

# Monte Carlo methods

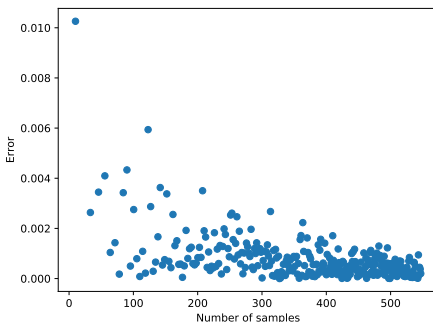
October 24, 2019

# Objective

- ▶ We will discuss Monte Carlo methods
- ▶ It will be sometimes technical.
- ▶ However you don't need to understand all technical details in order to apply these ideas to the project (if you are interested in doing so, which is not mandatory).
- ▶ We will discuss possible applications to the game.

## Defining the problem

- Facing a random process, we would like to compute its **expected value**.



# Defining the problem

- ▶ Facing a random process, we would like to compute its **expected value**
- ▶ For instance
  - ▶ what is the mean amount of rain one can expect in october in Paris ?

# Defining the problem

- ▶ Facing a random process, we would like to compute its **expected value**
- ▶ For instance
  - ▶ what is the mean amount of rain one can expect in october in Paris ?
  - ▶ If I play a game, what is my expected gain ?

# Defining the problem

- ▶ We are in a situation where we have some information about the random process.
  - ▶ We know its **probability density** or **distribution**.

# Defining the problem

- ▶ We are in a situation where we have some information about the random process.
  - ▶ We know its **probability density** or **distribution**.
  - ▶ However, it is not straightforward to explicitly compute the **expected value**.

# Defining the problem

- ▶ We are in a situation where we have some information about the random process.
  - ▶ We know its **probability density** or **distribution**.
  - ▶ However, it is not straightforward to explicitly compute the **expected value**.
  - ▶ We will need to compute an **approximate value** for the expectation.



# Defining the problem

- ▶ The Monte-Carlo method uses **simulated random variables** to compute such an approximate value.

## Question

- ▶ But why should we use a method involving randomness ?

# Overview

Expected values

The Monte Carlo Method

- The law of large numbers

- Central limit theorem

- Random variables simulations

Why is Monte-Carlo useful ?

- Notion of algorithmic complexity

Application to the game

## Expected values

- ▶ Let us study the **expected value**

## Expected values

- ▶ The **expected value** (or expectation) is a **weighted average** of a random variable.

## Example 1

- ▶ The expected value is a **weighted average** of a random variable.
- ▶ What is the expectation of a single throw of an unbiased dice ?



Figure: Dice

## Formal definition

- Is the random variable  $X$  can take a finite number of values  $x_i$  with probabilities  $p_i$ , then the expected value is :

$$E(X) = \sum_{i=1}^n p_i x_i \quad (1)$$

# Deterministic computation

## Exercise 1: Computing an expected value.

- ▶ Let us consider the following situation. We have  $n$  computers.
  - ▶ Computer 1 transmits a message to computer 2.
  - ▶ Computer 2 transmits the received message to computer 3.
  - ▶ ...



# Deterministic computation

## Exercise 1: Computing an expected value.

- ▶ Let us consider the following situation. We have  $n$  computers.
  - ▶ Computer 1 transmits a message to computer 2.
  - ▶ Computer 2 transmits the received message to computer 3.
  - ▶ ...
- ▶ At each step, the probability that there is a mistake in the transmission is  $p$ .

# Deterministic computation

## Exercise 1: Computing an expected value.

- ▶ Let us consider the following situation. We have  $n + 1$  computers.
  - ▶ Computer 1 transmits a message to computer 2.
  - ▶ Computer 2 transmits the received message to computer 3.
  - ▶ ...
- ▶ At each step, the probability that there is a mistake in the transmission is  $p$ .
- ▶ Let  $X$  be the total number of mistakes done during the transmission to the last computer. (we have  $n$  transmissions between  $n$  computers)

# Deterministic computation

## Exercise 1: Computing an expected value.

- ▶ What is the law of  $X$  ?
- ▶ ie: for each  $k \in [0, n]$ , what is  $P(X = k)$  ?

