Probability

Tullia Padellini

tulliapadellini.github.io
tullia.padellini@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

short probability glossary

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

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

Exercise

something to get you started

Two coins are tossed. Each coin has two possible outcomes H (Heads) and T (Tails).

- Determine the sample space and its size (i.e. how many elements are in the set)
- Formalize in terms of possible outcomes the event $E =$"the faces appearing on the two coins are different" and determine its size.

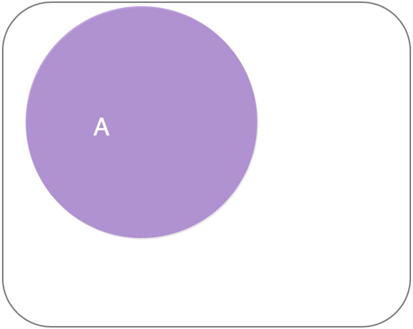
A card is drawn at random from a deck of 52 cards.

- Determine the sample space
- Formalize in terms of possible outcomes the event $E =$"the card drawn is a spade"

Recap of Set theory

basic operations on sets

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



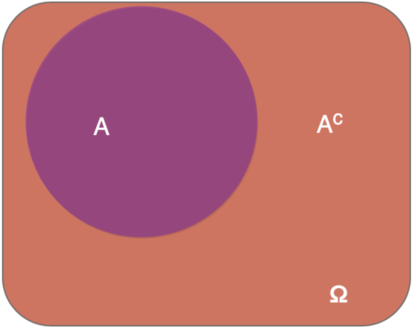
Example $A =$ "the die returns an even number"

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

Recap of Set theory

basic operations on sets

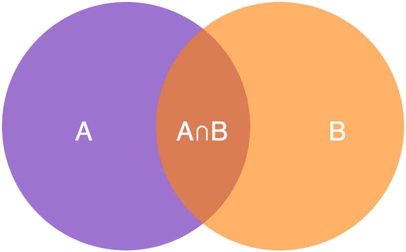
- **Complement** (A^c or $\bar{A}$) everything that is not in A.



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.



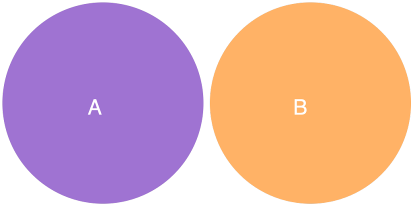
Example $A =$ "the die returns an even number", $B =$ "the die returns a number smaller than 5"

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

Recap of Set theory

basic operations on sets

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



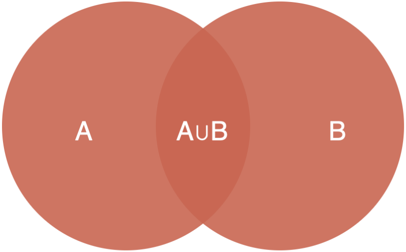
Example $A =$ “the die returns an even number”, $B =$ “the die returns a 5”

$\Rightarrow A \cap B = \emptyset$ $\quad$ A and B are **disjoint**

Recap of Set theory

basic operations on sets

- **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”

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

Some useful relationships

computing basic probabilities

Probability Axiomes...

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

Some useful relationships

computing basic probabilities

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

How do we define probability?

something we don't really focus about

- Classical approach: assigning probabilities based on the assumption of equally likely outcomes.
- Frequency approach: assigning probabilities as the limit of the relative frequency of the event happening in infinite repetition of the random experiment.
- Subjective approach: assigning probabilities based on the assignor's judgment / historical data.

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.

Exercises

based on the Classical approach to probability evaluation

Which of the following events has probability 0?

- Choosing an odd number from 1 to 10.
- Getting an even number after rolling a single 6-sided die.
- Choosing a white marble from a jar of 25 green marbles.
- 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?

What is the probability that an Italian newborn is a girl?

Exercises

based on the Classical approach to probability evaluation

Which of the following events has probability 0?

- Choosing an odd number from 1 to 10.
- Getting an even number after rolling a single 6-sided die.
- Choosing a white marble from a jar of 25 green marbles.
- 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?

What is the probability that an Italian newborn is a girl?

1/2? f you know that women are the 51.3% of the Italian population, does your guess changes?

Exercises

based on the Classical approach to probability evaluation

Which of the following events has probability 0?

- Choosing an odd number from 1 to 10.
- Getting an even number after rolling a single 6-sided die.
- Choosing a white marble from a jar of 25 green marbles.
- 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?

What is the probability that an Italian newborn is a girl?

1/2? If you know that women are the 51.3% of the Italian population, does your guess changes?

Conditional Probability

encoding new information

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.

Exercise

your turn

The blue M&M was introduced in 1995. Before then, the color mix in a bag of plain M&Ms was 30% Brown, 20% Yellow, 20% Red, 10% Green, 10% Orange and the rest was Tan.

Starting from 1995, the color mix is 24% Blue, 16% Orange, 14% Yellow, 13% Red, 13% Brown and the rest is Green.

Tullia is still jealously preserving M&Ms bags from before 1995. She is about to give her friend Carlo two M&M candies, choosing one at random from a fresh package and the other one at random from a bag produced in 1994.

- Determine the probability of selecting a tan M&M from the 1994 bag.
- Determine the probability of selecting a green M&M from the fresh bag.
- Determine the probability that none of the M&Ms selected is blue.
- Determine the probability that Tullia gives Carlo two green M&Ms.

