated.

## Exercise [1/2]

- › Let  $X$  and  $Y$  be two random variables with marginal distribution functions

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - \exp(-x) & \text{if } x \geq 0 \end{cases}$$

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 - \exp(-y) & \text{if } y \geq 0 \end{cases}$$



## Exercise [2/2]

- › Determine if  $X$  and  $Y$  are independent given the joint distribution function:

$$F_{X,Y}(x,y) = \begin{cases} 0 & \text{if } x < 0 \text{ or } y < 0 \\ 1 - \exp(-x) - \exp(-y) + \exp(-x - y) & \text{if } x \geq 0 \text{ and } y \geq 0 \end{cases}$$