Basics of Probability

Lecture 2

Francesco Dotto
francesco.dotto@uniroma3.it

2/9/2021

Introduction

Probability statements are everywhere around us. Examples of probability statements include

1. There is a 60 percent chance of its raining today.
2. The chance of me winning the lottery is one in 80 million.
3. There is a 50-50 chance of observing a head when a fair coin is tossed.

Just what is meant by chance in these statements? Chance is a measure of uncertainty, and we call this measure probability. In this lecture we will study this concept of probability.

Probability: a step by step definition (I)

Randomness suggests unpredictability. A simple example of *randomness*: when a coin is tossed, the outcome is uncertain.

The outcome could be either an observed head (H) or an observed tail (T). Because the outcome of the toss cannot be predicted for sure, we say that it displays randomness.

Probability: a step by step definition (II)

A random experiment is an experiment in which the outcome on each trial is uncertain and distinct.

Examples of random experiments are rolling a die, selecting items at random from a manufacturing process to examine for defects, selection of numbers by a lottery machine, etc.

Probability: a step by step definition (III)

When we toss a coin, we have two possible outcomes, summarized by $\{H, T\}$.

Such a list is called a **Sample Space**. Generally the sample space is indicated with the Greek letter Ω

Example : A fair regular six-sided die is rolled. List the sample space for this random experiment. Solution: Let Ω represent the sample space. Then $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Probability: a step by step definition (IV)

We may only be interested in part of the sample space. For example we may be concerned with even numbers $A = \{2, 4, 6\}$. This is a *subset* of the sample space. Such subsets are called **events**

If the outcomes in a sample space are equally likely to occur, then the **classical probability of an event A** is defined to be

$$P(A) = \frac{\text{Number of simple events in A}}{\text{Total number of simple events in the sample space}}$$

Probability: Some simple examples

When a child is born, the gender of the child is either a boy (B) or a girl (G), summarized by $\Omega = \{B, G\}$. If we consider a two-child family, the possibilities can be summarized by $\Omega = \{BB, BG, GB, GG\}$.

If a two-child family is selected at random, what is the probability of there being two boys in the family?

the event of two boys occurs once, and there are four simple events in the sample space, thus

$$P(BB) = \frac{1}{4} = 0.25$$

Set theory

Basic notions

Mathematically speaking events are sets.

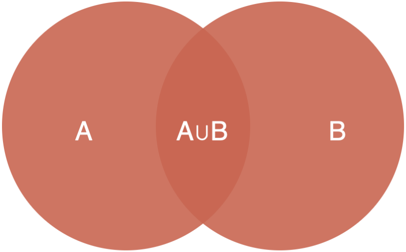
Definition: A set is a finite or infinite collection of objects

Sets can be **finite** (contains a finite number of objects) or **infinite** (contains an infinite number of objects)

The **cardinality** of a given set is the number of objects that belong to such set. I.e. if $E = \{1, 2, 3\}$, the cardinality of E, $\#E = 3$,

Recap of Set theory

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

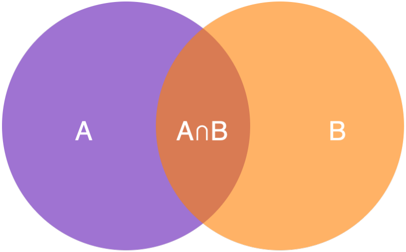


Example $A =$ "the die returns an even number", $B =$ "the die returns a 5"

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

Recap of Set theory

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



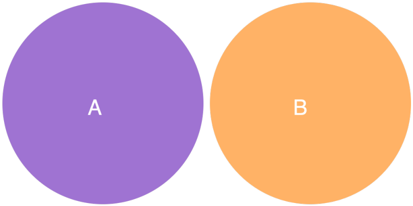
Example $A =$ "the die returns an even number", $B =$ "the die returns a number smaller than 5"

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

Recap of Set theory

Empty intersection

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

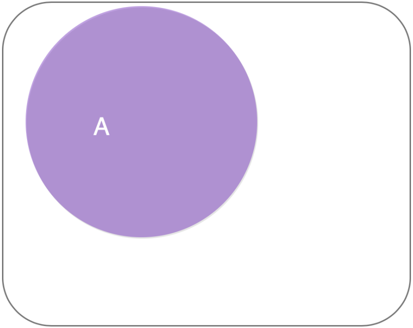


Example $A =$ "the die returns an even number", $B =$ "the die returns a 5"