# Deterministic computation

## Exercise 1: Computing an expected value.

- ▶ Can we check that our result is correct ?
- ▶ We need that :
  - ▶  $\forall k \ p_k \geq 0$
  - ▶  $\sum_{k=0}^n p_k = 1$

# Deterministic computation

## Exercise 1 : Computing an expected value.

- ▶ Please write a program that computes the expected value of  $X$  !

## Law of X

- ▶ This law is called the **binomial law**

## Remark

- ▶ If  $X$  is a random variable, any function  $f(X)$  of  $X$  is also a random variable.

# Generalisation

- ▶ Up to now, we studied **discrete, finite** random variables.



# Generalisation

- ▶ Up to now, we studied **discrete, finite** random variables.
- ▶ But we often encounter **continuous** random variables.

## Generalisation

- ▶ Up to now, we studied **discrete, finite** random variables.
- ▶ But we often encounter **continuous** random variables.
- ▶ The gaussian law  $\mathcal{N}(\mu, \sigma^2)$  is continuous, defined by a **density**  $f(x)$ .

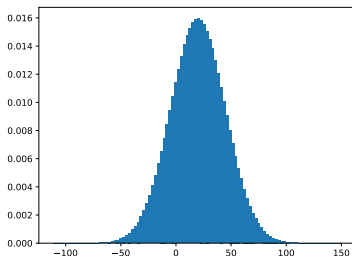


Figure: Normalized histogram

## Expected value of continuous variables

- How can we express the expected value of a continuous variable  $X$  that has a density  $f(X)$  ?

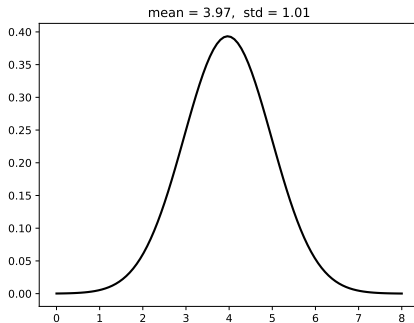


Figure: Probability density (normal law)

## Expected value of continuous variables

- ▶ How can we express the expected value of a continuous variable  $X \in \mathbb{R}$  that has a density  $f(X)$ .



$$E(X) = \int_{\mathbb{R}} xf(x)dx \quad (2)$$

## Expected value of continuous variables

- For instance the expected value for the gaussian law writes :

$$E(X) = \int_{\mathbb{R}} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = ? \quad (3)$$

# Careful !

- ▶ Sometimes the expected value does not exist !

# Careful !

- ▶ Sometimes the expected value does not exist !
- ▶ Can you think of examples ?

# Careful !

- ▶ Let us consider the random variable  $Y$  defined by
  - ▶  $Y = e^{X^3}$
  - ▶ where  $X \sim N(\mu, \sigma^2)$



# Careful !

- ▶ Let us consider the random variable  $Y$  defined by
  - ▶  $Y = e^{X^3}$
  - ▶ where  $X \sim N(\mu, \sigma^2)$
- ▶ The expected value would be

$$\int_{\mathbb{R}} e^{x^3} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = +\infty \quad (4)$$

- ▶ There is **no expected value**

# Variance

- ▶ The **variance** of a random variable is a measure of the variations around the mean.

# Variance

- ▶ The **variance** of a random variable is a measure of the variations around the mean.



$$V(X) = E((X - E(X))^2) \quad (5)$$

## Exercise 2 : Famous rule

- Please show that :

$$V(X) = E(X^2) - E(X)^2 \quad (6)$$

## Back to our problem

- ▶ Until now, we studied random variables where we can either explicitly compute the expectation, or write a very simple program to compute it.

