

Computational Methods in Finance

Pedro R. S. Antunes

<mailto:prsantunes@tecnico.ulisboa.pt>

May 27, 2025

Some definitions from probability theory

In this section we recall some definitions and elementary results from the probability theory.

A random variable on a space \mathcal{S} is a (measurable) function $X : \mathcal{S} \rightarrow \mathbb{R}$. More informally, a random variable is a numerical quantity that is “random”, in the sense that its value depends on the outcome of a random experiment.

Typically we use capital letters X, Y, \dots to denote random variables and lower case letters x, y, \dots to denote the corresponding real values that are realizations of a random variable.

A random variable whose set of values is finite or countably infinite is said to be a discrete random variable, otherwise it is said to be a continuous random variable.

Example

Consider the problem of tossing a (fair) coin twice. In each of the coin tosses we can obtain a head (H) or a tail (T). Thus, the sample space for this random experiment is

$$\mathcal{S} = \{HH, HT, TH, TT\}.$$

Now let X be a random variable that records the number of tails obtained in an outcome. All the four events in the sample space \mathcal{S} will occur with the same probability (equal to $\frac{1}{4}$), so the probability mass function if given by

$$p_X(x) = \begin{cases} \frac{1}{4} & x = 0 \\ \frac{1}{2} & x = 1 \\ \frac{1}{4} & x = 2. \end{cases}$$

Let X be a discrete random variable. The probability mass function is a function p_X that gives the probability that X is equal to some value x , that is $p_X : \mathbb{R} \rightarrow [0, 1]$ defined by $p_X(x) = P(X = x)$.

We have

- $0 \leq p_X(x) \leq 1$.
- $\sum_i p_X(x_i) = 1$.
- we can define the cumulative distribution $F(x) = P(X \leq x)$.
- $F(x)$ is a non-decreasing function and $0 \leq F(x) \leq 1$.
- $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow +\infty} F(x) = 1$.

- The expected value is defined by

$$E(X) = \sum_i x_i p_X(x_i)$$

and more generally, given a function g , the expectation of $g(X)$ is

$$E(g(X)) = \sum_i g(x_i) p_X(x_i).$$

For instance, in the previous example of tossing a coin twice and counting the number of tails we have

$$E(X) = 0 \times p_X(0) + 1 \times p_X(1) + 2 \times p_X(2) = \frac{1}{2} + \frac{1}{2} = 1$$

- The expectation is a linear operator

$$E(\alpha X + \beta) = \alpha E(X) + \beta, \quad \forall \alpha, \beta \in \mathbb{R}, \text{ and}$$

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n),$$

for any random variables X_1, X_2, \dots, X_n .

- The variance of X is given by

$$\begin{aligned}Var(X) &= E[(X - E(X))^2] \\&= E(X^2) - (E(X))^2 \\&= \sum_i x_i^2 p_X(x_i) - \left(\sum_i x_i p_X(x_i) \right)^2.\end{aligned}$$

Moreover, we have

$$Var(X + a) = Var(X), \quad \forall a \in \mathbb{R}.$$

Examples of discrete distributions are

- **Bernoulli distribution** with parameter $0 < p < 1$ (we use the notation $X \sim Ber(p)$) that models an experiment with only two possible outcomes that can be encoded as 0 and 1 (“Bernoulli trial”). In this case

$$p_X(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0. \end{cases}$$

For instance the outcome from tossing a coin can be modeled by a random variable $X \sim Ber(\frac{1}{2})$.

In the case of a random variable $X \sim Ber(p)$, we have

$$E(X) = \sum_i x_i p_X(x_i) = 1 \times \underbrace{p_X(1)}_p + 0 \times p_X(0) = p.$$

The variance is given by

$$Var(X) = E(X^2) - (E(X))^2$$

and we have

$$E(X^2) = \sum_i x_i^2 p_X(x_i) = p.$$

Therefore,

$$Var(X) = p - p^2 = p(1 - p).$$

- **Binomial distribution** ($X \sim Bin(n, p)$) with probability mass function defined by

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

where

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

for which we can prove that $E(X) = np$ and
 $Var(X) = np(1 - p)$.