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

Recap of Set theory

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

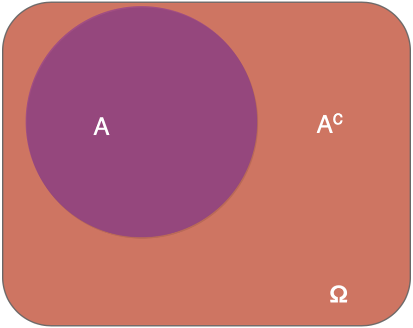


Example $A =$ "the die returns an even number"

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

Recap of Set theory

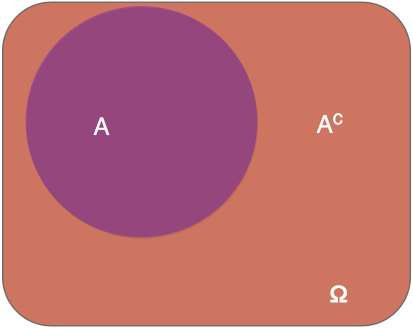
- **Complement** (A^c or $\bar{A}$) everything that is not in A.



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

Recap of Set theory

- **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"

Probability

A formal definition of probability is hard to be given. Practically speaking, the **probability** is a **set function** on which the following holds.

Probability Axioms...

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

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

The basic ingredients - an evergreen Example

Random phenomenon: throw of a die

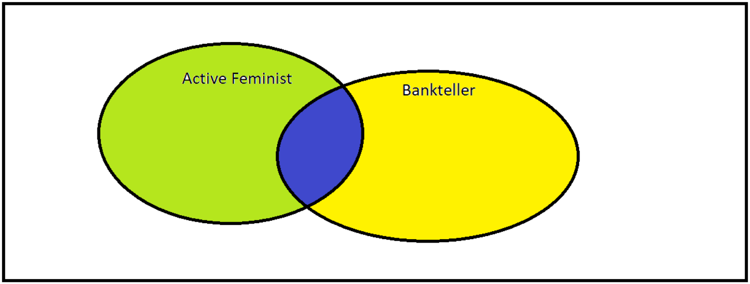
- **Sample Space**: all the possible outcomes
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
- **Event**: "the die returns an even number"
  - $E = \{2, 4, 6\}$
- **Probability**:
  - $P(E) = \frac{\#E}{\#\Omega} = \frac{1}{2}$

Exercise :: challenging

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

- Which is more probable?
  - Linda is a bank teller.
  - Linda is a bank teller and is active in the feminist movement.

Linda in a Venn diagram



Set Theory is important

For computing probabilities

Exercise "students" 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.

1. Compute probabilities for these events, and depict this information on a Venn diagram.

Let C be the event that a student owns a car, and let D be the event that a student owns a stereo. Thus $P(C) = 0.6$, $P(D) = 0.3$, and $P(C \cap D) = 0.1$

2. Compute the probability that a student has car but hasn't a stereo

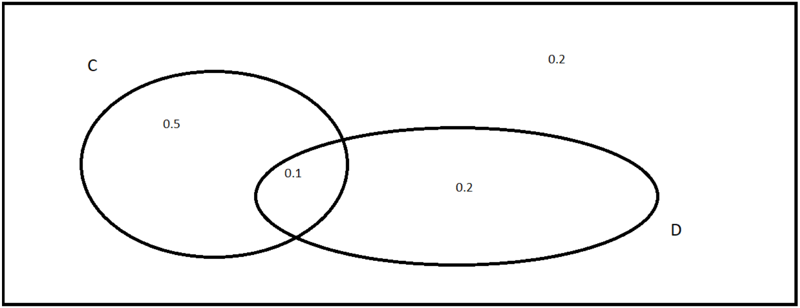
$$P(\text{“Only a Car”}) = P(C) - P(C \cap D) = 0.6 - 0.1 = 0.5$$

Exercise “students”

Venn Diagram

3. Compute The probability that a student has either a car or a Stereo

$$P(\text{“A Car OR A Stereo”}) = P(C \cup D) = P(C) + P(D) - P(C \cap D) =$$
$$= 0.6 + 0.3 - 0.1$$



Exercises

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

1. Get the sample space.
2. Find the probability that two heads are obtained.

- A card is drawn at random from a deck of cards. Find the probability of getting a diamond.

