Statistics Pre-course

Part 2

Fundamentals of Probability

Alfonso Russo

Department of Economics and Finance

Tor Vergata University of Rome

September 2024

Part II Syllabus

1. Basic definitions and recap of Set Theory
2. Random Variables
3. Discrete Probability Distributions
4. Continuous Probability Distributions
5. Expected value and Variance of a Random Variable
6. Main Probability Distributions (Bernoulli, Binomial, Poisson, Uniform, Normal, Exponential, Student-t ...)
7. Basics of Asymptotics (Central Limit Theorem, Law of Large Numbers)

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 possible outcomes

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

Examples of random phenomena

- The moment when it will first start rain tomorrow
- The number of tweets Trump is going to post tomorrow
- The result of a football match
- Tomorrow's price of a stock
- ...

There follows some basic definitions we are going to use in dealing with randomness

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

An evergreen (albeit boring) example

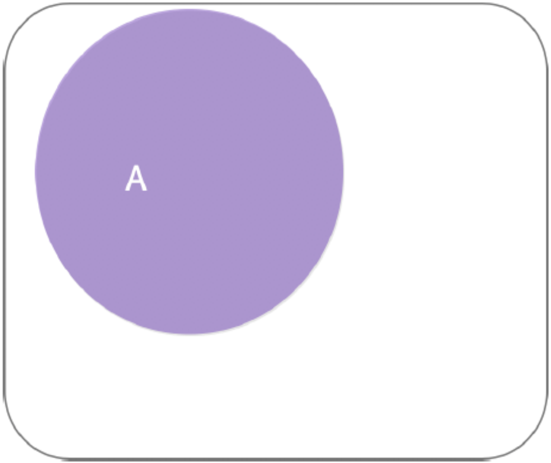
Random phenomenon: throw of a fair die

- **Event space:** all of the possible outcomes
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
- **Event**: "the die returns an even number"
  - $E = \{2, 4, 6\}$
- **Probability**:
  - $\mathbb{P}(E) = \frac{3}{6} = \frac{1}{2}$

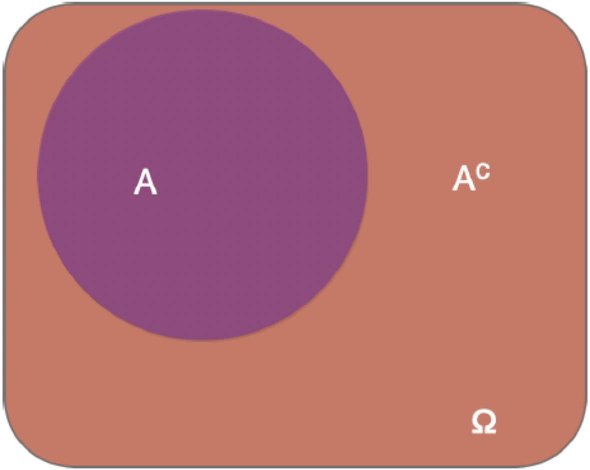
Set Theory

- **Events** are mathematically treated as **Sets**.
- Sets can be **finite** (contain a finite number of objects) or **infinite** (consist of infinite elements).
- The **cardinality** of a given set is the measurement of objects that the set contains. E.g. if $E = \{1, 2, 3\}$ then the cardinality of E, denoted as $\#E = 3$.

Consider a generic set A included in an event space Ω

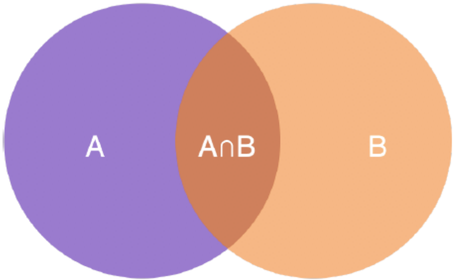


Complement: (A^c or $\bar{A}$) everything that is not in A



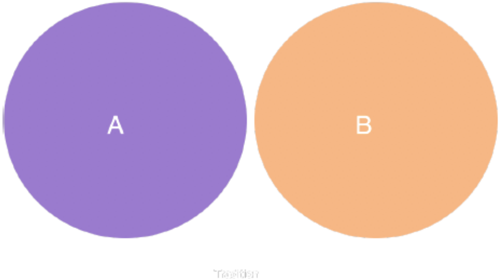
Example: $A =$ "the die returns an even number"; $A^c =$ "the die returns an odd number"

Intersection: $(A \cap B)$ everything that is **both** in A and B



Example: $A =$ "the die returns an even number"; $B =$ "the die returns a number less than 5" $\implies A \cap B = \{2, 4\}$

Intersection: $(A \cap B)$ everything that is **both** in A and B

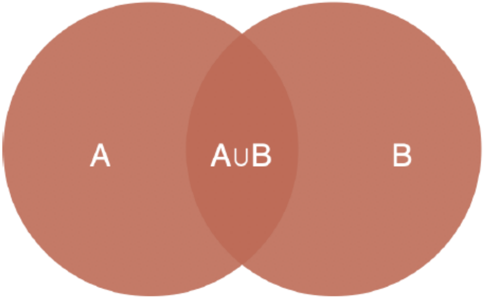


Example: $A =$ "the die returns an even number"; $B =$ "the die returns a 5"

$$\implies A \cap B = \emptyset$$

A and B are **disjoint**

Union: $(A \cup B)$ everything that is **either** in A in B or both



Example: $A =$ "the die returns an even number"; $B =$ "the die returns a 5"

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

Probability Axioms

and some trivial consequences

Given a generic set A in an event space Ω

- $0 \leq \mathbb{P}(A) \leq 1$
- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(\emptyset) = 0$

As a consequence

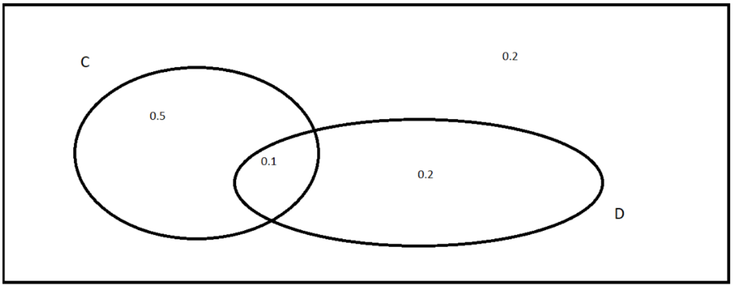
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- If A and B are disjoint then $\mathbb{P}(A \cap B) = 0$. Hence $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Exercise

- In a sample of 100 college students, 60 said that they own a car, 30 said that they own a stereo, and 10 said that they own both a car and a stereo.
- Compute the probability that a student owns a car but **not** a stereo.
- Compute the probability that a student owns either a car **or** a stereo.
- Depict this information on a Venn diagram.

Solution

- Let C representing the event "the student owns a car". Let D be the event "the student owns a stereo".
- We know that $\mathbb{P}(C) = 0.6$, $\mathbb{P}(D) = 0.3$ and $\mathbb{P}(C \cap D) = 0.1$.
- $\mathbb{P}(\text{"car but NOT stereo"}) = \mathbb{P}(C) - \mathbb{P}(C \cap D) = 0.6 - 0.1 = 0.5$.
- $\mathbb{P}(\text{"car OR stereo"}) = \mathbb{P}(C \cup D) = \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(C \cap D) = 0.6 + 0.3 - 0.1 = 0.8$



How do we define probability?