For instance, the distribution $Bin(n, \frac{1}{2})$ describes the number of tails (or heads) that we obtain in n coin tosses.

The joint probability mass function of two discrete random variables X and Y is defined as

$$p_{XY}(x, y) = P(X = x \text{ and } Y = y)$$

and the joint cumulative distribution function is defined by

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

We define the expectation of the product XY by

$$E(XY) = \sum_i \sum_j x_i y_j p_{XY}(x_i, y_j).$$

We say that two discrete random variables X and Y are independent if

$$p_{XY}(x, y) = p_X(x)p_Y(y), \quad \forall x, y$$

which implies that

$$E(XY) = E(X)E(Y), \quad (1)$$

because assuming independence of X and Y , we have

$$\begin{aligned} E(XY) &= \sum_i \sum_j x_i y_j p_{XY}(x_i, y_j) \\ &= \sum_i \sum_j x_i y_j p_X(x_i)p_Y(y_j) \\ &= \sum_i x_i p_X(x_i) \sum_j y_j p_Y(y_j) \\ &= E(X)E(Y). \end{aligned}$$

We also know that for any $a, b \in \mathbb{R}$,

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)$$

where

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Taking into account (1) we conclude that the covariance of two independent random variables is equal to zero, which implies that if X and Y are independent random variables, then

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y), \forall a, b \in \mathbb{R}. \quad (2)$$

Let X be a continuous random variable with density function f defined in the interval $[a, b]$ (eventually unbounded). Then, we have

- $f(x) \geq 0, \forall x \in [a, b].$
- $\int_a^b f(x)dx = 1.$
- we can define the cumulative distribution function
 $F(x) = \int_a^x f(t)dt.$
- $P(c \leq X \leq d) = F(d) - F(c) = \int_c^d f(x)dx$ and more generally, for a set A , we have $P(X \in A) = \int_A f(x)dx.$

- The expected value is defined by

$$E(X) = \int_a^b xf(x)dx$$

and more generally, given a function g , we have

$$E(g(X)) = \int_a^b g(x)f(x)dx.$$

- The expectation is a linear operator

$$E(\alpha X + \beta) = \alpha E(X) + \beta, \quad \forall \alpha, \beta \in \mathbb{R}, \text{ and}$$

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n),$$

for any random variables X_1, X_2, \dots, X_n .

- The variance of X is given by

$$\begin{aligned} Var(X) &= E[(X - E(X))^2] = E(X^2) - (E(X))^2 \\ &= \int_a^b x^2 f(x)dx - \left(\int_a^b xf(x)dx \right)^2 \end{aligned}$$

Some of the most common continuous distributions are

- **uniform distribution** over a bounded interval $[a, b]$ (and we use the notation $X \sim \mathcal{U}([a, b])$) for which the probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

In this case, we have

$$\begin{aligned} E(X) &= \int_a^b xf(x)dx \\ &= \int_a^b \frac{x}{b-a} dx \\ &= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_{x=a}^{x=b} \\ &= \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}. \end{aligned}$$

The variance is given by

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

and we have

$$\begin{aligned} E(X^2) &= \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_{x=a}^{x=b} \\ &= \frac{b^3 - a^3}{3(b-a)} \\ &= \frac{(b-a)(a^2 + ab + b^2)}{3(b-a)} = \frac{a^2 + ab + b^2}{3}. \end{aligned}$$

Thus,

$$\text{Var}(X) = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2} \right)^2 = \dots = \frac{(b-a)^2}{12}.$$

- **normal (Gaussian) distribution** ($X \sim \mathcal{N}(\mu, \sigma^2)$, with expected value μ and variance σ^2) with probability density function defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{2}\right\}, \quad x \in \mathbb{R}.$$

A particular case is the

- **standard normal (Gaussian) distribution** ($X \sim \mathcal{N}(0, 1)$) with probability density function defined by