Exercise pt II

BONUS

- Tullia gives Carlo a yellow and a green M&M. In order not to get sick, he should not eat the one taken from the 1994 bag. What is the probability that she is trying to Poison her friend with a yellow M&M taken from the 1994 bag?

you may want to check Bayes Theorem for this

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 coint three times. X is the random variable representing the *number of Tails*

ω	$P(\omega)$	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	1/8×1 = 1/8
1	1/8×3 = 3/8
2	1/8×3 = 3/8
3	1/8×1 = 1/8

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

Exercise

your turn

Two dice are rolled:

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

Distribution of a Discrete Random Variable:

discrete = how many

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

- **Probability mass distribution**

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

- **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!*)

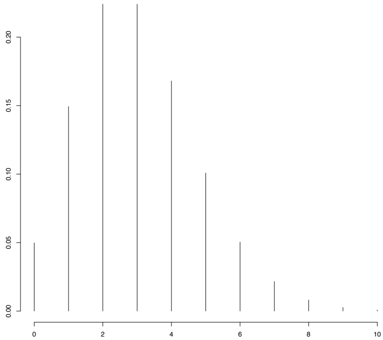
$$P(X = 2 \cup X = 0) = P(X = 2) + P(X = 0) = p_0 + p_2 = 4/8$$

Properties

of discrete distribution functions

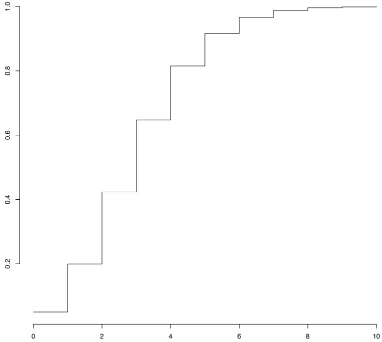
Probability mass distribution

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



Cumulative distribution function

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



Exercise

your turn

Let X be a discrete random variable with the following cumulative distribution function

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

- compute the corresponding probability mass function
- compute the following probabilities:
  - $P(X=6)$
  - $P(X=5)$
  - $P(2<X<5.5)$
  - $P(0 \leq X<5.5)$

Distribution of a Continuous Random Variable:

continuous = how much

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

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

- **Cumulative distribution function**

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

- **Probability density distribution**

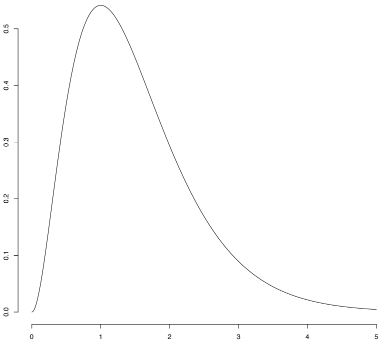
$$f_X(x) = \frac{dF_X(x)}{dx} \qquad\qquad \forall x \in \mathcal{X}$$

Properties

of continuous distribution functions

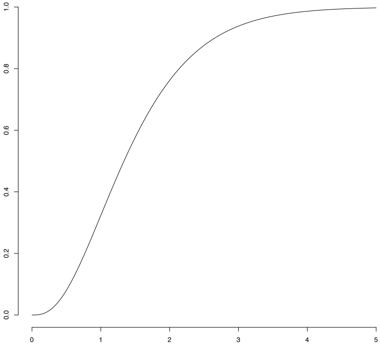
Probability density function

- $f_X(x) \geq 0$
- $f_X(x)$ **needs not** be ≤ 1
- $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$



Cumulative distribution function

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



Exercise

your turn

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

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

- determine c so that $f(x)$ is a valid pdf
- compute $P(X = 0.5)$
- compute $P(X > 0.5)$

Comparison

discrete vs continuous

- 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 $f_X(x)$
- $P(X \in A) = \int_A f_X(x) \mathsf{d}x$

if $A = [a, b]$ then

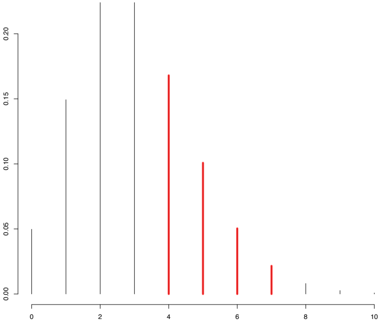
$$\begin{aligned} P(X \in A) &= \int_a^b f_X(x) \mathsf{d}x \\ &= F_X(b) - F_X(a) \end{aligned}$$

Comparison

discrete vs continuous

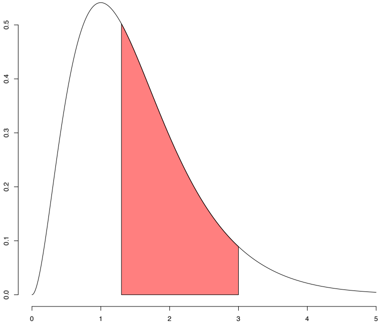
$A = \{x_1, \ldots, x_k\}$

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



$A = [a, b]$

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



Summaries

measuring the center of the distribution

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

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

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

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) \mathsf{d}x$$

WATCH OUT The expected value *may not exist!*

Properties

of the expected value

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

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

- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- $\mathbb{E}[X - \mathbb{E}[X]] = 0$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 0$

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

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

Measuring Variability

another way of summarizing a distribution

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. $f_X(x)$

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

Properties

of the variance

- 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 sd$(X) = \sqrt{\mathbb{V}[X]}$ is called **standard deviation**. Roughly, ? 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-frendlier** alternative definition of the variance is:

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

Exercise

your turn

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

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

- compute its expected value and variance

Exercise

your turn

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

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

- compute its expected value and variance

Covariance

measuring dependence

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

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

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

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

Multidimensional random variables

distribution functions

Let (X,Y) be a two dimensional random variable:

- Joint Cumulative Distribution Function:

$$F_{X,Y}(x,y) = P(X \leq x \cap Y \leq y)$$

- Joint Probability Mass Function (only for X,Y **discrete**)

$$p_{X,Y}(x,y) = P(X = x \cap Y = y)$$

- Joint Probability Density Function (only for X,Y **continuous**)

$$f_{X,Y}(x,y) = dF_{X,Y}(x,y)/dxdy$$

Univariate distributions

derived from the joint distribution

Let (X,Y) be a two dimensional random variable:

- Marginal Probability Mass Function (only for X,Y **discrete**)

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

- Marginal Probability Density Function (only for X,Y **continuous**)

$$f_X(x) = \int f_{X,Y}(x,y)dy$$

- Conditional Mass / Density Function:

$$p_{X|Y=y}(x) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \qquad f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Toy Example