- Classical approach: assigning probabilities based on the assumption of equally likely events
- Frequency approach: assigning probabilities as the limit of the relative frequency of the event assuming having observed infinite repetitions of the random experiment
- Subjective approach: assigning probabilities based on assignor's judgment or external information

Regardless of the followed approach, **probability is still a measure of uncertainty**. In other words, it quantifies how much we do not know and it **strongly depends on the information available** about the random phenomenon.

- The probabilistic relative frequency of an event's occurring is the proportion of times the event occurs over a given number of trials. If A is the event of interest, then the probabilistic relative frequency of A, denoted as $\mathbb{P}(A)$, is defined as

$$\mathbb{P}(A) = \frac{\text{number of occurrences}}{\text{number of trials}}$$

- Among the first 43 Presidents of the United States, 26 were lawyers. What is the probability of the event $A =$ "selecting a President who is also a lawyer"?

$$\mathbb{P}(A) = \frac{26}{43} \approx 0.605$$

Exercises

- Is there an intruder? Why?
  - Choosing at random an even number from 1 to 10.
  - Getting a diamond card from a deck of 52 cards.
  - Drawing a red ball from a jar of 500 blue balls.
  - Pick exactly our Sun at random from a jar with the names of all the Stars in the observable universe.
- In a room there are 6 volleyball players, 4 basketball players and 10 football players. If one of them is selected at random:
  - what is the probability that the selected one is an athlete?
  - what is the probability that the selected one is either a volleyball or a football player?
  - what is the probability that the selected one is not a basketball player?
- What is the probability that an Italian newborn is a girl?

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

- Let A and B be two events. If we knew that B happened, we could update the probability of A as follows

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{1}$$

Exercise: Back to the students, find the probability that a Student own a stereo given possession of a car.

If knowing about an event B does not affect our probability evaluation of another event A we say that A and B are **independent**.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \tag{2}$$

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

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A) \implies \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{3}$$

- **Problem I:** two coins are tossed. Each coin has two possible outcomes, head (H) and tail (T).
  - Determine the event space and its size
  - Find the probability of the event $A =$ "the faces appearing on the two coins are different"
  - Find the probability of the event $B =$ "the faces appearing on the two coins are two heads"

- **Problem II:** which of the following numbers cannot be a probability?
  1. 0.5
  2. -0.001
  3. 1
  4. 0
  5. 1.01

- **Problem III:** two fair dice are rolled. Find the probabilities of the following events
  - the sum is equal to 1
  - the sum is equal to 4
  - the sum is less than 13
- **Problem IV:** a card is drawn at random from a deck of 52 cards. Find the probabilities of the following events
  - the card is a 3 of diamond
  - the card is a queen

- To compute probabilities we might often need a method to assess in **how many different ways** a certain phenomenon can happen. E.g. "how many times will I obtain two Heads in two tosses of a coin?".
- The table of all the 4 possible outcomes.

1	H	H
2	H	T
3	T	H
4	H	H

- **Combinatorics** is a branch of mathematics that is about counting.

FUNDAMENTAL PRINCIPLE OF COMBINATORICS

If you have an experiment with n possible outcome and add a second experiment with m possible outcomes, then the combination of the two experiments has $n \times m$ possible outcomes.

- In the previous example: each coin toss has two possible outcomes; then two tosses of a coin have 2×2 possible outcomes.

- Imagine there are 9 students attending the Statistics course. Suppose further that there are 9 chairs available positioned on a straight line where the students can sit. How many different lines can be formed by changing the position of the students?

- The first student can choose his sit in 9 different ways, the second has 8 possible choices, the third can sit in only 7 alternative ways and so on. Therefore, there are

$$9 \times 8 \times 7 \times 6 \times \ldots \times 1 = 9!$$

possible ways to place 9 students on a line.

PERMUTATION OF SET ELEMENTS

Given a set of n elements, a given *ordering* of its components is a permutation.

There are $n!$ possible permutations of n elements.

- Suppose we want to place 23 Students on 23 chairs in a Maths class. If you have 4 classes a week and there are 52 weeks in one year, how long would it take to get through all the possible sit permutation?

PERMUTATION OF SET ELEMENTS

Given a set of n elements, a given *ordering* of its components is a permutation.

There are $n!$ possible permutations of n elements.

- Suppose we want to place 23 Students on 23 chairs in a Maths class. If you have 4 classes a week and there are 52 weeks in one year, how long would it take to get through all the possible sit permutation?

  Answer: 10 million times the current age of the Universe!

- Suppose again we have $n = 9$ Students but this time we have to place them on $k = 6$ chairs only. How many ways are there to dispose these students on the available chairs regardless of the ordering?

Combinations

Given a set of n elements, a combination is a subset of k elements chosen without repetition and regardless of their ordering. The number of possible combinations of k elements out of a total of n is given by

$$\frac{n!}{(n-k)!k!} = \binom{n}{k}$$

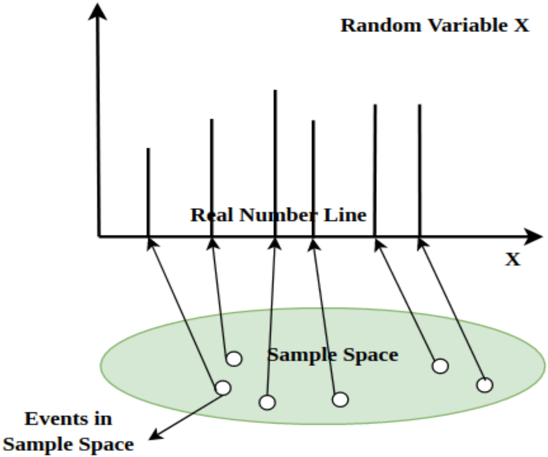
Typically, we are not interested in a single outcome or events themselves but in a *function* of them

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

- **Examples:**
  - Toss a coin three times and count the tails
  - Roll two dice and sum the values on the faces

Random Variables

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



- X the random variable: the random function before it is observed
- x a realization of the random variable: the number we observe
- $\mathcal{X}$ the support of the random variable: the set of the possible values that X can assume
- **Example:** toss a coin three times and count the number of heads
  - $\mathcal{X} = \{0, 1, 2, 3\}$

Toss a coin three times. X is the random variable representing the *number of tails*

The distribution of the random variable p_x is a just a convenient way to summarize outcomes probabilities.

Exercise

M&M sweets are of varying colours that occur in different proportions. The proportions are as follows:

blue $= 0.3$, red $= 0.2$, yellow $= 0.2$, green $= 0.1$, orange $= 0.1$, tan $= ?$

You draw an M&M at random from the package:

- Determine the value of the missing proportion
- Find the probability of getting either a blue or a red one
- Find the probability of getting one which is not yellow
- Find the probability of getting one which neither orange nor tan
- Find the probability of getting one which is either blue or red or yellow or orange or green or tan

Distribution of a Discrete Random Variable

discrete = how many

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

- **Probability mass function**

$$p_x = \mathbb{P}(X = x) \quad \forall\, x \in \mathcal{X} \tag{4}$$

- **Cumulative distribution function**

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{y \leq x} \mathbb{P}(X = y) = \sum_{y \leq x} p_y \tag{5}$$

Note: statements like $X = 1$ or $X \leq 2$ are *events* and we can use unions, intersections, complements are all the operations we have seen before!