Exercises

- Which of the following is an impossible event?

1. Choosing an odd number from 1 to 10.
2. Getting an even number after rolling a single 6-sided die.
3. Choosing a white marble from a jar of 25 green marbles.
4. None of the above.

- There are 4 parents, 3 students and 6 teachers in a room. If a person is selected at random, what is the probability that it is a teacher or a student?

1. $4/13$
2. $7/13$
3. $9/13$
4. None of the above.

Probability and Relative Frequencies

The relative frequency probability of an event's occurring is the proportion of times the event occurs over a given number of trials. If A is the event in which we are interested, then the relative frequency probability of A's occurring, denoted by $P(A)$, is computed from

$$P(A) = \frac{\text{frequency of occurence}}{\text{number of trials}}$$

Probability and Relative Frequencies

For the first 43 presidents of the United States, 26 were lawyers. What is the probability of randomly selecting from those 43 presidents a president who was a lawyer?

Solution: Let A represent the event of a president being a lawyer. Thus, since there are 43 presidents and 26 were lawyers, then

$$P(A) = \frac{26}{42} = 0.605$$

Conditional Probability

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

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

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

Conditional probability

An example with Student Exercise

Known Data Let C be the event that a student owns a car, and let D be the event that a student owns a stereo. Thus $P(C) = 0.6$, $P(D) = 0.3$, and $P(C \cap D) = 0.1$

Problem what is the probability of a student's having a stereo given the student has a car?

$$P(D|C) = \frac{P(D \cap C)}{P(C)} = \frac{0.1}{0.6} = 0.167$$

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.

Dependence

There is a relation between events

Recalling the conditional probability formula

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

Exercises

- Three cards are chosen at random from a deck without replacement. What is the probability of choosing an eight, a seven and a six, in order?
  - 6/35152
  - 1/2197
  - 8/16575
  - None of the above.

- A jar contains 5 red, 3 green, 2 purple and 4 yellow marbles. A marble is chosen at random from the jar. After replacing it, a second marble is chosen. What is the probability of choosing a purple and then a red marble?
  - 5/98
  - 1/2
  - 3/98
  - 2/49

Combinatory Calculus

To compute probabilities we will often need a method for detrmining in **How many ways** a given phenomenon can happen. E.g. Tossing a coin: *How many times will I obtain two Heads?*

The complete list of the possible outcomes is geven below. We can see that the table below has 4 possible outcomes.

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

Combinatorics is abranch of mathematics which is about *counting*

Fundamental principle of combinatorics

Fundamental principle of combinatorics

Given an experiment with n possible outcomes and and another experiment with m possible outcomes, then the combination of the two experiment has $m \times n$ possible outcomes.

In the previous case tossing coin 1 has 2 possible outcomes and the same holds for coin 2. Thus, there are 2×2 possible outcomes.

Permutations: An example

There are $n = 9$ students attending the statistics precourse (I hope that the number does not decrease as the time goes by!). Let us suppose that all the students have to seat on a chair and let us suppose that the chairs are disposed on a straight line. How many possible lines can be formed by changing the position of each student?

Answer The first student can choose his position in 9, different ways, the second student has 8 possibilities, the third student has 7 possibilities and so on. Thus, there are

$$9 \times 8 \times 7 \times 6 \times \cdots \times 1 = 9! \tag{1}$$

ways to place 9 students in a line.

Permutations: A formal definition

Permutation of objects

Given a set of n a **permutation** is a given ordering of those objects. The number of possible permutations of n objects is equal to $n!$

Exercise

In how many different ways could 23 children sit on 23 chairs in a Maths Class? If you have 4 lessons a week and there are 52 weeks in a year, how many years does it take to get through all different possibilities? Note: The age of the universe is about 14 billion years.