Car crashes

- X: extent of the injury
- Y: type of restraint

Joint Probability Mass Function

$p_{X,Y}$	None	Belt	Belt and Harness	$p_X(x)$
None	0.065	0.075	0.06	0.20
Minor	0.175	0.16	0.115	0.45
Major	0.135	0.10	0.065	0.30
Death	0.025	0.025	0.01	0.05
$p_Y(y)$	0.40	0.35	0.25	1

Conditional Probability Mass Function

$X\|Y$=None	None	Minor	Major	Death
$p_{X\|Y}$	0.162	0.438	0.337	0.063

Exercise
all together now

Let (X, Y) be a bivariate random variable with joint probability density function

$$f_{X,Y}(x, y) = e^{-y} \qquad\qquad 0 < x < y < \infty$$

- compute the marginal distribution of X
- compute the conditional distribution of $Y|X = x$
- compute the expected value of $Y|X = x$

Independence of Random Variables

assessing the relation between two variables

Two random variables X, Y are independent if

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

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

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

Independence of Random Variables

Factorization Criterion

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

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

your turn

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

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

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

Determine if X and Y are indipendent given the joint distribution function:

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

Famous random variables

you have help

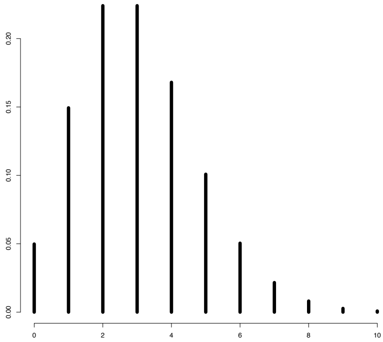
Typically you do not have to derive yourself the distribution function for a random variable.

You can choose from a **catalog** of famous random variable, whose distribution are know and well investigated, which one is more adequate to the phenomenon of interest.

Discrete Random Variables
a quick reminder

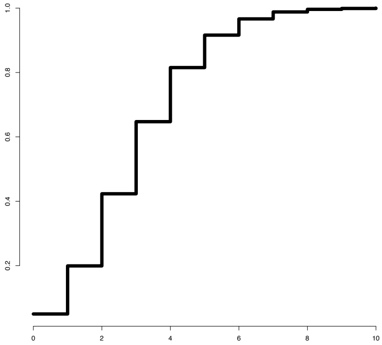
Probability mass distribution

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



Cumulative distribution function

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



Uniform

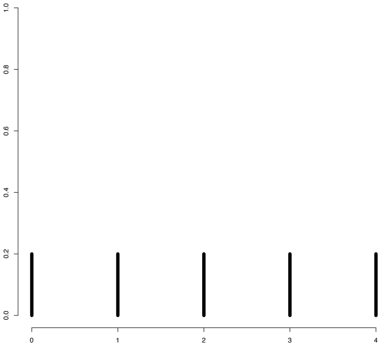
the intuition behind it

- The discrete uniform distribution is a symmetric probability distribution whereby a finite number of values are equally likely to be observed.
- If the support of X (i.e. the set of values that X can take) has size K, the probability of each of its elements is 1/K.
- The discrete uniform distribution itself is inherently non-parametric. It is convenient, however, to represent its values as all integers in an interval [a,b], so that a and b become the main parameters of the distribution.

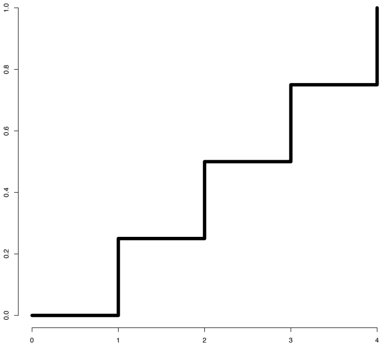
Uniform

$X \sim$ Unif$\{x_1, \ldots, x_K\}$. For today we assume $a = x_1 \leq \cdots \leq x_K = b$.

$$p_j = P(X = x_j) = 1/K$$



$$F_X(x) = P(X \leq x) = \frac{\lfloor x \rfloor - a + 1}{b - a + 1}$$



Formulas
expected value

$X \sim \mathsf{Uniform}(a, b)$

$$p_x = P(X \leq x) = \frac{\lfloor x \rfloor - a + 1}{b - a + 1}$$

$X \in [a, b] = [a, a + k, a + 2k, \dots, b] \quad \text{where} \quad b = a + (n-1)k$

$$\begin{aligned}
\mathbb{E}[X] = \sum_x x p_x &= \sum_{l=0}^{n-1} x \frac{1}{n} = \frac{1}{n} \sum_{l=0}^{n-1} a + lk = \\
&= \frac{1}{n} [na + k \sum_{l=0}^{n-1} l] = a + \frac{k(n-1)n}{2n} = \\
&= a + \frac{k(n-1)}{2} = a + \frac{b-a}{2} = \\
&= \frac{a+b}{2}
\end{aligned}$$

Formulas

expected value of X squared

$X \sim \mathsf{Uniform}(a, b)$

$$p_x = P(X \leq x) = \frac{\lfloor x \rfloor - a + 1}{b - a + 1}$$

$$\begin{aligned}
\mathbb{E}[X^2] = \sum_x x^2 p_x &= \frac{1}{b-a+1} \sum_{x=a}^{b} x^2 = \\
&= \frac{1}{b-a+1} (\frac{(b^2+b)(2b+1) - (a^2-a)(2a-1)}{6}) = \\
&= \frac{1}{b-a+1} (\frac{(2b^3+3b^2+b) - (2a^3-3a^2+a)}{6})
\end{aligned}$$

Formulas

variance

$X \sim \text{Uniform}(a, b)$

$$p_x = P(X = x) = \frac{\lfloor k \rfloor - a + 1}{b - a + 1}$$

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

$$\begin{aligned}
\mathbb{V}[X] &= \frac{1}{b-a+1}(\frac{(2b^3 + 3b^2 + b) - (2a^3 - 3a^2 + a)}{6}) - (\frac{a+b}{2})^2 = \\
&= \cdots = \\
&= \frac{(b-a+1)^2 - 1}{12}
\end{aligned}$$

Example

a uniform random variable you already know

Throw a fair die. The random variable X describing the value of the face up follows a **uniform discrete distribution**.