Example

Consider the example of tossing a coin three times

- What is the probability of getting **exactly** 1 head? $p_1 = 3/8$
- What is the probability of getting **at most** 2 heads?

  $\mathbb{P}(X \leq 2) = F_X(2) = p_0 + p_1 + p_2 = 7/8$
- What is the probability of **not getting** 1 head?

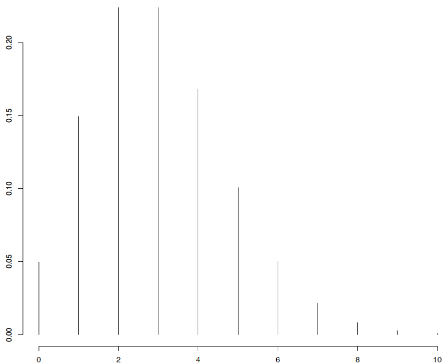
  $\mathbb{P}(X \neq 1) = \mathbb{P}[(X = 1)^c] = 1 - \mathbb{P}(X = 1) = 1 - p_1 = 5/8$
- What is the probability of **at least** 2 heads?

  $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - F_X(1) = 1 - (p_0 + p_1) = 4/8$
- What is the probability of getting **either** 0 **or** 2 heads?

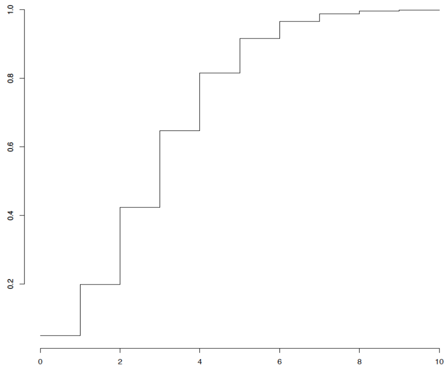
  $\mathbb{P}(X = 2 \cap X = 0) = \mathbb{P}(X = 2) + \mathbb{P}(X = 0) = p_2 + p_0 = 4/8$

Properties

- **Probability mass function**
  - $p_x \geq 0$
  - $p_x \leq 1$
  - $\sum p_x = 1$



- **Cumulative distribution function**
  - $0 \leq F(X) \leq 1$
  - $F(X)$ is non-decreasing
  - $F(X)$ is right-continuous



- A lottery is organised each year in Manchester. A thoudand tickets are sold at the price of 1£ each. Each ticket has the same probability of winning the lottery. First price is set at 300£, second price at 200£ and third price is 100£.
- Let $\mathcal{X}$ denote the gain from purchasing one ticket. Construct the distribution of $\mathcal{X}$. Find the probability of winning any money from the lottery.

Example

Suppose a random variable X has the following probability distribution

x	1	3	4	7	9	10	14	18
$\mathbb{P}(X = x)$	0.11	0.07	0.13	0.28	0.18	0.05	0.12	?

- Fill in the missing value
- Write down the distribution function
- Evaluate the following probabilities:
  - X is at least 10
  - X is more than 10
  - X is less than 4

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

If $\mathcal{X}$ is not countable, is not possible to put mass on any values of $\mathcal{X}$, meaning that

$$\mathbb{P}(X = x) = 0 \quad \forall\, x \in \mathcal{X} \tag{6}$$

Cumulative distribution function:

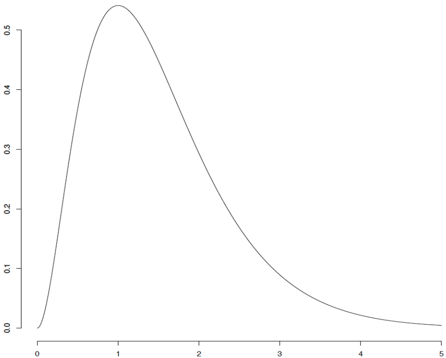
$$F_X(x) = \mathbb{P}(x \leq x) = \int_{-\infty}^{x} f_X(x)\, dx \quad \forall\, x \in \mathcal{X} \tag{7}$$

Probability density function:

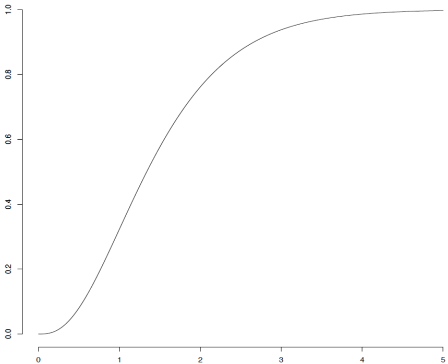
$$f_X(x) = \frac{\partial F_X(x)}{\partial x} \quad \forall\, x \in \mathcal{X} \tag{8}$$

Properties

- **Probability density function**
  - $f_X(x) \geq 0$
  - $\int_{-\infty}^{+\infty} f_X(x) = 1$



- **Cumulative distribution function**
  - $0 \leq F(X) \leq 1$
  - $F(X)$ is non-decreasing
  - $F(X)$ is right-continuous



Exercise

Let X be a continuous random variable with the following probability density function

$$f_X(x) = \begin{cases} cx(1-x) & if\, 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

- determine c such that this is a proper probability density function
- evaluate $\mathbb{P}(X = 0.5)$
- evaluate $\mathbb{P}\left(X \leq \frac{1}{2}\right)$

Exercise

Let Y be a continuous random variable with the following cumulative distribution function

$$F_Y(y) = \begin{cases} 1 & \text{if } y \geq 1 \\ 3y^2 - 2y^3 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

- evaluate $\mathbb{P}\left(Y \leq \frac{1}{2}\right)$ using $F_Y(y)$

Comparison

discrete vs continuous

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

If $A = \{x_1, \ldots, x_k\}$ then

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

- X continuous rv with pdf $f_X(x)$
- $\mathbb{P}(X \in A) = \int_A f_X(x)\, dx$

If $A = [a, b]$ then

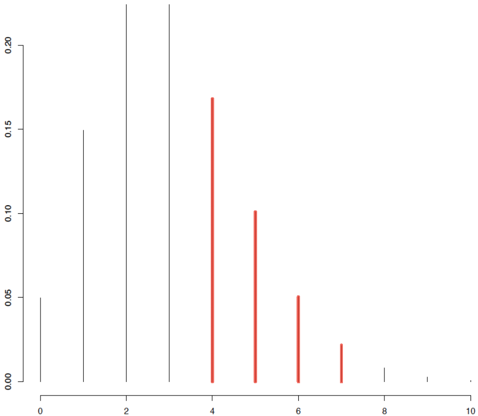
$$\mathbb{P}(X \in A) = \int_a^b f_X(x)\, dx = F_X(b) - F_X(a)$$

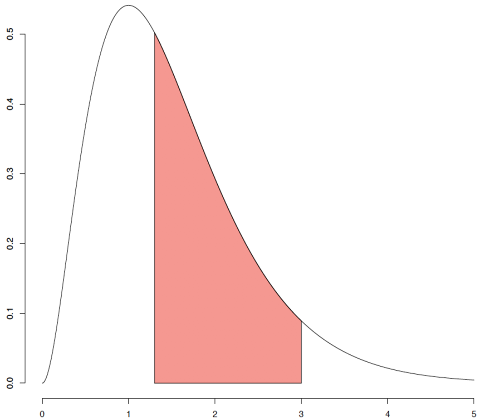
Comparison

discrete vs continuous

$$A = \{x_1, \dots, x_k\}$$
$$\mathbb{P}(X \in A) = \textstyle\sum_{i=1}^k p_{x_i}$$

$$A = [a, b]$$
$$\mathbb{P}(X \in A) = \textstyle\int_a^b f_X(x)\, dx$$





The distribution of a random variable fully characterize it but it may not be immediate to gain insight from it.