Permutations and universe's age

For $n = 23$ children to sit on 23 chairs there are $23! = 25,852,016,738,884,800,000,000$ possibilities (this number is too big to be displayed on a calculator screen). Trying all possibilities would take

$$\frac{23!}{4 \times 52} = 124,288,542,000,000,000,000 \text{ years.} \tag{2}$$

This is nearly 10 million times as long as the current age of the universe!

Permuting a subset of observations: Dispositions

Let us consider the example of the students enrolled in the statistics precourse. Let us now suppose that only $k = 6$ chairs are available for $n = 9$ students. How many ways can I dispose $n = 9$ students on $k = 6$ chairs?.

The first student can choose a chair in 9 different ways, the second one can choose his chair in 8 different ways and so on. Thus there are $9 \times 8 \times 7 \times 6 \times \cdots \times 4$ ways to dispose 9 students on six chairs.

Dispositions: A formal definition

Dispositions

Given a set of n objects a **disposition** is a way of choosing k elements out of a set of n elements without repetitions and taking in account of their order. The number of possible of k disposition out of n objects is equal to

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

The way we wrote dispositions' formula *may* look more complicate than the way we actually found the number of possible disposition. Obviously they are equivalent, infacts, the following holds:

$$\frac{n!}{(n-k)!} = \frac{n \times (n-1) \times (n-k+1) \times (n-k) \times \dots \dots (n-n+1)}{n \times (n-1) \times (n-k+1) \times (n-k) \times \dots (n-n+1)}$$

Combinatorics: Combinations (i)

Does the ordering matter?

Let's go back to the example of the nine students enrolled in the statistics precourse. Let us suppose that we only have six chairs on which the students can sit. In how many different ways can I dispose $n = 9$ students on $k = 6$ chairs regardless of the order of the chairs?

Answer We found that, if we keep in account the *ordering*, we can dispose the students in $\frac{n!}{(n-k)!}$ different ways. Now we are not interest in their ordering, thus disposition as 1,2,3,4,5,6 is equivalent to 6,2,3,4,5,1. Thus we have to divide $\frac{n!}{(n-k)!}$ by the number of possible permutations of k objects which is equal to $k!$

Combinatorics: Combinations (i)

A formal definition

Permutation of objects

Given a set of n objects a combination is a way of choosing k elements out of a set of n elements without repetitions and **without** taking in account of their. The number of possible k combinations out of n objects is equal to

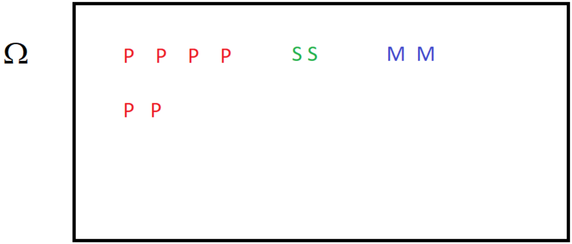
$$\frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Random variables

From the sample space to random variables

So far we introduced the sample space Ω and we computed probabilities associated to subset of the sample space.

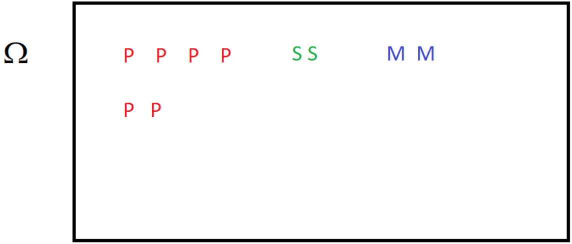
P= “Primary School”, S=“Secondary School” M=“Master”



For example we learnt how to compute $P(``P")$ or $P(``M")$ and so on...