- $X \sim \mathsf{Unif}\{1, 6\}$
- $\{a = x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5, x_6 = 6 = b\}$
- $p_x = P(X = x) = 1/6$ for every $x \in \{1, 2, 3, 4, 5, 6\}$

Exercise

something to keep you busy

A die is rolled.

- List the possible outcomes in the sample space.
- What is the probabilty of getting a number which is even?
- What is the probabilty of getting a number which is greater than 4?
- What is the probabilty of getting a number which is less than 3? What is its complement?
- If two dice are thrown and their values added, is the resulting distribution still a discrete uniform?

Bernoulli

the idea behind it

Assume your random experiment has two possible outcomes (typically addressed as *Success* and *Failure*).

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

Since Success/Failure are the only two possible outcomes we have that:

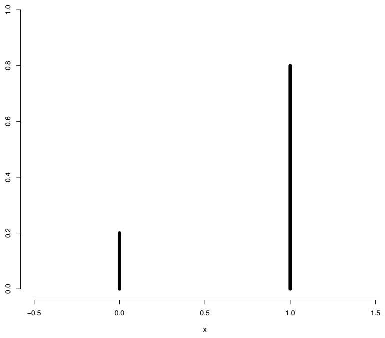
- Probaility of success: $P(X = 1) = p$
- Probability of failure: $P(X = 0) = 1 - p$.

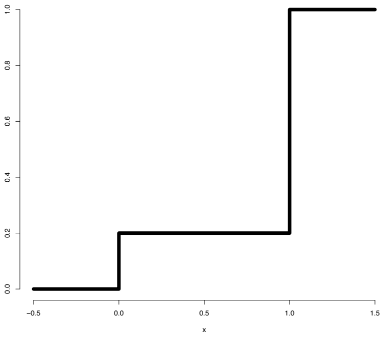
Example: Support to a political party, Result of an exam (pass/fail)

Bernoulli

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

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





Exercise

easiest expected value and variance you can derive

$X \sim$ Bernoulli(p)

- Compute $\mathbb{E}[X]$ and $\mathbb{V}[X]$

Expected Values

suuuper easy

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

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_x x p_x \\
&= 0 \times p_0 + 1 \times p_1 \\
&= 0 \times (1 - p) + 1 \times p = p
\end{aligned}
$$

Variance

trickier but still doable

Remember the *quick* formula to compute the variance

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

Compute the $\mathbb{E}[X^2]$, the 2nd moment of the distribution:

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_x x^2 p_x \\ &= 0 \times p_0 + 1 \times p_1 \\ &= 0 \times (1-p) + 1 \times p = p\end{aligned}$$

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= p - p^2 \\ &= p(1-p)\end{aligned}$$

Toy Example

What is a Bernoulli?

Let X be the random variable representing the price behavior of Microsoft Stock in the next month.

- $X = 1$ if the price next month of Microsoft stock goes up
- $X = 0$ if the price goes down (assuming it cannot stay the same).

The price will go up with probability $p = 3/5$. Then X follows a Bernoulli distribution with parameter $p = 3/5$.

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

Binomial

from one Bernoulli to many

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

Example: Flip a coin 3 times, ask 10 people about their political preferences

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

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

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

Binomial

conditions under which it may be used

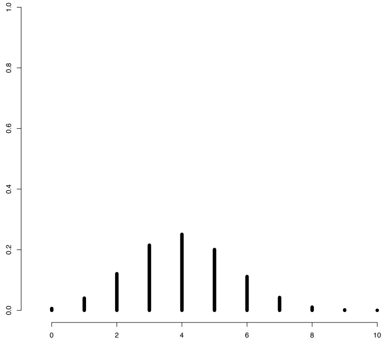
- Each of n trials has two possible outcomes. The outcome of interest is called a success and the other outcome is called a failure.
- Each trial has the same probability of a success. This is denoted by p, so the probability of a success is p and the probability of a failure is $1 - p$.
- The n trials are independent. That is, the result for one trial does not depend on the results of other trials.

Then the random variable X representing the number of successes in the n trials follows a Binomial distribution with parameters n and p.

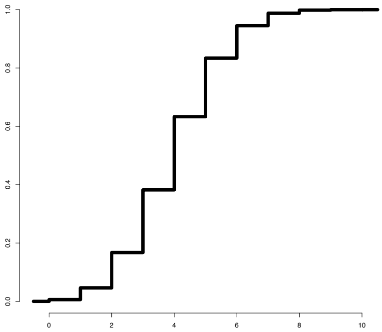
Binomial

$X \sim$ Binomial(n, p)

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



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

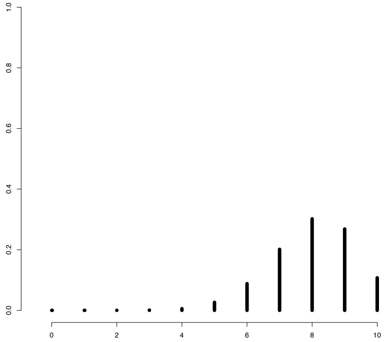


Binomial

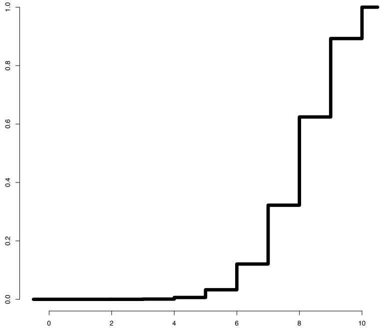
large p

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

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



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

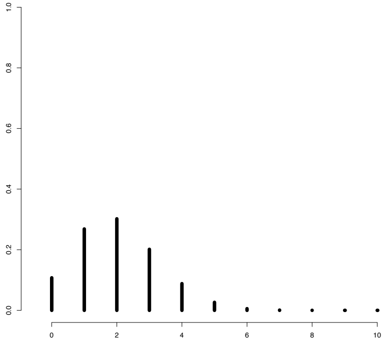


Binomial

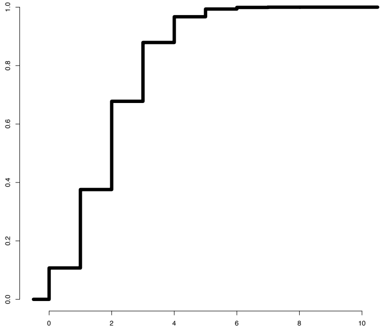
small p

$X \sim$ Binomial(n, p)