There is a bunch of alternatives to summarize the information contained in the distribution:

- **Mode:** the value that is the "most likely" (maximises the density)
- **Median:** the value that "splits in half" the distribution, denoted by m

$$\mathbb{P}(X \leq m) = \mathbb{P}(X > m) = 0.5 \tag{11}$$

Expected Value

the king of all summaries

The **Mean** or **Expected Value** is the "average" of the elements in the support of X, weighted by the probabilities of each outcome.

The Expected Value gives a rough idea of what to expect as the average of the observed outcomes in a **large repetition** of the random experiment (not what we are going to get after a single trial!!)

- X discrete rv with pmf p_x

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x p_x \tag{12}$$

- X continuous rv with pdf $f_X(x)$

$$\mathbb{E}(X) = \int_{x \in \mathcal{X}} x \, f_X(x) \, dx \tag{13}$$

Watch out: the EV may not exist

Properties of Expected Value

- $\mathbb{E}(c) = c$ for any constant c
  - $\mathbb{E}[\mathbb{E}(X)] = \mathbb{E}(X)$
- $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
- $\mathbb{E}[X - \mathbb{E}(X)] = 0$
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

Given a continuous random variable X (respectively discrete) whose expectation exists and is finite, and any function g we have that

$$\mathbb{E}[g(X)] = \int_{\mathcal{X}} g(x)\, f_X(x)\, dx \qquad \left(\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) p_x\right) \tag{14}$$

The **Expected Value** gives a rough idea about the centre of the distribution but it does not provide any information about the dispersion of the possible observable values

Example: two investment plans that gives exactly the same expected payout; we would like to chose the one with lower variability

We need some further definitions and concepts since:

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

The **variance** of a random variable X

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \tag{15}$$

tells us **how much** the rv oscillates around its mean.

- X discrete rv with pmf p_x

$$\mathbb{V}[X] = \sum_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2 p_x \tag{16}$$

- X continuous rv with pdf $f_X(x)$

$$\mathbb{V}[X] = \int_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2 f_X(x) dx \tag{17}$$

Properties of the variance

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

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X) \tag{18}$$

- a **computationally-friendlier** formula for the variance

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \tag{19}$$

(i) Show that $\mathbb{V}(X)$ can be calculated by equation (19).

(ii) Let X be the number showing if we roll a die. Calculate expected value and variance.

If we have two random variables X and Y the **covariance** gives us a measure of the association between them

$$\mathbb{C}ov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \tag{20}$$

- 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

Independence of Random Variables

Two random variables X and Y are independent if

$$\begin{aligned} F_{X,Y}(x,y) &= \mathbb{P}(X \leq x \cap Y \leq y) \\ &= \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) \\ &= F_X(x)F_Y(y) \qquad \forall\, x,y \in \mathbb{R} \end{aligned} \tag{21}$$

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

Ramark: If $X_1, \ldots, X_n$ are independent then

- $p_{x_1,x_2,\ldots,x_n} = p_{x_1} \cdot p_{x_2} \cdots p_{x_n}$
- $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n)$

Factorisation Criterion

$$F_{X,Y}(x.y) = F_X(x)F_Y(y) \qquad \forall\, x, y \in \mathbb{R} \tag{22}$$

If X and Y are independent then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

As a consequence

$$\mathbb{C}ov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 \tag{23}$$

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

Exercise

(i) Prove formula (23) (iii) Let X and Y be two random variables with marginal distribution functions

$$F_X(x) = \begin{cases} 0 & if\ x < 0 \\ 1 - e^{-x} & if\ x \geq 0 \end{cases} \tag{24}$$

$$F_Y(y) = \begin{cases} 0 & if\ y < 0 \\ 1 - e^{-y} & if\ y \geq 0 \end{cases} \tag{25}$$

Determine if the two random variable are independent given that

$$F_{X,Y}(x,y) = \begin{cases} 0 & if\ x, y < 0 \\ 1 - e^{-x} - e^{-y} + e^{-x-y} & if\ x, y \geq 0 \end{cases} \tag{26}$$

Often you do not have to derive the distribution of a random variable on your own.

You can choose form a **catalogue** of known random variables whose functions are known and deeply investigated. You then select the one that is the more adequate to the phenomenon under analysis.

Known Discrete Random Variables

- Bernoulli
- Binomial
- Poisson
- Geometric
- Hypergeometric
- Degenerate
- ...

Bernoulli

Assume that a random experiment has two possible outcomes (typically adressed as *success* or *failure*)

The random variable X representing the result of the experiment can take either 0 or 1 as values.

We have that

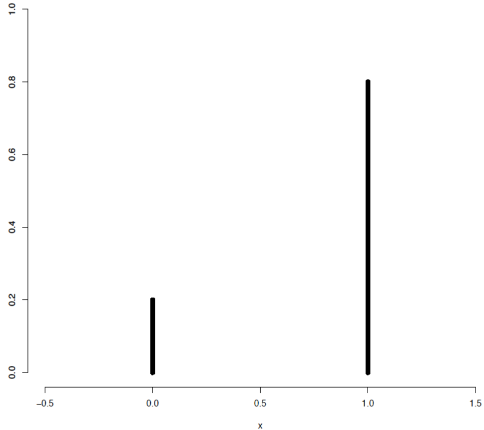
- Probability of success $\mathbb{P}(X = 1) = p$
- Probability of failure $\mathbb{P}(X = 0) = 1 - p$

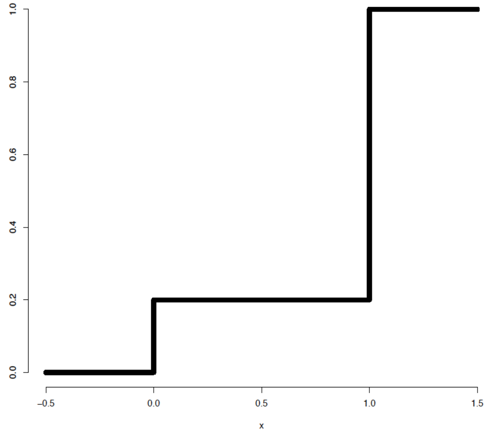
Example: result of an exam (pass or fail)

Bernoulli

$X \sim \textit{Bernoulli}(p)$

$$X = \begin{cases} 0 & \text{with probability } 1-p \\ 1 & \text{with probability } p \end{cases}$$





- $X \sim Bernoulli(p)$
- Compute expected value and Variance

Let X be the random variable representing the price behaviour of a Microsoft's stock.