Random variables

Coneptually speaking



Now our probabilistic experiment is to choose a student at random and record the number of years schooling (Y)

1. $P \rightarrow 5$
2. $S \rightarrow 10$
3. $M \rightarrow 15$

Random variables

Black Box

You can think Y as an abstract box that takes as input a student and produces a number y which is the number of years schooling of that particular student

$$\text{Student} \rightarrow \boxed{Y} \rightarrow y \text{ (years of schooling)}$$

just like a standard mathematical function

$$x \rightarrow \boxed{f} \rightarrow f(x) \text{ (years of schooling)}$$

Notation

Y : Random Variable y : Numerical Value

Random variable

Going back to the years of schooling

The random variable (r.v.) Y **randmoly picks up** the student and gives back in output his/her years of schooling. Thus there is a probability law associated to its values y (this i generally called **probability distribution**)

$$P(x) = \begin{cases} 6/10 & \text{if} \quad x = 5 \\ 2/10 & \text{if} \quad x = 10 \\ 2/10 & \text{if} \quad x = 15 \end{cases}$$

Random variable

Little bit more formally

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

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

In formulas:

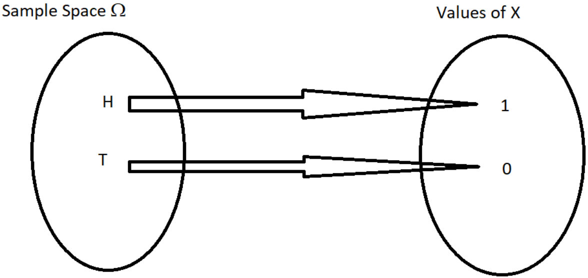
$$X : \Omega \to \mathbb{R}$$

N:B: As the random variable is defined on **the sample space** we can associate a probability to the values that the random variable assumes

Random variable

how to define it

Example: Single toss of a Coin. $\Omega = \{H, T\}$. We define X as the random variable that counts the number of Heads



Random Variables

Some additional 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!

Random variables: discrete and continous

- A **discrete** random variable is a variable that represents numbers found by counting. For example: number of marbles in a jar, number of students present or number of heads when tossing two coins.
- When we have to use intervals for our random variable or all values in an interval are possible, we call it a continuous random variable. Thus, continuous random variables are random variables that are found from measuring - like the height of a group of people or distance traveled while grocery shopping or student test scores. In this case, X is continuous because X represents an infinite number of values on the number line.

Distribution of a random variable

an example of how to derive it

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

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	$1/8 \times 1 = 1/8$
1	$1/8 \times 3 = 3/8$
2	$1/8 \times 3 = 3/8$
3	$1/8 \times 1 = 1/8$

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

Constructing the probability distribution

- Ω represents the sets of all the possible outcomes after tossing two coins
- X Represents the number of tails after tossing 2 coins
- $\mathcal{X} = \{0, 1, 2, 3\}$

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

$$P(x) = \begin{cases} 1/8 & \text{if} \quad x = 0 \\ 3/8 & \text{if} \quad x = 1 \\ 3/8 & \text{if} \quad x = 2 \\ 1/8 & \text{if} \quad x = 3 \end{cases}$$

Exercise

- Two dice are rolled:

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

Distribution of a Discrete Random Variable:

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

- **Probability mass distribution**

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

- **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?

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

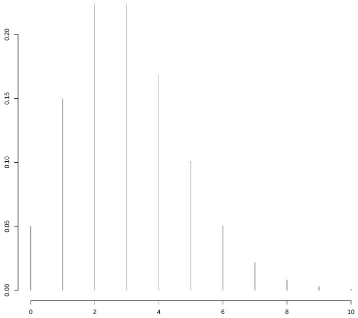
- What is the probability of 0 **or** 2 heads? (*disjoint events!*)

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

Properties

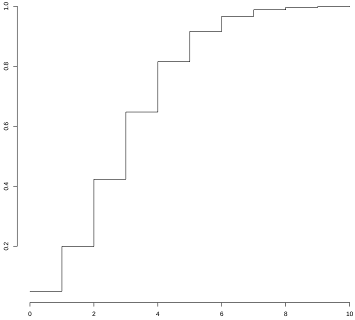
Probability mass distribution

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



Cumulative distribution function

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



Exercise

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

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

- Calculate the probability function.
- Calculate the following probabilities

a. P(X=6)
b. P(X=5)
c. P(2<X<5.5)
d. P(0 X<4)

Discrete Random Variable: Expected Value

In words

A very important concept in probability is the idea of expected values. The expected value for a random variable is the long-term mean or average value of the random variable. If the random variable is observed over a long period of time, the expected value should be close to the average value of the observations generated by the random process. The larger the number of observations, the closer the expected value will be to the average value of the observations.