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



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



Expected Value

not as easy as before...

$$
\begin{aligned}
\mathbb{E}[X] = \sum_x x p_x &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \binom{n}{x} p^x (n-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{z=0}^{n-1} \frac{n(n-1)!}{z!(n-z-1)!} p^{z+1} (1-p)^{n-z-1} \\
&= np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!(n-z-1)!} p^z (1-p)^{n-1-z} = np
\end{aligned}
$$

Towards the Variance

still challenging

$$
\begin{aligned}
\mathbb{E}[X^2] = \sum_x x^2 p_x &= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{z=0}^{n-1} (z+1) \frac{n(n-1)!}{z!(n-z-1)!} p^{z+1} (1-p)^{n-z-1} \\
&= np[\sum_{z=0}^{n-1} z \binom{n-1}{z} p^z (1-p)^{n-1-z} + \sum_{z=0}^{n-1} \binom{n-1}{z} p^z (1-p)^{n-1-z}] \\
&= np(n-1)p + np
\end{aligned}
$$

Variance

the end result

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

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

Shortcut

an easy way out

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

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

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} p = np$$

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

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

Example

something you know already

Flip a fair coin 3 times. The random variable X representing the number of heads follows a Binomial distribution.

- The Bernoulli trial is the coin flip, hence $n = 3$
- The coin is *fair*, hence $p = 1/2$
- The support of X is $\{0, 1, 2, 3\}$
- $P(X > 2) = 1 - F_X(1) = 1 - (p_X(0) + p_X(1)) = 0.5$

since $p_X(0) = \binom{3}{0}(0.5)^3 = 0.125$ and $p_X(1) = \binom{3}{1}(0.5)^1(0.5)^2 = 0.375$

Exercises

your turn!

A quiz in statistics course has four multiple-choice questions, each with five possible answers. A student needs three or more correct answers to pass. Allison has not studied for the quiz, so she answer completely at random to all the questions.

- Find the probability she lucks out and answers all four questions correctly.
- Find the probability that she passes the quiz.

Each newborn baby has a probability of approximately 0.49 of being female and 0.51 of being male. For a family of four children, let X = number of children who are girls.

- Explain why the Binomial represent the phenomenon and identify its parameters.
- Compute the mean and the variance of X.
- Find the probability that the family has two girls and two boys.

Poisson

the underlying intuition

The **Poisson** distribution is called the *distribution of rare events*.

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

Examples:

- # of clients calling a call-center
- # of defects in a square meter of a manufactured good
- # of patients arriving to the emergency hospital in the last hour
- # of earthquakes in a given interval of time

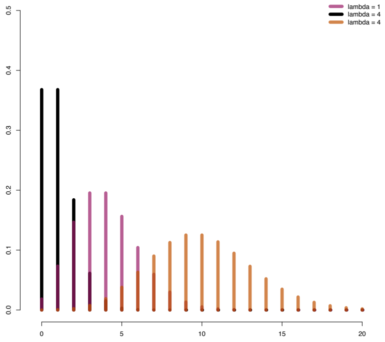
Poisson

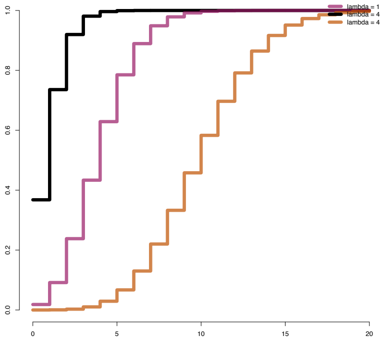
the formalization

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

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

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





Expected Value

pretty mathy

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

Variance

still non trivial

$$\mathbb{E}[X^2] = \sum_x x^2 p_x = \sum_{x=0}^{\infty} x^2 \frac{\lambda^x e^{-\lambda}}{x!} =$$

$$= \sum_{x=1}^{\infty} x \frac{\lambda^x e^{-\lambda}}{(x-1)!} =$$

$$= \sum_{z=0}^{\infty} (z+1) \frac{\lambda^{z+1} e^{-\lambda}}{z!} =$$

$$= \lambda(\sum_{z=0}^{\infty} z \frac{\lambda^(z) e^{-\lambda}}{z!} + \sum_{z=0}^{\infty} \frac{\lambda^(z) e^{-\lambda}}{z!}) =$$

$$= \lambda(\lambda + 1) = \lambda^2 + \lambda$$

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

$$\mathbb{V}[X] = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Toy Example

a basic one

Your instagram account receives on average 10 likes per day. In order to gain more information you are interested in modeling the distribution of

X = "number of likes per day".

- $X \sim$ Poisson(λ), with $\lambda = 10$
- Probability of having a bad day? $P(X = 3) = \frac{10^3 e^{-10}}{3!} = 0.0076$

Exercises

your turn

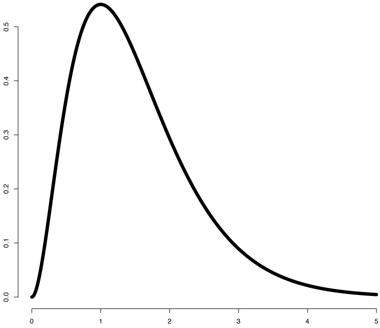
The average number of homes sold by the Acme Realty company is 2 homes per day. What is the probability that exactly 3 homes will be sold tomorrow?

Suppose the average number of lions seen on a 1-day safari is 5. What is the probability that tourists will see fewer than 4 lions on the next 1-day safari?