## Back to our problem

- ▶ Until now, we studied random variables where we can either explicitly compute the expectation, or write a very simple program to compute it.
- ▶ But we are interested in a situation where it is not easy to compute the expectation. For instance when we want the expectation of some function  $g$  of a random variable of density  $f$  :

$$E[g(X)] = \int_{\mathbb{R}} f(x)g(x)dx \quad (7)$$

# Objective

- We want an approximation of this object :

$$E[g(X)] = \int_{\mathbb{R}} f(x)g(x)dx \quad (8)$$

# Objective

- ▶ We want an approximation of this object :

$$E[g(X)] = \int_{\mathbb{R}} f(x)g(x)dx \quad (9)$$

- ▶ Several methods exist :
  - ▶ Deterministic methods
  - ▶ Random methods (such as Monte-Carlo)



## Random methods

- ▶ Now let us discuss today's topic, the Monte Carlo method

# The law of large numbers

- ▶ The fundamental idea behind the Monte Carlo method is the following theorem

## Theorem

*Let  $(X_n)_{n \in \mathbb{N}}$  be a sequence of real random variables, independent and identically distributed. We assume that  $E(|X_1|) < +\infty$ . Then*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{a.s.} E(X) \quad (10)$$

# The law of large numbers

- ▶ Let us apply this idea to our problem. If  $X$  is a random variable distributed with a probability density  $f(x)$ . We want to compute, the expectation  $E[g(X)]$  for some function  $g$ .
- ▶ Is  $(X_i)_{i \in \mathbb{N}}$  is a sequence of **i.i.d** random variables with density  $f$ , then

$$\frac{1}{n} \sum_{i=1}^n g(X_i) \rightarrow E[g(X)] \quad (11)$$

## Remark

- If we simply want to compute the expectation  $E[X]$ , then

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow E[X] \quad (12)$$

# Method

- So all we need to do is being able to **simulate** i.i.d. random variables with the relevant density  $f$ .

# Applying Monte Carlo

## Exercise 3: Computing an expectation

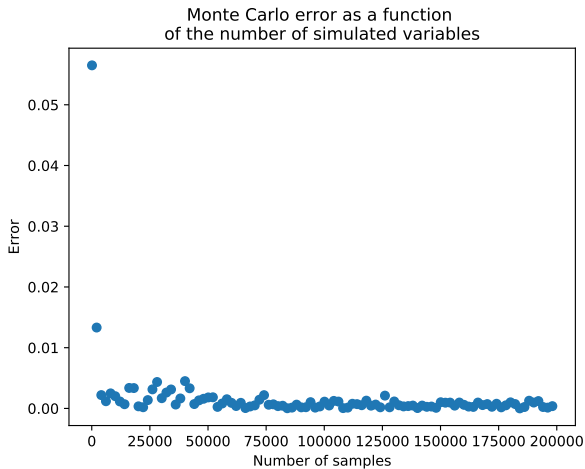
- ▶ Let us compute the expectation of the random variable  $U^2$ , where  $U$  is a **uniform random variable on**  $[0, 1]$
- ▶ Hence, you will need to **simulate** a uniform random variable.

# Applying Monte Carlo

## Exercise 3: Computing an expectation

- ▶ Please plot the **error of the estimation** as a function of the number of samples used.

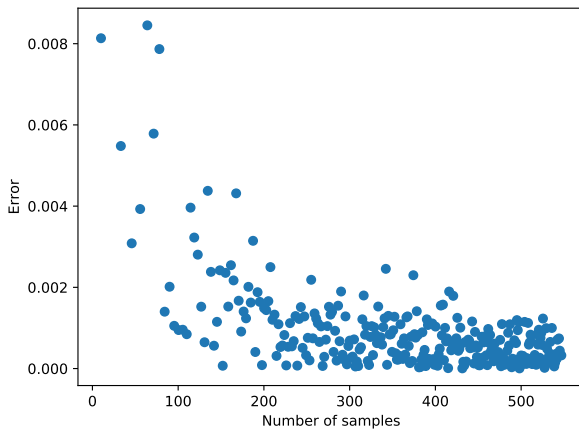
# Applying Monte Carlo





# Applying Monte Carlo

Monte Carlo error as a function of the square root of number of simulated variables



## Error and number of samples

- ▶ We need a result that tells us how much simulation we need to perform in order to trust our result.

# Speed of convergence

- ▶ How many variables  $X_i$  should we simulate ?
- ▶ i.e : what  $n$  should we choose ?