Discrete Random Variable: Expected Value

In formulas

The expected value for a discrete random variable X is the mean value of the random variable. It is denoted by $\mathbb{E}[X]$ and is obtained by computing

Expectation: Formula

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot p(x)$$

\end{tcolorbox}

Discrete Random Variable: Expected Value

An example (i)

Consider the following game.

*You flip a fair coin. Every time you get a head, you lose $1\$$. Every time you get a Tail you get $2\$$

Would you play to this game?

Discrete Random Variable: Expected Value

An example (ii)

The game considered in the previous example **is a random variable** whose distribution is given by

$$P(x) = \begin{cases} 1/2 & \text{if} \quad x = -1 \\ 1/2 & \text{if} \quad x = 2 \end{cases}$$

and such that

$$\mathbb{E}[X] = -1 \cdot 1/2 + 2 \cdot 1/2$$

In such a game you are expected to gain money over time, so you should play this type of game

Expectation: a method to evaluate if a game is fair

The Roulette is maid of 38 numbers (from 0 to 37). Suppose every player bets $1\$$ on a given number. If the extracted number is equal to the one chosen by the player the player wins $35\$$ dollars. Otherwise he loses the dollar he bet. The distribution of random variable "Amount of money the player wins at Roulette'' is given by

$$P(x) = \begin{cases} 1/38 & \text{if} \quad x = 35 \\ 37/38 & \text{if} \quad x = -1 \end{cases}$$

Thus its expectation is given by

$$\mathbb{E}[X] = -1 \cdot 37/38 + 35 \cdot 1/38 = -0.0525$$

This means that over the long run, you should expect to lose on average about 5 cents each time you play this game. Yes, you will win sometimes. But you will lose more often.

Expectation: Final remarks

Expectation: Interpretation

Warning: The expected value really ought to be called the expected mean. It is NOT the value you most expect to see but rather the average (or mean) of the values you see over the course of many trials.

Variance of a Random Variable

in words

Recap: The expected value of a random variable X tells us that if we observe N random values of X, then the mean of the N values will be approximately equal to $\mathbb{E}[X]$ for large N.

Besides of the value around which a given random variable is oncentrated, we may also be interested in how the random variable X is spread out.

A measure of how spread out a random variable is, is given by its **variance**

Variance of a Random Variable

in words

Variance of a Random Variable

The variance is the mean squared deviation of a random variable from its own mean. Its formula is given by:

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \qquad (3)$$

Morning Reminescences

Remind that, this morning, we defined the variance of a variable collected on a set of statistical units as

$$s^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \bar{y}^2$$

Continuous Random Variable:

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

Example For example, the amount (in inches) of rainfall in your community during the month of March is an example of a continuous random variable. If X is the amount it rained during the month of March, then the possible values for X will be in the interval $(0, \infty)$. That is, the amount can vary from zero inches to an infinite number of inches. Theoretically, the number of inches of rainfall can go to infinity ∞, but from a practical standpoint, this may never happen. A practical continuous interval may be $[0, 12]$ inches.