- $X = 1$ if the price goes up
- $X = 0$ if the price goes down

(assuming it cannot stay fixed)

The price can go up with probability $3/5$.

Then X follows a Bernoulli distribution with parameter $p = 3/5$.

$$X \sim Bernoulli(3/5) \qquad X = \begin{cases} 0 & \text{with probability } \frac{2}{5} \\ 1 & \text{with probability } \frac{3}{5} \end{cases}$$

Typically, we are interested in the outcome of a Bernoulli experiment **on many** random repetitions, rather than just one.

Example: flip a coin T times, ask N people about their political preferences

The random variable of interest then becomes $X =$ "number of successes":

$$X = \sum_{i=1}^{n} Y_i \tag{27}$$

where $Y_1, \ldots, Y_n$ are independent Bernoulli random variables with parameter p

- Each of the n trials has only two possible outcomes. The outcome we are interested in is called *success* and the other *failure*
- Each trial has the same probability of success. The probability of a success is p then the probability of a failure is $1 - p$.
- The n trials are independent. The result of one does not affect the results of other trials.

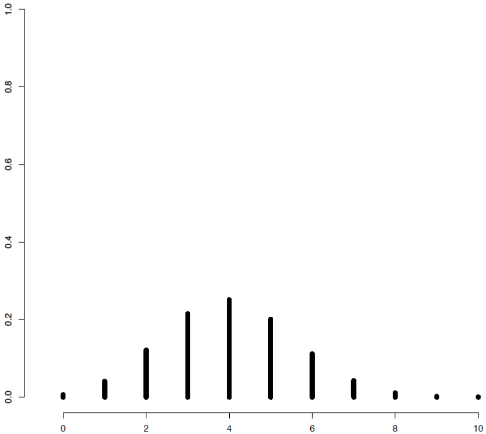
Then X follows a Binomial distribution with parameters n and p.

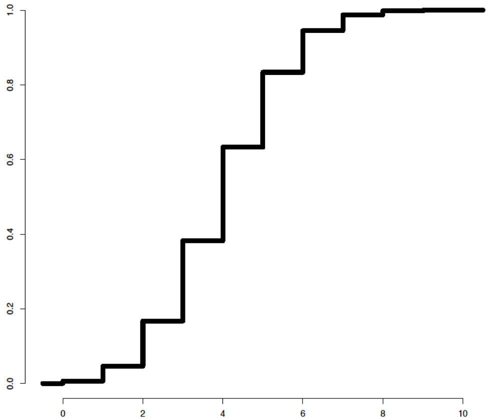
Binomial

$X \sim \textit{Binomial}(n, p)$

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$F_X(x) = \sum_{k \le x} p_X(k)$$





Binomial

Expected Value

$$
\begin{aligned}
\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x p_x &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= np \sum_{z=0}^{s} \frac{s!}{z!(s-z)!} p^z (1-p)^{s-z} = np
\end{aligned}
\tag{28}
$$

Binomial

Towards the variance

$$
\begin{aligned}
\mathbb{E}[X(X-1)] = \sum_{x\in\mathcal{X}} x(x-1)p_x &= \sum_{x=0}^{n} x(x-1)\begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x} \\
&= \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x!)} p^x(1-p)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x(1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2}(1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!((n-2)-(x-2))!} p^{x-2}(1-p)^{(n-2)-(x-2)} \\
&= n(n-1)p^2 \sum_{z=0}^{s} \frac{s!}{z!(s-z)!} p^{z}(1-p)^{s-z} = n(n-1)p^2
\end{aligned}
\tag{29}
$$

Binomial

Towards the variance

$$\mathbb{E}[X(X-1)] = \mathbb{E}(X^2 - X) = \mathbb{E}(X^2) - \mathbb{E}(X) \tag{30}$$

$$\mathbb{E}(X^2) = \mathbb{E}[X(X-1)] + \mathbb{E}(X) = n(n-1)p^2 + np \tag{31}$$

$$\begin{aligned} \mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 &= n(n-1)p^2 + np - (np)^2 \\ &= np\Big[np - p + 1 - np\Big] = np(1-p) \end{aligned} \tag{32}$$

Consider $X \sim Binomial(n, p)$ as the sum of n independent Bernoulli random variables Y_i.

Remember that if $Y_1, \dots, Y_n \sim Bernoulli(p)$ then $\mathbb{E}[Y_i] = p$ and $\mathbb{V}[Y_i] = p(1-p) \quad \forall\, i$, which is enough to prove

$$\mathbb{E}(X) = \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{E}(Y_i) = np \tag{33}$$

Moreover, since $Y_1, \dots, Y_n$ are independent we have that

$$\mathbb{V}(X) = \mathbb{V}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{V}(Y_i) = np(1-p) \tag{34}$$

Exercise

Garden records report that 65% of some rare plants grown there will not blossom. What is the probability that out of 10 randomly selected plants, 6 will have flowers?

Katniss Everdeen (The Hunger Games) is know to hit the target 4 times out of 5. If she shots 6 arrows, what is the probability of:

A. exactly 4 hits
B. at least 1 hit

Geometric Distribution

The **Geometric Distribution** gives the distribution of the number X of Bernoulli trials needed to get one success

If the probability of success in each trial is p, then the probability of observing a success on the xth trial (after $x-1$ failures) is

$$\mathbb{P}(X = x) = (1-p)^{x-1}p \tag{35}$$

The Geometric distribution is a suitable model for a random variable X if

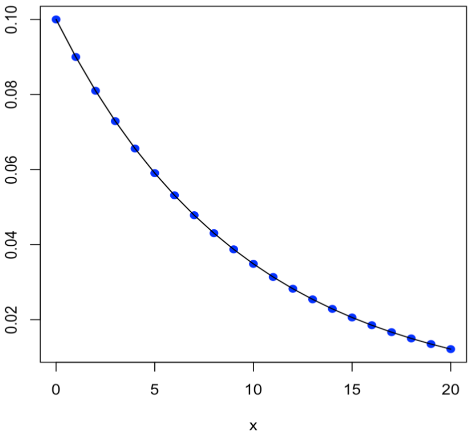
- X is the result of an experiment which requires a sequence of independent trials
- There are only two possible outcomes for each trials (success or failure)
- Each trial has the same probability of success p

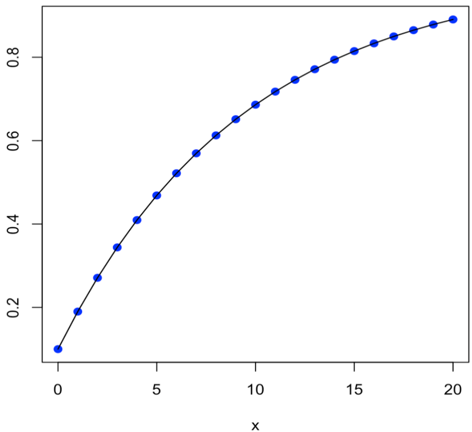
Geometric

$X \sim \textit{Geometric}(p)$

$$p_X(x) = \mathbb{P}(X = x) = (1-p)^{x-1}p$$

$$F_X(x) = 1 - (1-p)^x$$





Expected Value

$$\mathbb{E}(X) = \sum_{x\in\mathcal{X}} x p_X(x) = \sum_{x=0}^{\infty} x(1-p)^{x-1}p$$
$$= \sum_{x=0}^{\infty} x(1-p)^{x-1}p = p\sum_{x=0}^{\infty} x(1-p)^{x-1} \tag{36}$$
$$= p\sum_{x=0}^{\infty} -\frac{d}{dp}(1-p)^{x} = -p\frac{d}{dp}\sum_{x=0}^{\infty}(1-p)^{x}$$