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \quad x \in \mathbb{R}$$

for which we can prove that $E(X) = 0$ and $Var(X) = 1$. In this case we usually denote the cumulative distribution function by Φ , that is,

$$P(X \leq z) = \Phi(z).$$

Two random variables X and Y defined respectively on the intervals $I_X = [a_x, b_x]$ and $I_Y = [a_y, b_y]$ are said to be jointly continuous if there exists a nonnegative function

$f_{XY} : I_X \times I_Y \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, which is called the joint probability density function, such that for any $A \subset I_X \times I_Y$, we have

$$P((X, Y) \in A) = \int_A f_{XY}(x, y) dx dy.$$

Of course, we have

$$\int_{I_X} \int_{I_Y} f_{XY}(x, y) dx dy = 1.$$

We can define the marginal probability density functions,

$$f_X(x) = \int_{a_Y}^{b_Y} f_{XY}(x, y) dy, \quad \forall x \in I_X$$

and

$$f_Y(y) = \int_{a_X}^{b_X} f_{XY}(x, y) dx, \quad \forall y \in I_Y.$$

The joint cumulative distribution function of two random variables X and Y is defined as

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = \int_{a_X}^x \int_{a_Y}^y f_{XY}(x, y) dy dx.$$

We recall also the conditional probability formula, for any events A and B ,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ provided } P(B) > 0$$

and the conditional probability density function of X given Y ,

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

We say that X and Y are independent if

$$f_{X|Y}(x|y) = f_X(x),$$

that is, if

$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

Next, we recall the central limit theorem.

Theorem

(Central limit theorem) Let X_1, X_2, \dots be independent and identically distributed random variables with

$$\mu = E(X_i), \quad \sigma^2 = \text{Var}(X_i)$$

and define

$$S_n = \sum_{i=1}^n X_i.$$

Then, for large n , for all $z \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) = \Phi(z),$$

that is $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ follows approximately a standard normal distribution.

Theorem

Let X_1, X_2, \dots, X_n be independent random variables following respectively the distributions $\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2), \dots, \mathcal{N}(\mu_n, \sigma_n^2)$. Then, we have

$$\sum_{i=1}^n X_i \sim \mathcal{N} (\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2) .$$

Moreover,

Theorem

(Weak law of large numbers) Let X_1, X_2, \dots be independent and identically distributed random variables with

$$\mu = E(X_i), \quad \sigma^2 = \text{Var}(X_i)$$

and define

$$S_n = \sum_{i=1}^n X_i.$$

Then, for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \leq \epsilon\right) = 1.$$

The binomial method for pricing options

The calculation of the value of an option can be done through the solution of a terminal/boundary value problem involving Black-Scholes equation.

In this section we describe a different technique due to Cox, Ross and Rubinstein, known as binomial method. It allows to calculate the value of an option at current market price S_0 , $V(S_0, 0)$ without the need to calculate $V(S, t)$ for the entire domain to extract just the required value $V(S_0, 0)$, as typically happens for the finite difference method.

We start defining a discretization of the time interval $[0, T]$, by taking $M + 1$ equally spaced points $t_i = ih_t$, $i = 0, 1, \dots, M$, where $h_t = \frac{T}{M}$. Let's denote by S_i the market price at a certain instant of time t_i .

The binomial method relies on three assumptions:

- (A1) - The market price over a period of time h_t (that is for $t = t_{i+1}$) can only take two values, either it evolves up to S_u or down to S_d , where $0 < d < 1 < u$.
- (A2) - The probability of the price to evolve up is p .
- (A3) - The expectation and variance of S of this discrete model when $h_t \rightarrow 0$ shall coincide with the expected value/variance of the continuous model, evaluated for the risk-free interest rate r .

The assumptions (A1) and (A2) correspond to the framework of a binomial process, where we toss a (biased) coin where the outcome (say) “head” (up movement) occurs with probability p , as illustrated in Figure 1.