Continuous Distributions:
a small recap

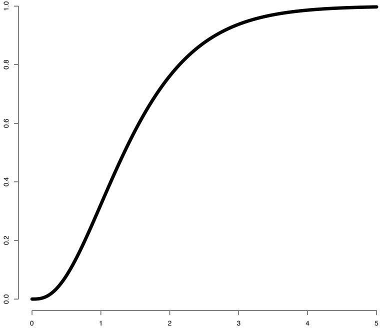
Probability density function

- $f_X(x) \geq 0$
- $f_X(x)$ **needs not** be ≤ 1
- $\int_{-\infty}^{\infty} f_X(x)\mathsf{d}x = 1$



Cumulative distribution function

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



Continuous Uniform Distribution

the intuition

The Continuous Uniform Distribution can be used to model phenomena that

- A random variable X is **uniformly distributed** between a and b, if X takes value in any interval of a given size with equal probability.
  Discrete case: it takes any value in the support with equal probability
- the probability of X being in an interval, is proportional to the length of the interval.
  Discrete case: probability of a set is proportional to its size

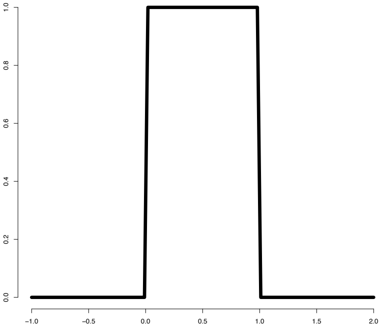
Example: the arrival of the bus 20 between the moment you get to the bus stop and midnight.

Continuous Uniform Distribution

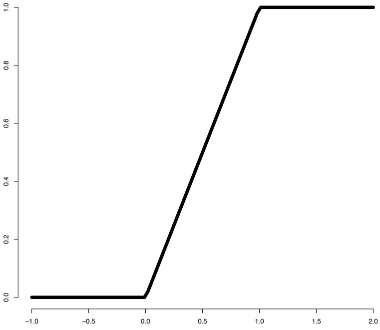
the formalization

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

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



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



Uniform c.d.f

how to use it

- In the case of a Uniform random variable there is a closed form (& easy to derive) expression for the c.d.f.:

$$
\begin{aligned}
F_X(x) &= \int_a^x \frac{1}{b-a} dt = \frac{1}{b-a} \left(t|_a^x \right) \\
&= \frac{x-a}{b-a} \qquad a \leq x \leq b
\end{aligned}
$$

- It is trivial to see that the probability of a set only depends on its size:

$$
\begin{aligned}
P(X \in [a_1, b_1]) &= F_X(b_1) - F_X(a_1) \\
&= \frac{b_1 - a}{b-a} - \frac{a_1 - a}{b-a} \\
&= \frac{b_1 - a_1}{b-a}
\end{aligned}
$$

Mean and Expected Value

do try this at home

› Expected Value of $X \sim \text{Unif}(a, b)$

$$\mathbb{E}[X] = \frac{a+b}{2}$$

Since it is a *location/center* summary, the expected value depends on the specific values the random variable assumes.

› Variance of $X \sim \text{Unif}(a, b)$

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

Since it is a *scale/dispersion* summary, the variance depends only on the size of the support.

Example

As the name suggest, a pay-per-kilo clothes shop (something like Pifebo) charges the customer based on the weight of what they are buying.

Empirical evidence suggest that a client typically buys between 200 and 800 gr of clothes.

- Probability Density Function:

$$f_X(x) = \begin{cases} \frac{1}{600} & 200 \leq x \leq 800 \\ 0 & \text{otherwise} \end{cases}$$

Exercise

your turn!

- What is the average amount of clothes bought?
- What is its variance?
- What is the probability that a customer buys less than 300 gr of clothes?

Exponential Distribution

the intuition

A random variable X is said to have an **Exponential Distribution** with parameter $\lambda > 0$, if its probability distribution can be written as

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

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

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

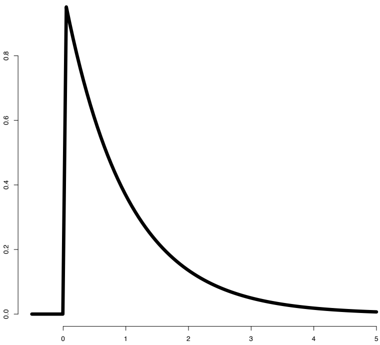
Examples: the amount of time until an earthquake occurs, the amount of money customers spend in one trip to the supermarket, the value of the change that you have in your pocket