# Central limit theorem

- ▶ This theorem tells us that with the same hypothesis as before **and** a new condition  $E(X_1^2) < +\infty$  :

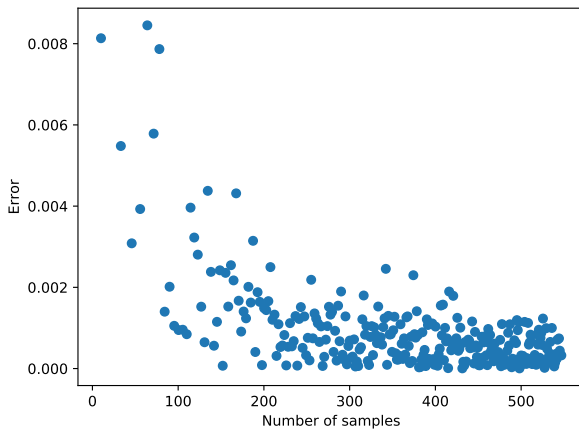
$$\frac{\sqrt{n}}{\sigma} \left( \frac{X_1 + \cdots + X_n}{n} - E(X_1) \right) \xrightarrow[n \rightarrow +\infty]{distribution} \mathcal{N}(0, 1) \quad (13)$$

# Error

- ▶ The theorem tells us that the error decays as a function of  $\sqrt{n}$

# Error

Monte Carlo error as a function of the square root of number of simulated variables



# Central limite theorem

- ▶ Let  $\epsilon_n$  be the error  $(\frac{X_1 + \dots + X_n}{n} - E(X_1))$
- ▶ The Central limit theorem tells us that **in distribution**,

$$\frac{\sqrt{n}}{\sigma} \epsilon_n \xrightarrow[n \rightarrow +\infty]{distribution} \mathcal{N}(0, 1) \quad (14)$$

# Central limite theorem

## Exercise 4 : Value of n

- ▶ Let  $\epsilon_n$  be the error  $(\frac{X_1 + \dots + X_n}{n} - E(X_1))$
- ▶ The Central limit theorem tells us that **in distribution**,

$$\frac{\sqrt{n}}{\sigma} \epsilon_n \xrightarrow[n \rightarrow +\infty]{\text{distribution}} \mathcal{N}(0, 1) \quad (15)$$

- ▶ For what value of  $n$  can we say that the error is smaller than 0.01 with probability 0.95 ?



## Remark

- ▶ The variance  $\sigma$  of the random variables appears in the estimator !

# Simulation of non uniform random variables

- ▶ Let us now assume that we need the expectation of a random variable that is **not uniform**.

# Cumulative distribution function

- To do so, we will need the Cumulative distribution function

$$F(x) = P(X \leq x) \quad (16)$$

# Cumulative distribution function

- ▶ To do so, we will need the Cumulative distribution function

$$F(x) = P(X \leq x) \quad (17)$$

- ▶  $F$  is monotonically increasing

# Pseudo inverse

- We introduce the pseudo inverse  $F^{-1}$ .

$$\forall x \in [0, 1], F^{-1}(u) = \inf\{y \in \mathbb{R}, F(y) \geq u\} \quad (18)$$

# Pseudo inverse

- We can show that  $\forall u \in [0, 1], x \in \mathbb{R}$

$$F^{-1}(u) \leq x \Leftrightarrow u \leq F(x) \quad (19)$$

# Pseudo inverse

- ▶ We can show that  $\forall u \in [0, 1], x \in \mathbb{R}$

$$F^{-1}(u) \leq x \Leftrightarrow u \leq F(x) \quad (20)$$

- ▶ and that if  $U$  is a uniform law on  $[0, 1]$ , then the random variable  $F^{-1}(U)$  is a random variable with a cumulative distribution function of  $F$ .

# Exponential law

## Exercise 5 : Computing a pseudo inverse.

- ▶ Let us introduce the **exponential law**.
- ▶ Its density is

$$f(x) = \lambda \exp(-\lambda x) \quad (21)$$

for  $x \geq 0$  and 0 otherwise.

- ▶ Please compute its cumulative distribution function.