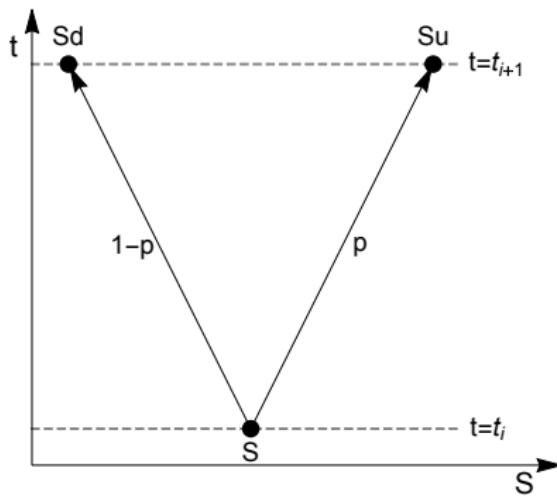


Figure: Binomial process.

They imply that the probability mass function of a random variable S_{i+1} defining the price at the instant of time $t = t_{i+1}$ is given by

$$p_{S_{i+1}}(x) = \begin{cases} p & x = S_i u \\ 1 - p & x = S_i d. \end{cases}$$

Note that the three parameters u , d and p are yet to be determined. They will be calculated in such a way that the discrete model is consistent with the continuous model, when $h_t \rightarrow 0$, which is assumption (A3). Next, we discuss three equations to be imposed in order to determine the three parameters.

It can be proven that in the continuous model we have

$$E(S_{i+1}) = S_i e^{rh_t} \quad (3)$$

and

$$E(S_{i+1}^2) = S_i^2 e^{(2r+\sigma^2)h_t} \quad (4)$$

from which we conclude that

$$\begin{aligned} Var(S_{i+1}) &= E(S_{i+1}^2) - (E(S_{i+1}))^2 \\ &= S_i^2 e^{(2r+\sigma^2)h_t} - S_i^2 e^{2rh_t} \\ &= S_i^2 e^{2rh_t} (e^{\sigma^2 h_t} - 1). \end{aligned} \quad (5)$$

The expectation of the discrete model is given by

$$\begin{aligned} E(S_{i+1}) &= \sum_i x_i p_{S_{i+1}}(x_i) \\ &= S_i u \ p_{S_{i+1}}(S_i u) + S_i d \ p_{S_{i+1}}(S_i d) \\ &= S_i u \ p + S_i d \ (1 - p). \end{aligned} \tag{6}$$

Equating (3) and (6), we get

$$\begin{aligned} S_i u \ p + S_i d \ (1 - p) &= S_i e^{rh_t} \Rightarrow \\ u \ p + d \ (1 - p) &= e^{rh_t} \end{aligned} \tag{7}$$

which implies that

$$p = \frac{e^{rh_t} - d}{u - d} \tag{8}$$

and since we must have $0 \leq p \leq 1$, we shall have $d \leq e^{rh_t} \leq u$, which relates upward, downward movements of the asset price and the interest rate r .

Also in the discrete model we have

$$\begin{aligned} E(S_{i+1}^2) &= \sum_i x_i^2 p_{S_{i+1}}(x_i) \\ &= (S_i u)^2 p_{S_{i+1}}(S_i u) + (S_i d)^2 p_{S_{i+1}}(S_i d) \\ &= (S_i u)^2 p + (S_i d)^2 (1 - p) \end{aligned} \tag{9}$$

from which we conclude that in the discrete model

$$\begin{aligned} \text{Var}(S_{i+1}) &= E(S_{i+1}^2) - (E(S_{i+1}))^2 \\ &= (S_i u)^2 p + (S_i d)^2 (1 - p) - [S_i u p + S_i d (1 - p)]^2 \end{aligned} \tag{10}$$

and equating (5) and (10) we get

$$\begin{aligned}(S_i u)^2 p + (S_i d)^2 (1 - p) - [S_i u p + S_i d (1 - p)]^2 &= S_i^2 e^{2rh_t} \left(e^{\sigma^2 h_t} - 1 \right) \Rightarrow \\ u^2 p + d^2 (1 - p) - \underbrace{[u p + d (1 - p)]^2}_{=e^{2rh_t}, \text{ by (7)}} &= e^{2rh_t} \left(e^{\sigma^2 h_t} - 1 \right) \Rightarrow \\ u^2 p + d^2 (1 - p) - e^{2rh_t} &= e^{2rh_t} e^{\sigma^2 h_t} - e^{2rh_t} \Rightarrow \\ u^2 p + d^2 (1 - p) &= e^{2rh_t} e^{\sigma^2 h_t}. \end{aligned}\tag{11}$$