Exponential distribution

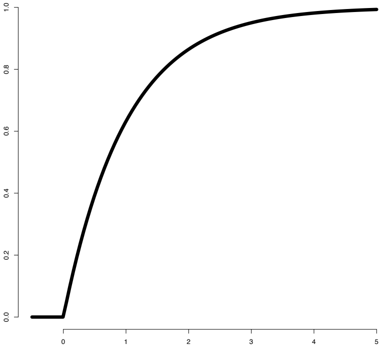
the formalization

$$X \sim \mathsf{Exp}(\lambda), \qquad \lambda > 0, \qquad x \geq 0$$

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



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

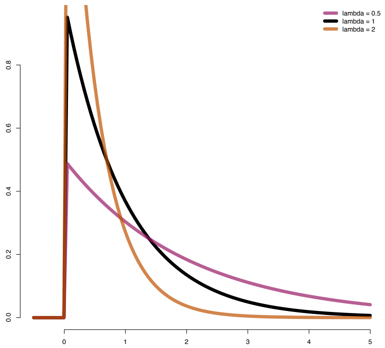


Exponential distribution

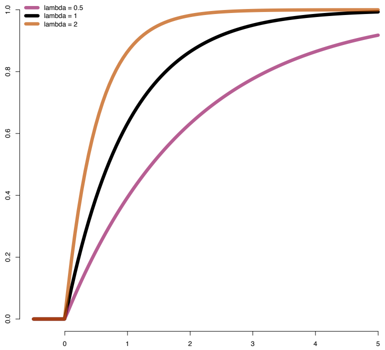
the formalization

$$X \sim \mathsf{Exp}(\lambda), \qquad \lambda > 0, \qquad x \geq 0$$

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



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

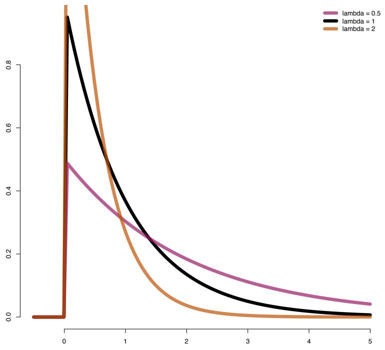


Expected Value and Variance

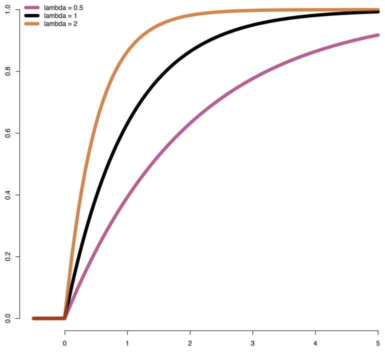
in pictures

$$X \sim \mathsf{Exp}(\lambda), \qquad \lambda > 0$$

$$\mathbb{E}[X] = \frac{1}{\lambda}$$



$$\mathbb{V}[X] = \frac{1}{\lambda^2}$$



Expected Value and Variance

in formulas

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x\lambda \exp\{-\lambda x\} dx \\
&= [-x \exp\{-\lambda x\}]_0^\infty + \int_0^\infty \exp\{-\lambda x\} dx \\
&= (0 - 0) + \left[-\frac{1}{\lambda} \exp\{-\lambda x\}\right]_0^\infty \\
&= 0 + \left(0 + \frac{1}{\lambda}\right) \\
&= \frac{1}{\lambda}
\end{aligned}
$$

Expected Value and Variance

in formulas

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_0^\infty x^2 \lambda \exp\{-\lambda x\} dx \\
&= [-x^2 \exp\{-\lambda x\}]_0^\infty + \int_0^\infty 2x \exp\{-\lambda x\} dx \\
&= (0 - 0) + \left[-\frac{2}{\lambda} x \exp\{-\lambda x\}\right]_0^\infty + \frac{2}{\lambda} \int_0^\infty \exp\{-\lambda x\} dx \\
&= (0 - 0) + \frac{2}{\lambda} \left[-\frac{1}{\lambda} \exp\{-\lambda x\}\right]_0^\infty \\
&= \frac{2}{\lambda^2}
\end{aligned}
$$

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

$$
\mathbb{V}[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
$$

Exercise

Bus waiting time

The waiting time (in minute) for the bus 20 is distrubuted as an Exponential random variable, with expected value 10 minutes.

- determine the parameter λ of the distribution
- determine the probability of waiting at least 15 minutes
- suppose you have been waiting for 10 minutes already, what is the probability of waiting at least an additional 15 minutes?

Exercise

Bus waiting time

$X \sim \text{Exp}(0.1)$

- determine the probability of waiting at least 15 minutes

$$P(X > 15) = 1 - F_X(15) = 1 - (1 - e^{-0.1*15}) = e^{-0.1*15}$$

- suppose you have been waiting for 10 minutes already, what is the probability of waiting an additional 15 minutes?

$$\begin{aligned} P(X > 15 + 10 | X > 10) &= \frac{P(X > 15 + 10 \cap X > 10)}{P(X > 10)} \\ &= \frac{P(X > 15 + 10)}{P(X > 10)} = \frac{e^{-0.1*(15+10)}}{e^{-0.1*10}} \\ &= e^{-0.1[(15+10)-10]} = e^{-0.1*15} \end{aligned}$$

Exercise

Bus waiting time

$X \sim \text{Exp}(0.1)$

- determine the probability of waiting at least 15 minutes

$$P(X > 15) = 1 - F_X(15) = 1 - (1 - e^{-0.1*15}) = e^{-0.1*15}$$

- suppose you have been waiting for 10 minutes already, what is the probability of waiting an additional 15 minutes?

$$\begin{aligned} P(X > 15 + 10 | X > 10) &= \frac{P(X > 15 + 10 \cap X > 10)}{P(X > 10)} \\ &= \frac{P(X > 15 + 10)}{P(X > 10)} = \frac{e^{-0.1*(15+10)}}{e^{-0.1*10}} \\ &= e^{-0.1[(15+10)-10]} = e^{-0.1*15} \end{aligned}$$

it's the same in both cases!

Properties

formalizing something we have seen already

- The Exponential is memoryless distribution

$$
\begin{aligned}
P(T \geq t + s | T \geq s) &= \frac{e^{\lambda(s+t)}}{e^{\lambda(s)}} \\
&= e^{\lambda t} = P(T \geq t)
\end{aligned}
$$

- The Exponential represent the waiting time between Poisson events

Normal Distribution

The **Normal** or **Gaussian** Distribution is the *queen* of the random variables, and this is because:

- it represents many natural and economic phenomena
- it approximates other distributions
- it is key to inference in sampling

A random variable $X \sim \text{Norm}(\mu, \sigma^2)$ has an interpretable parametrization:

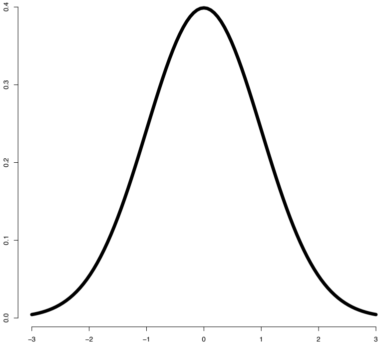
$$\mu = \mathbb{E}[X] \qquad \sigma^2 = \mathbb{V}[X]$$