# Exponential law

## Exercise 5 : Computing a pseudo inverse.

- ▶ Let us introduce the **exponential law**.
- ▶ Its density is

$$f(x) = \lambda \exp(-\lambda x) \quad (22)$$

for  $x \geq 0$  and 0 otherwise.

- ▶ Please compute its cumulative distribution function  $F$ .
- ▶ What is the pseudo-inverse of  $F$  ?

# Monte Carlo II

## Exercise 6 : Computing an expectation.

- ▶ Let us consider the lifespan of a transistor. We will say that this lifespan is a random variable  $T$  following an exponential law of parameter  $\frac{1}{3}$ . Let us assume (unrealistically) that the user could process  $T^2$  tasks using the machine.
- ▶ Please use the Monte Carlo method in order to approximate the expectation of this random variable.

## Deterministic vs stochastic ?

- ▶ So which method is better : deterministic or stochastic ?

# Algorithmic complexity

- ▶ The **complexity** of an algorithm is a measure of its **cost**. It is the number of elementary operations necessary for the algorithm to run.

## Complexity examples

- ▶ 1) What is the complexity of enumerating all the elements in a set of size  $n$  ?

## Complexity examples

- ▶ 2) What is the complexity searching a given name in a stack of **ranked**  $n$  folders ?

## Complexity examples

- ▶ 3) What is the complexity of enumerating all the permutations of a set of size  $n$  ?

# Complexities

- ▶ linear, polynomial complexities are OK
- ▶ exponential complexities are not OK



## Monte Carlo vs deterministic complexity

- ▶ Let  $n$  be the number of simulated variables for MC and the number of steps for the Riemann method.
- ▶ Let  $d$  be the **dimensionality** of the problem (we worked with dimension 1). If you work with **random vectors** the dimension might be  $> 1$ .

## Monte Carlo vs deterministic complexity

- ▶  $n \simeq$  computation cost
- ▶ Deterministic method : the precision is  $n^{-\frac{1}{d}}$ .
- ▶ Monte Carlo : the precision is  $n^{-\frac{1}{2}}$ .

## Monte Carlo vs deterministic complexity

- ▶  $n \simeq$  computation cost
- ▶ Deterministic method : the precision is  $n^{-\frac{1}{d}}$ .
- ▶ Monte Carlo : the precision is  $n^{-\frac{1}{2}}$ .
- ▶ Which method is better ?

## Monte Carlo vs deterministic

- ▶ Monte Carlo is better is the dimension is bigger than 3.
- ▶ Its precision does not depend on the dimensionality.
- ▶ Monte Carlo is mostly used in large dimensions when the precision required is smaller.
- ▶ The speed of convergence is in  $\frac{1}{\sqrt{n}}$  which is quite slow.

- └ Why is Monte-Carlo useful ?
- └ Notion of algorithmic complexity

# Speeding up Monte Carlo

- ▶ There are several methods to accelerate the convergence
- ▶ The most famous one is the **Variance reduction method**

# Speeding up Monte Carlo

- ▶ There are several methods to accelerate the convergence
- ▶ The most famous one is the **Variance reduction method**
- ▶ The idea is to use, instead of  $X$ , another random variable with the same expectation but with smaller variance.

$$E[Y] = E[X], \quad V(Y) \leq V(X) \quad (23)$$

## Monte Carlo and game

- ▶ How could we apply these ideas when building a fantom or an inspector ?



## Monte Carlo and game

- We could compute the probability of winning after being in some state  $s$  with a Monte Carlo approximation.





## Monte Carlo and game

- ▶ We could compute the probability of winning after being in some state  $s$  with a Monte Carlo approximation.
- ▶ Example of state to consider :  $(I, N)$  where :
  - ▶  $I$  is the number of isolated suspects.
  - ▶  $N$  is the number of non-isolated ones.



## Monte Carlo and game

- ▶ We could compute the probability of winning after being in some state  $s$  with a Monte Carlo approximation.
- ▶ Example of state to consider :  $(I, N)$  where :
  - ▶  $I$  is the number of isolated suspects.
  - ▶  $N$  is the number of non-isolated ones.
- ▶ But other choices are possible.