Therefore, we obtained two equations ((8) and (11)) to determine u , d and p in terms of the parameters r , σ and h_t which are known. Therefore, we can impose an additional equation. One usual assumption is that $u d = 1$, which implies that we must solve the system

$$\begin{cases} p = \frac{e^{rh_t} - d}{u - d} \\ u^2 p + d^2 (1 - p) = e^{2rh_t} e^{\sigma^2 h_t} \\ u d = 1. \end{cases} \quad (12)$$

Using the first two equations, we get

$$\begin{aligned} e^{2rh_t} e^{\sigma^2 h_t} &= u^2 \underbrace{\frac{e^{rh_t} - d}{u - d}}_p + d^2 \left(1 - \underbrace{\frac{e^{rh_t} - d}{u - d}}_p \right) \\ &= \frac{e^{rh_t} u^2 - u^2 d}{u - d} + d^2 \frac{u - e^{rh_t} + d}{u - d} \\ &= \frac{e^{rh_t} u^2 - u^2 d + u d^2 - e^{rh_t} d^2}{u - d} \\ &= \frac{e^{rh_t} (u^2 - d^2) - u^2 d + u d^2}{u - d} \\ &\stackrel{(ud=1)}{=} \frac{e^{rh_t} (u - d)(u + d) - (u - d)}{u - d} \\ &= e^{rh_t} (u + d) - 1. \end{aligned}$$

Therefore, we get

$$e^{rh_t} e^{\sigma^2 h_t} = u + \frac{1}{u} - \frac{1}{e^{rh_t}} \Rightarrow u^2 - u \left(\frac{1}{e^{rh_t}} + e^{rh_t} e^{\sigma^2 h_t} \right) + 1 = 0 \quad (13)$$

and defining

$$\beta := \frac{1}{2} \left(\frac{1}{e^{rh_t}} + e^{rh_t} e^{\sigma^2 h_t} \right),$$

which is a non-negative parameter, we can write (13) as

$$u^2 - 2\beta u + 1 = 0 \Rightarrow u = \frac{2\beta \pm \sqrt{4\beta^2 - 4}}{2} = \beta \pm \sqrt{\beta^2 - 1}.$$

Note that the product of these two solutions leads to

$$(\beta + \sqrt{\beta^2 - 1})(\beta - \sqrt{\beta^2 - 1}) = \beta^2 - (\sqrt{\beta^2 - 1})^2 = 1.$$

Thus, since we have $u > d$ we define

$$u = \beta + \sqrt{\beta^2 - 1} \text{ and } d = \beta - \sqrt{\beta^2 - 1}$$

which implies that we have the following parameters

$$\begin{aligned}\beta &:= \frac{1}{2} \left(\frac{1}{e^{rh_t}} + e^{rh_t} e^{\sigma^2 h_t} \right) \\ u &= \beta + \sqrt{\beta^2 - 1} \\ d &= \beta - \sqrt{\beta^2 - 1} \\ p &= \frac{e^{rh_t} - d}{u - d}. \end{aligned} \tag{14}$$

Now, we can construct a scheme to build a tree that will allow to determine $V(S_0, 0)$.

We start at $t = t_0$ and assume that the market price of the asset is $S = S_0$. For now, let's assume that we have just one time level ($M=1$), that is, $t_1 = T$. The value of the option at initial instant of time, $V(S_0, 0)$ has just two possibilities: either it moves upward to a value, say $V^{(u)}$ or downward to the value $V^{(d)}$.