Normal Distribution

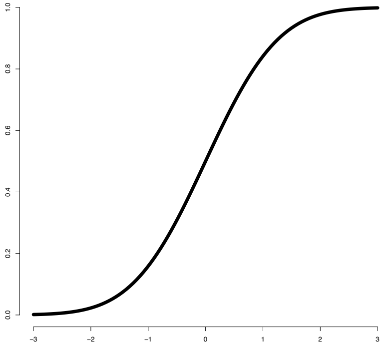
the formalization

$$X \sim \mathsf{Norm}(\mu, \sigma^2), \qquad \sigma^2 > 0, \mu \in \mathbb{R}$$

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



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



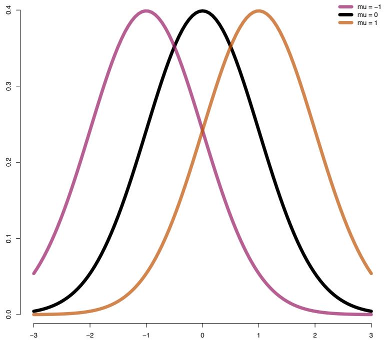
Normal Distribution

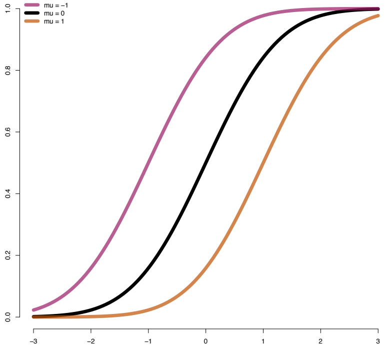
as the mean varies

$$X \sim \text{Norm}(\mu, \sigma^2), \qquad \sigma^2 > 0, \mu \in \mathbb{R}$$

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

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





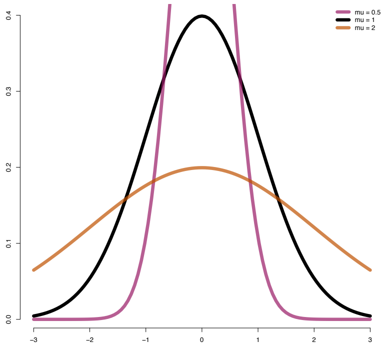
Normal Distribution

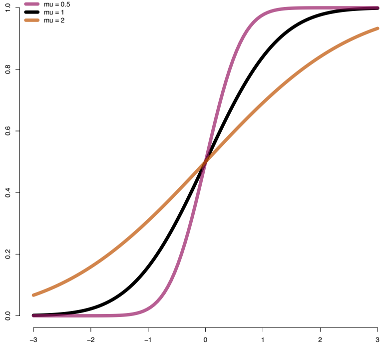
as the variance varies

$$X \sim \text{Norm}(\mu, \sigma^2), \qquad \sigma^2 > 0$$

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

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





Properties

the Normal is a *stubborn* distribution

- a linear transformation of a Normal random variable is still a Normal random variable:

  $X \sim \text{Norm}(\mu, \sigma^2)$, if $Y = aX + b$, where $a, b \in \mathbb{R}$

$$Y \sim \text{Norm}(a\mu + b, a^2\sigma^2)$$

- a linear combination of Normal random variables is still a Normal random variable:

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

$$Y = \sum_{i=1}^{n} a_i X_i \sim \text{Norm}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right),$$

Standard Normal

When $\mu = 0$ and $\sigma^2 = 1$, the random variable Norm$(0, 1)$ is called a **standard Normal** and it is denoted by Z.

Every Normal distribution can be turn into a standard Normal by means of **standardization**

If $X \sim$ Norm(μ, σ^2), then

$$Z = \frac{X - \mu}{\sigma} \sim \text{Norm}(0, 1)$$

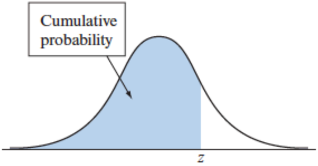
This is just a linear tranformation of a Normal, so it is easy to show:

$$\mathbb{E}[Z] = \mathbb{E}\left[\frac{X - \mu}{\sigma}\right] = \frac{\mathbb{E}[X] - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$
$$\mathbb{V}[Z] = \mathbb{V}\left[\frac{X - \mu}{\sigma}\right] = \frac{\mathbb{V}[X]}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

Tables of a standard Normal

what is the fuss about Standard Normal

Someone computed for you all the values of the cumulative distribution function of a Standard Normal and store them into **tables**.



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

Toy Example

The time (in minutes) you need to solve the exercises I gave you, X, is Normally distributed with mean $\mu = 5$ and standard deviation $\sigma = 10$.

Formally $X \sim \text{Norm}(5, (10)^2)$.

When I prepared this exercise at home, yesterday, it took me 6.2 minutes to solve it.

- **?** What is the probability to find someone faster than me, i.e. $P(X \leq 6.2)$

$$
\begin{aligned}
P(X \leq 6.2) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{6.2-5}{10}\right) \\
&= P(Z \leq 0.12) = 0.5478
\end{aligned}
$$

Exercise

The length of Black Mirror episodes (in minutes), is known to be Normally distributed with mean $\mu = 50$ and standard deviation $\sigma = 5$.

A new episode just got out:

- determine the probability that its length is exactly 50 minutes;
- determine the probability that its length is between 48 and 51 minutes;

A whole new season made of 8 episode is scheduled to be released next fall:

- determine the probability distribution of the length (in minutes) of the whole season;
- determine the expected length (in hours) of the whole season and its variance.

Central Limit Theorem

the intuition

Suppose you have $X_1, \dots, X_n$ random variables independent and with the same distribution.

Identical distribution implies that all the variables have the same expected value $\mu = \mathbb{E}[X_i]$ and variance $\sigma = \mathbb{V}[X_i]$

The average of this collection is also a random variable

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Even if we don't know the distribution of $\bar{X}$, the **Central Limit Theory** tell us that as $n \to \infty$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow Z$$

Central Limit Theorem

the consequences

- if $X_1, \dots, X_n$ are already Normals, then the result of the CLT is *exact*, that is, it works for any n

- even if we have no idea of what distribution generated the collection $X_1, \dots, X_n$, we can *always* (albeith **asymptotically**) derive a distribution for its mean

- the CLT is very useful in statistical inference. We typically consider our data as realization of a collection of random variables $X_1, \dots, X_n$ whose distribution we do not know; it is crucial to have a summary whose distribution we know in order to draw inferential conclusions.