Monte Carlo methods

July 8, 2019

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 july 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 july 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 ?

Ressources

- https://github.com/nlehir/summerschool contains our slides and exercices.
- When doing exercises, we will be using **python 3**

Overview

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 ?

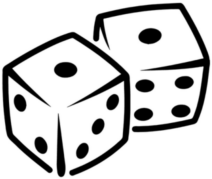


Figure: Dice

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 :

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

Example 2

- 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.
  - ...

Example 2

- 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.

Example 2

- 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 transimition to the last computer. (we have n transmisions between n computers)

Exercise 1

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

Exercise 2

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

Exercise 3

- 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, definied by a **density** $f(x)$.

Expected value of continous variables

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

Expected value of continous variables

- How can we express the expected value of a continous variable $X \in \mathbb{R}$ that has a density $f(X)$.
- $$E(X) = \int_{\mathbb{R}} x f(x) dx \quad (2)$$

Expected value of continuous variables

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

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

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

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

- 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.
- $$V(X) = E((X - E(X))^2) \tag{5}$$

Exercise 3

- Please show that :

$$V(X) = E(X^2) - E(X)^2 \tag{6}$$

Back to our problem

- Until now, we studies 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 studies random variables where we can either explicitely 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 :

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

Objective

- We want an approximation of this object :

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

Objective

- We want an approximation of this object :

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

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

Riemann sum

- Before presenting the Monte-Carlo method, let us discuss a more direct method to compute such an integral.

Riemann sum

- Let us consider a function f defined over an interval $[a, b]$. Assume that f is continuous.
- Then :

$$\frac{b-a}{n}\sum_{k=1}^{n} f(a+k\frac{b-a}{n}) \to \int_a^b f(x)dx \qquad (10)$$

Exercise

- Please use Riemann sum to give an approximation of the integral

$$\int_3^7 \cos^2(x)dx \tag{11}$$

Exercise

- Let us double check the result.

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 dsitributed. We assume that $E(|X_1|) < +\infty$. Then

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to+\infty]{a.s.} E(X) \tag{12}$$

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

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \rightarrow E[g(X)] \tag{13}$$

Remark

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

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X] \tag{14}$$

Method

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

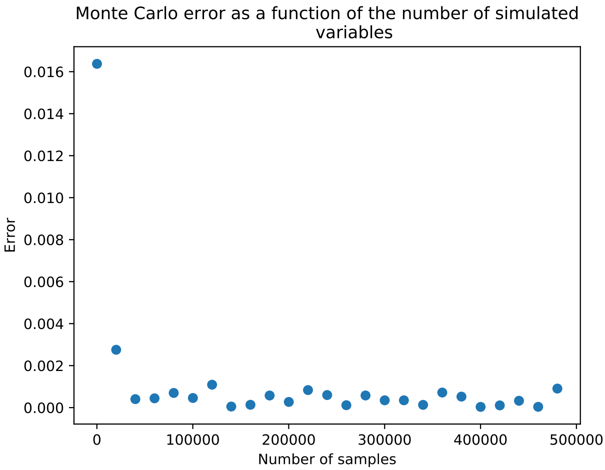
Exercise 4

- Let us compute the expectancy 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.

Exercise 5

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

Exercise 5


Monte Carlo error as a function of the number of simulated variables
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
Error
0
100000
200000
300000
400000
500000
Number of samples

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

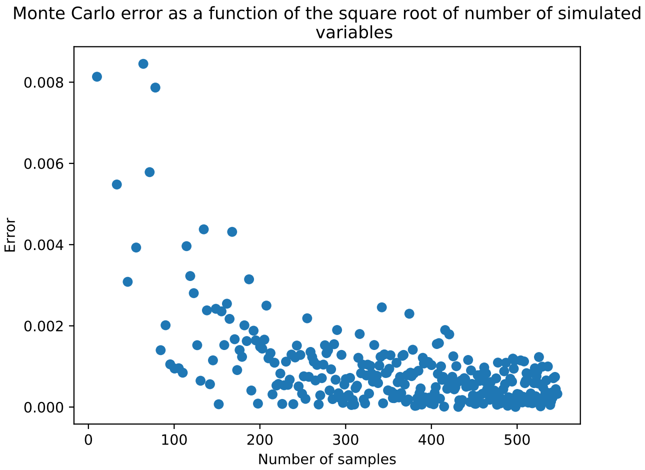
- This theorem tells us that with the same hypothesis as before **and** a new condition $E(X_1^2) < +\infty$:

$$\frac{\sqrt{n}}{\sigma}\left(\frac{X_1 + \cdots + X_n}{n} - E(X_1)\right) \underset{n \to +\infty}{\xrightarrow{distribution}} \mathcal{N}(0,1) \qquad (15)$$

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
Error
0.000
0.002
0.004
0.006
0.008
0
100
200
300
400
500
Number of samples

Exercise 6

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

$$\frac{\sqrt{n}}{\sigma}\epsilon_n \underset{n\to+\infty}{\overset{distribution}{\longrightarrow}} \mathcal{N}(0,1) \tag{16}$$

Exercise 6

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

$$\frac{\sqrt{n}}{\sigma}\epsilon_n \underset{n\to+\infty}{\overset{distribution}{\longrightarrow}} \mathcal{N}(0,1) \tag{17}$$

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

Remark

- The variance σ of the random variables appears in the estimator !

Simulation of non uniform random variables

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

Cumulative distribution function

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

$$F(x) = P(X \leq x) \tag{18}$$

Cumulative distribution function

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

$$F(x) = P(X \leq x) \tag{19}$$

- F is monotonically increasing

Pseudo inverse

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

$$\forall x \in [0,1], F^{-1}(u) = \inf\{y \in \mathbb{R}, F(y) \geq u\} \qquad (20)$$

Pseudo inverse

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

$$F^{-1}(u)| \leq x \Leftrightarrow u \leq F(x) \tag{21}$$

Pseudo inverse

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

$$F^{-1}(u)| \leq x \Leftrightarrow u \leq F(x) \tag{22}$$

- 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.

Exercise 7

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

$$f(x) = \lambda \exp -\lambda x \tag{23}$$

for $x \geq 0$ and 0 otherwise.

- Let us compute its cumulative distribution function.

Exercise 7

- Let us introduce the **exponential law**.
- Its density is
$$f(x) = \lambda \exp -\lambda x \tag{24}$$
for $x \geq 0$ and 0 otherwise.
- Let us compute its cumulative distribution function F.
- What is the pseudo-inverse of F ?

Exercise 8

- Let us consider the lifespan of a transistor. We will say that this lifespan is a random variable T following an exponenial 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 elementaty operations necessary for the algorithm to run.

Algorithmic complexity

- The **complexity** of an algorithm is a measure of its **cost**. It is the number of elementaty 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 by 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** this is what you will encounter.

Monte Carlo vs deterministic complexity

- Let n 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

- Let n 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.

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.

$$E[Y] = E[X], \; V(Y) \leq V(X) \tag{25}$$