It can be proven that we can write

$$V_0 = e^{-rh_t} \left(p V^{(u)} + (1 - p) V^{(d)} \right). \quad (15)$$

We can generalize this analysis to a tree with several time levels.

We assume that for $t = t_0$ the market price of the asset is $S = S_0$. Then, for $t = t_1$, we have two possibilities: either the price moves upward to a value S_0u or downward to a value S_0d . At next level corresponding to $t = t_2$, again, each of the values from the previous level can move upward or downward.

Since we have $S_0du = S_0ud$, at level $t = t_2$ the asset price can only take three values (not four) and the same idea applies to next time levels.

For the instant of time $t = t_i$, we have $i + 1$ possible values for the asset price, $S_{j,i} := S_0 u^j d^{i-j}$, $j = 0, 1, \dots, i$, as illustrated in Figure 2.

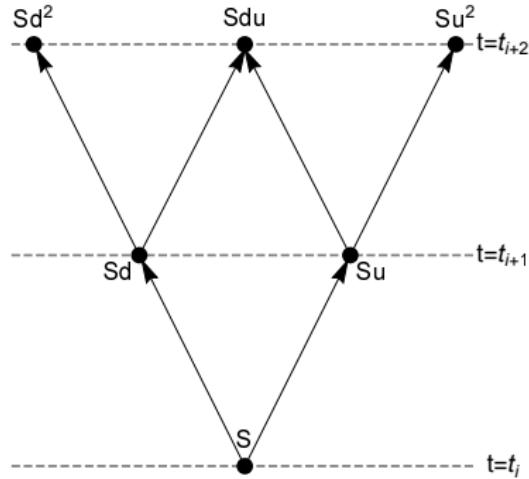


Figure: Construction of the tree of the binomial method.

We apply the same construction up to the maturity $t = t_M = T$, obtaining a tree like the one presented in Figure 3 which was obtained for $S_0 = 5$, $T = 1$ and $M = 20$.

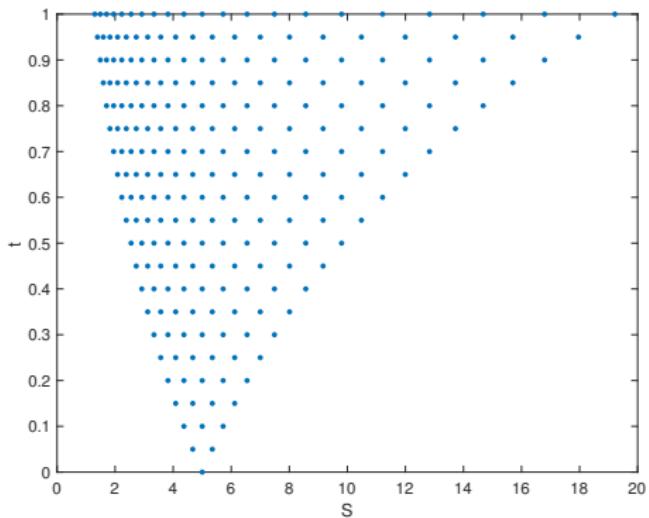


Figure: The tree of the binomial method obtained for $S_0 = 5$, $T = 1$ and $M = 20$.

Now, we need to calculate the option values $V_{j,i} = V(t_i, S_{j,i})$. For $t = t_M = T$, these values can be immediately calculated from the payoff,

- European call:

$$V_{j,M} = V(t_M, S_{j,M}) = V(T, S_{j,M}) = \max\{S_{j,M} - K, 0\}$$

- European put:

$$V_{j,M} = V(t_M, S_{j,M}) = V(T, S_{j,M}) = \max\{K - S_{j,M}, 0\}.$$

Then, we move backward in time and using (15) we can write that

$$V_{j,i} = e^{-rh_t} (pV_{j+1,i+1} + (1 - p)V_{j,i+1}), \quad j = 0, 1, \dots, i \quad (16)$$

which means that each value $V_{j,i}$ at time level $t = t_j$ is calculated by using two values of the option at time level $t = t_{j+1}$.