Using the rule for geometric series we get

$$\mathbb{E}(X) = -p\frac{d}{dp}\frac{1}{p} = -p - \frac{1}{p^2} = \frac{1}{p} \tag{37}$$

Variance

$$\begin{aligned}\mathbb{E}(X^2) = \sum_{x=0}^{\infty} x^2 p_X(x) &= \sum_{x=0}^{\infty} x^2 (1-p)^{x-1} p \\ &= p \sum_{x=0}^{\infty} x^2 (1-p)^{x-1}\end{aligned} \tag{38}$$

Substitute $q = (1 - p)$ and solve the infinite series

$$\begin{aligned}\mathbb{E}(X^2) = p \sum_{x=0}^{\infty} x^2 q^{x-1} &= p \, \frac{1+q}{(1-q)^3} \\ = p \, \frac{2-p}{p^3} &= \frac{2-p}{p^2}\end{aligned} \tag{39}$$

Variance

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \tag{40}$$

$$\mathbb{V}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} \tag{41}$$

The lifetime risk of developing psoriasis is about 1 out of 78 (1.28%). Let X be the number of people you ask before one says she suffers of psoriasis.

- What is the probability that you ask 9 people before one she has psoriasis?
- Find the mean and standard deviation

A baseball player has a batting average of 0.320.

- What is the probability that he gets the first hit on the third trip to bat?
- How many trips to the bat do you expect before the hitter gets her first hit?

Poisson

The Poisson distribution is known as the distribution of **rare events**

It is typically used to model **counts**, i.e. the number of events in a given interval of time (or space)

Examples:

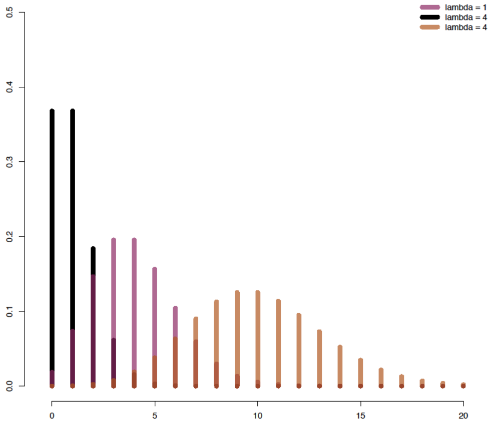
- number of clients calling a call centre
- number of defects of a square meter of manufactured goods
- number of patience arrived at the A&E in the last hour
- number of earthquakes in a year

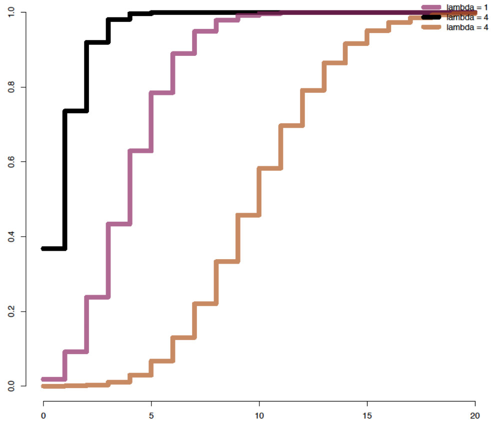
Poisson

$X \sim \textit{Poisson}(\lambda)$

$$p_X(x) = \mathbb{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$F_X(x) = \sum_{k \leq x} p_X(k)$$





$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in \mathcal{X}} x p_x = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!} = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= e^{-\lambda} \lambda \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} = \lambda
\end{aligned} \tag{42}
$$

Recall:

$$
e^{\alpha} = \sum_{s=0}^{\infty} \frac{\alpha^s}{s!} \tag{43}
$$

Poisson

Towards the variance

$$
\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{x \in \mathcal{X}} x(x-1)p_x = \sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=1}^{\infty} x(x-1)\frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} \\
&= e^{-\lambda}\lambda^2 \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} = \lambda^2
\end{aligned} \tag{44}
$$

$$
\mathbb{E}[X(X-1)] = \mathbb{E}[X^2 - X] = \mathbb{E}[X^2] - \mathbb{E}[X] \tag{45}
$$

$$
\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda \tag{46}
$$

$$
\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \tag{47}
$$

Exercise

Suppose it has been observed that, on average, 180 cars per hour pass a specified point on a particular road in the morning rush hour. Due to impending roadworks it is estimated that congestion will occur closer to the city centre if more than 3 cars pass the point in any one minute. What is the probability of congestion occurring?

Known Continuous Random Variables

- Uniform
- Exponential
- Normal
- Student's t
- Gamma
- Beta
- ...

- A random variable X is **uniformly distributed** between a and b, if X take value in any interval of a given size with equal probability
- The probability of X being in an interval is proportional to the length of the interval

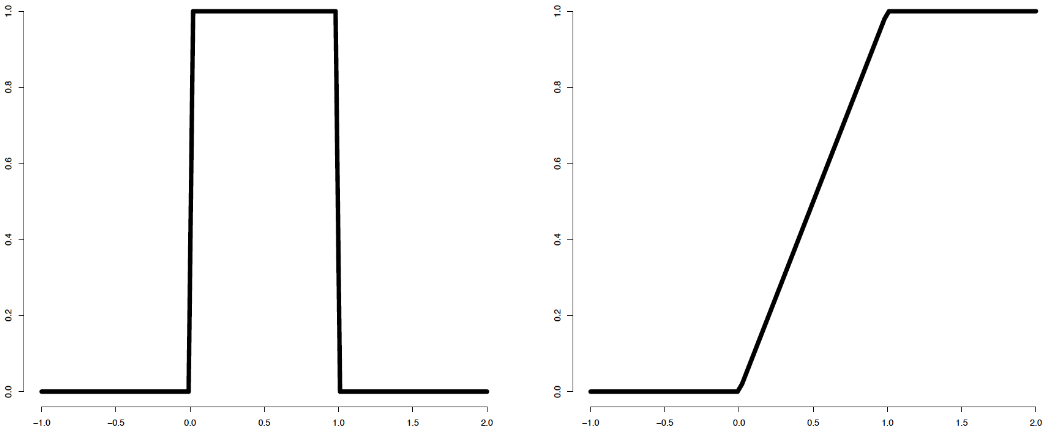
Example: the arrival of the bus between the moment you get to the stop and midnight

Uniform

$X \sim \mathit{Unif}(a, b)$

$f_X(x) = \frac{1}{b-a}$

$F_X(x) = \frac{x-a}{b-a}$



Uniform

Expected Value

$$\begin{aligned}\mathbb{E}[X] = \int_a^b x f_X(x)\,dx &= \int_a^b x \frac{1}{b-a}\,dx \\ &= \frac{1}{b-a}\int_a^b x\,dx \\ &= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\ &= \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{a+b}{2}\end{aligned} \tag{48}$$

Exercise: Prove that $\mathbb{V}[X] = \frac{(b-a)^2}{12}$

The current (in mA) measured in a piece of copper wire is known to follow a uniform distribution over the interval $[0, 48]$. Write down the formula for the probability density function of the random variable X representing the current. Calculate the mean and variance of the distribution and find the cumulative distribution function.