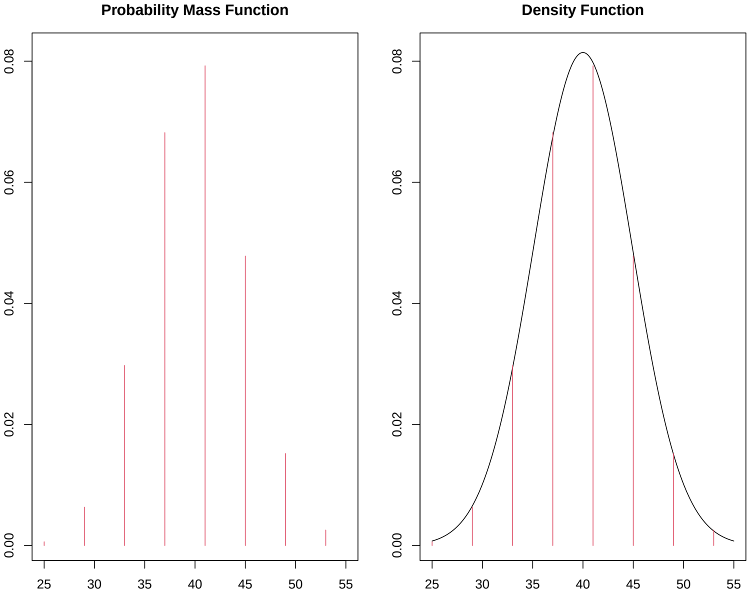
Comparison

In words

1. A discrete random variable takes values in a discrete set. Its PMF tells us exactly how much probability is assigned to each point. To calculate the probability that the random variable lies in a given interval we have to sum the masses associated to the points composing the interval
2. A continous random variable take values over a continous range. In this case the mass assigned to each point is given by the so called density function. To calculate the probability that the random variable lies in a given interval we have to calculate the area under the density curve.

Comparison

with plots



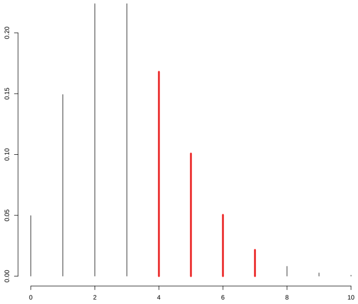
Comparison

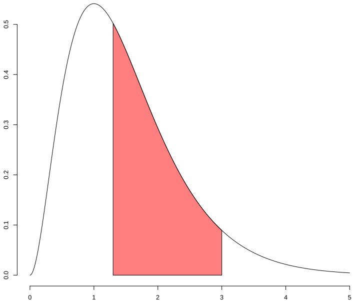
$$A = \{x_1, \dots, x_k\}$$

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

$$A = [a, b]$$

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





Continuous Random variables: Additional informations

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

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

- **Cumulative distribution function**

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

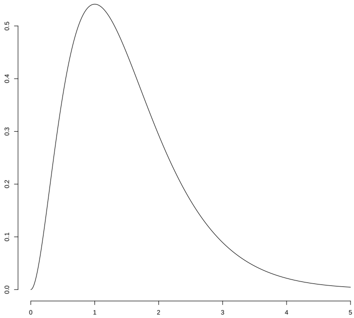
- **Probability density distribution**

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

Properties

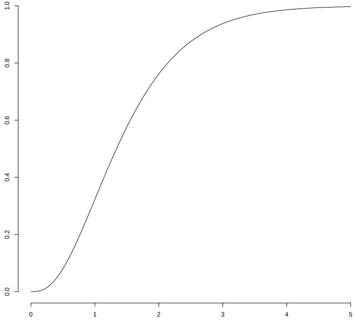
Probability density function

- $f_X(x) \geq 0$
- $f_X(x)$ **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

- 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.

Comparison

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

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

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

- X continuous rv with pdf $f_X(x)$
- $P(X \in A) = \int_A f_X(x)\mathrm{d}x$

if $A = [a, b]$ then

$$P(X \in A) = \int_a^b f_X(x)\mathrm{d}x$$
$$= F_X(b) - F_X(a)$$

Summaries

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

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

Candidates:

- *Mode*: the value that is "more likely", i.e. the value that maximizes 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

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

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

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

Measuring Variability

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

Example:

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

(Bad) Candidates:

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

Variance

queen of all summaries

The **variance** of a random variable X

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

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

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

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

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

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

Properties

- the variance is always **non-negative**, $\mathbb{V}[X] \geq 0$ and is 0 only when X is constant

- the square root of the variance $\text{sd}(X) = \sqrt{\mathbb{V}[X]}$ is called **standard deviation**. Roughly, describes how far values of the random variable fall, on the average, from the expected value of the distribution

- the variance is *insensitive to the location* of the distribution but **depends only on its scale**

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

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

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

Exercise

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

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

- Calculate mean and variance

Exercise

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

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

- Calculate mean and variance

Covariance

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

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

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

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

Independence of Random Variables

Two random variables X, Y are independent if

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

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

Remark: if $X_1, \ldots, X_n$ are independent then

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

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 [1/2]

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

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

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

Exercise [2/2]

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

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