Exercise

The amount of time, in minutes, that a person will wait at the post office is uniformly distributed between $[0, 25]$.

- Find mean and standard deviation
- What is the probability of waiting less than 16.5 minutes?
- Find the $90th$ percentile

The battery duration x of an iPhone is known to be uniformly distributed between $[20, 40]$ years.

- Write the probability density function
- Find mean and variance
- Find the cumulative distribution function
- What is the probability that the battery of an iPhone will last less than 35 years?

Exponential Distribution

A random variable X follows an **Exponential Distribution** with parameter $\lambda > 0$ if its probability density function can be written as

$$f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0 \tag{49}$$

The intuition behind an Exponential random variable is that the **larger** is a value, the **less likely** it is.

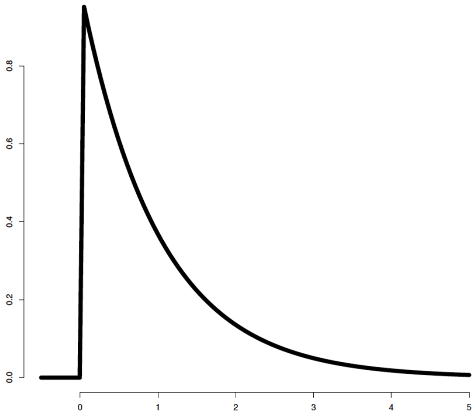
The Exponential distribution is typically used to model **time until some specific event** and the parameter λ affects the mean time between events.

Example: the amount of time until an earthquake strikes, the amount of money customers are going to spend in one trip to supermarket ...

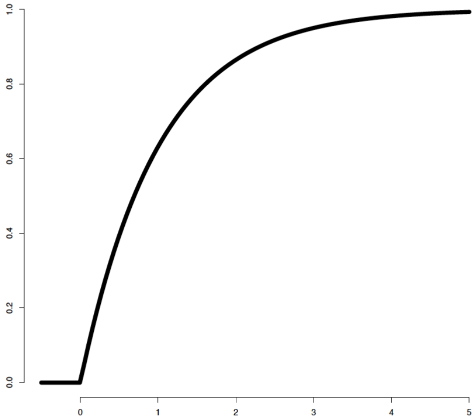
Exponential

$X \sim Exp(\lambda) \quad \lambda > 0$ and $x \geq 0$

$$f_X(x) = \lambda e^{-\lambda x}$$



$$F_X(x) = 1 - e^{-\lambda x}$$



Expected Value

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x f_X(x)\,dx = \int_0^\infty x\lambda e^{-\lambda x}\,dx \\
&= \lambda \int_0^\infty x e^{-\lambda x}\,dx
\end{aligned} \tag{50}
$$

Integrating by parts $f(x) = x$, $g'(x) = e^{-\lambda x} \implies f'(x) = 1 \implies du = dx$ and $g(x) = -\frac{e^{-\lambda x}}{\lambda}$

$$
\mathbb{E}[X] = \lambda \int_0^\infty x e^{-\lambda x}\,dx = \lambda \Big[-\frac{x e^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2} \Big]_0^\infty = \frac{1}{\lambda} \tag{51}
$$

Exercise

Find the variance

If jobs arrive every 15 seconds on average, $\lambda = 4$ per minute, what is the probability of waiting less than or equal to 30 seconds, i.e 0.5 min?

The amount of time Tor Vergata's researchers in Statistics spend studying Statistics can be modelled by an exponential distribution with the average time equal to 15 minutes (far way more per day!!). Write the distribution, state the probability density function. Find the probability that a randomly selected researcher spends one to two hours studying statistics.

A random variable X follows a **Gamma Distribution** if its probability density function can be written as

$$f_X(x) = \frac{1}{\Gamma(\alpha)}\beta^{\alpha}x^{\alpha-1}e^{-\beta x} \quad x \geq 0 \tag{52}$$

where $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}\,dt$ is the Gamma function

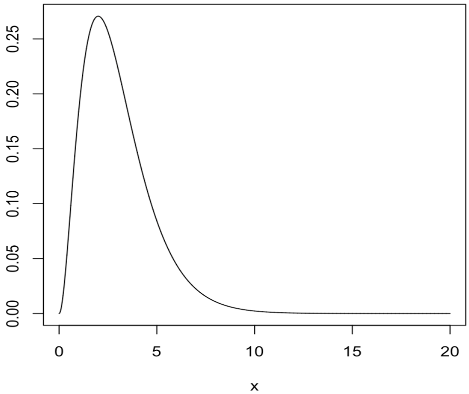
- Alternative parametrisations:
  - $f_X(x \mid k, \theta)$; $f_X(x \mid \theta, \mu)$
- Widely used in Econometrics to model waiting times
- Bayesian Statistics: conjugacy and relationship with the Inverse-Gamma distribution

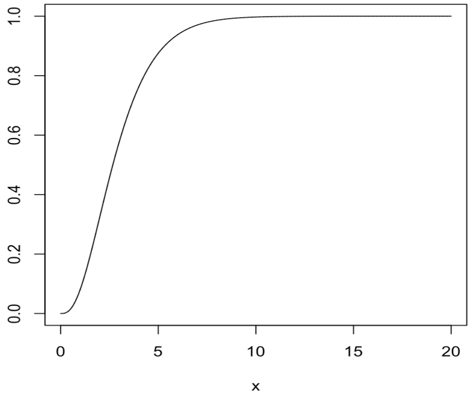
$X \sim \textit{Gamma}(\alpha, \beta) \quad \alpha, \beta > 0$ and $x \geq 0$

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{\alpha - 1} e^{-\beta x}$$

$$F_X(x) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \beta x)$$

where $\gamma(\alpha, \beta x)$ is the lower incomplete gamma function





Expected Value

$$
\begin{aligned}
\mathbb{E}[X] = \int_0^\infty x f_X(x)\, dx = \int_0^\infty x \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{\alpha-1} e^{-\beta x}\, dx \\
= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\beta x}\, dx
\end{aligned}
\tag{53}
$$

Substitute $t = \beta x \implies dx = dt/\beta$

$$
\begin{aligned}
\mathbb{E}[X] = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{t}{\beta}\right)^\alpha e^{-t}\, \frac{dt}{\beta} &= \frac{\beta^\alpha}{\beta^{\alpha+1}\Gamma(\alpha)} \int_0^\infty t^\alpha e^{-t}\, dt \\
&= \frac{1}{\beta\Gamma(\alpha)} \Gamma(\alpha+1) \\
&= \frac{\alpha}{\beta}
\end{aligned}
\tag{54}
$$

Variance

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_0^\infty x^2 \, \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{\alpha-1} e^{-\beta x} \, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-\beta x} \, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{t}{\beta}\right)^{\alpha+1} e^{-t} \, \frac{dt}{\beta} \\
&= \frac{1}{\beta^2 \Gamma(\alpha)} \int_0^\infty t^{\alpha+1} e^{-t} \, dt \\
&= \frac{\alpha(\alpha+1)}{\beta^2}
\end{aligned} \tag{55}
$$

$$
\begin{aligned}
\mathbb{V}[X] = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2 &= \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} \\
&= \frac{\alpha}{\beta^2}
\end{aligned} \tag{56}
$$

The **Normal** or **Gaussian** is the queen of all random variables.

- It is helpful in representing many natural and economic phenomena
- It can be used to approximate other distributions
- It is key to inference in sampling

A traditional parametrisation

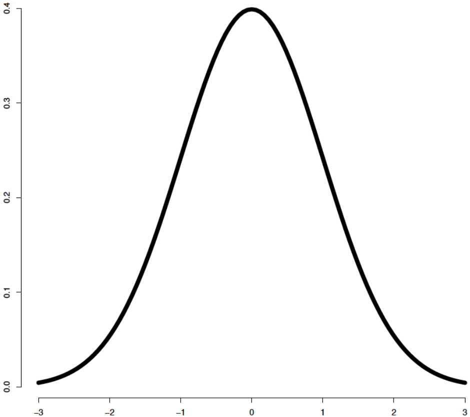
$\mathbb{E}[X] = \mu \qquad \mathbb{V}[X] = \sigma^2$

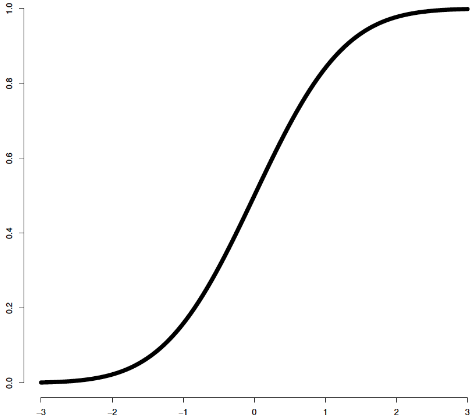
Normal

$X \sim \mathcal{N}(\mu, \sigma^2)$ $\sigma > 0$ and $\mu \in \mathbb{R}$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$F_X(x) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-\mu)^2}\, dt$$



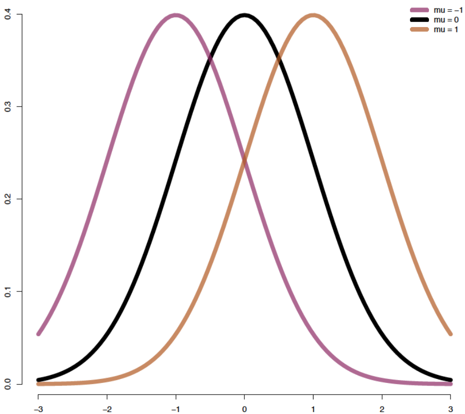


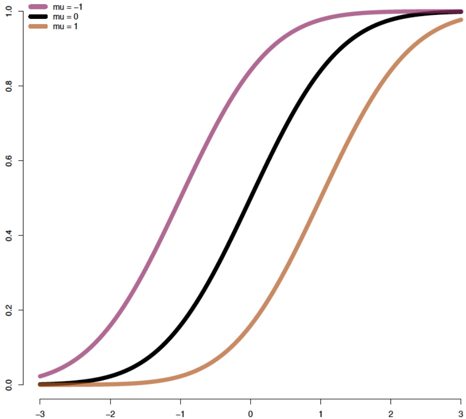
Normal

Varying mu

$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \sigma > 0$ and $\mu \in \mathbb{R}$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad F_X(x) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-\mu)^2}\, dt$$





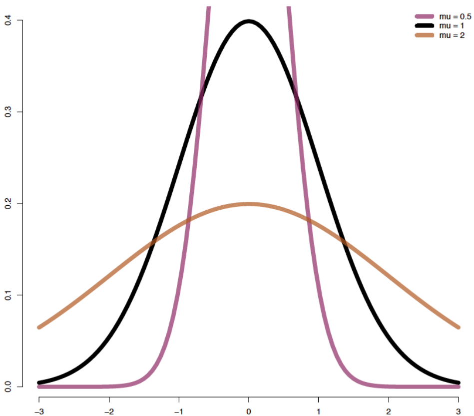
Normal

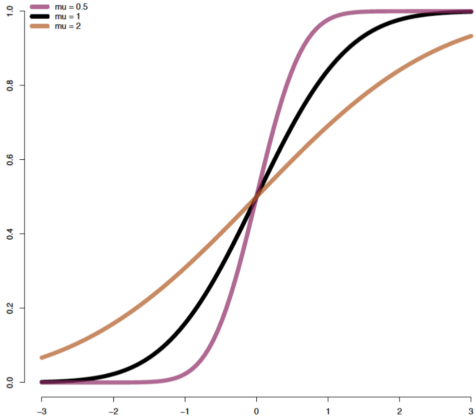
Varying sigma

$X \sim \mathcal{N}(\mu, \sigma^2)$ $\sigma > 0$ and $\mu \in \mathbb{R}$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$F_X(x) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-\mu)^2}\, dt$$





Properties

A linear transformation of a Normal random variable is still a Normal random variable

If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$ then

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2) \tag{57}$$

A linear combination of Normal random variables is still a Normal random variable

If $X_1, \ldots, X_n$ are independent random variables such that $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ then

$$Y = \sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right) \tag{58}$$

Standard Normal

When $\mu = 0$ and $\sigma^2 = 1$ the random variable $X \sim \mathcal{N}(0,1)$ is called a **standard normal random variable** and usually denoted by **Z**

Every Normal random variable can be turned into a standard Normal via **standardisation**

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

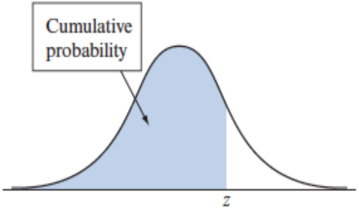
$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1) \tag{59}$$

This is just a linear transformation of X, thus it is easy to show that

$$\mathbb{E}[Z] = \mathbb{E}\left[\frac{X - \mu}{\sigma}\right] = \frac{\mathbb{E}[X] - \mu}{\sigma} = 0 \tag{60}$$

$$\mathbb{V}[Z] = \mathbb{V}\left[\frac{X - \mu}{\sigma}\right] = \frac{\mathbb{V}[X]}{\sigma^2} = 1 \tag{61}$$

Tables of a Standard Normal



Cumulative probability for *z* is the area under the standard normal curve to the left of *z*

Table A Standard Normal Cumulative Probabilities (*continued*)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549

Exercises

- The time (in minutes), X, that is needed to solve this Statistics exercise is normally distributed with mean 5 and standard deviation 10. When I solved it at home, it took me 6.2 minutes. What is the probability of a random PhD student faster than me?
- X is a normally distributed random variable with mean 30 and standard deviation 4.
  - Find $\mathbb{P}(X < 40)$
  - Find $\mathbb{P}(X > 21)$
  - Find $\mathbb{P}(30 < x < 35)$
- Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585. Will he be admitted to this university?