We apply this iterative procedure until reaching the time level $t = t_0$ and obtaining the value $V(S_0, 0)$, the fair value of the option at the current time.

The binomial method for pricing an European option can be defined by the following algorithm

Given r, σ, S_0, T, K and the type of option (put or call)

Calculate $h_t = \frac{T}{M}$ and u, d, p defined by (14)

Define $S_{0,0} = S_0$ and $S_{j,M} = S_{0,0} u^j d^{M-j}, j = 0, \dots, M$

Set $V_{j,M}$ from the payoff

$\left\{ \begin{array}{l} \text{For } i = M - 1, M - 2, \dots, 0 \text{ calculate} \\ V_{j,i} = e^{-rh_t} (pV_{j+1,i+1} + (1 - p)V_{j,i+1}) . \end{array} \right.$

Output $V_{0,0}$ which is the approximation for $V(S_0, 0)$.

Figure 4 illustrates the option pricing of a European put option with the parameters $r = 0.06$, $\sigma = 0.3$, $T = 1$, $K = 10$ and $M = 20$.

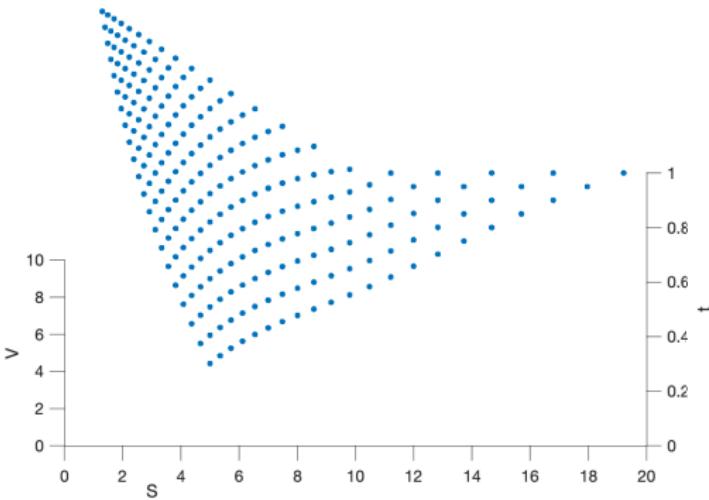


Figure: The tree of the binomial method obtained for a European put option with the parameters $r = 0.06$, $\sigma = 0.3$, $T = 1$, $K = 10$ and $M = 20$ and the corresponding value $V(S_0, 0)$.

We obtained the approximation for the value of the option at current instant of time $V(S_0, 0) \approx V_{0,0} \approx 4.4281$. Note that corresponding value obtained from the solution of Black-Scholes equation is 4.4304648. Table 44 shows the numerical approximations obtained by the binomial method for the value $V(S_0, 0)$ of a European put option with the parameters $r = 0.06$, $\sigma = 0.3$, $T = 1$, $K = 10$ and different values of M .

M	$V_{0,0}$
10	4.4284828
20	4.4281482
50	4.4296070
100	4.4302752
200	4.4303387
1000	4.4304459

The binomial method can also be applied for pricing an American option. In that case, we shall take into account that the option can be exercised at any time. At each instant of time t_j , the holder optimizes his/her position and decides if it preferable to hold or to exercise the option. So, the holder takes the maximum

$$\max \{ \text{payoff } (S_{j,i}), V_{j,i}^{\text{cont}} \},$$

where $V_{j,i}^{\text{cont}}$ is the “continuation value” defined by (16).

Therefore, we have

- American call:

$$V_{j,i} = \max \{ (S_{j,i} - K)^+, e^{-r h_t} (p V_{j+1,i+1} + (1-p) V_{j,i+1}) \}$$

- American put:

$$V_{j,i} = \max \{ (K - S_{j,i})^+, e^{-r h_t} (p V_{j+1,i+1} + (1-p) V_{j,i+1}) \}.$$

and we obtain the following algorithm

Given r, σ, S_0, T, K and the type of option (put or call)

Calculate $h_t = \frac{T}{M}$ and u, d, p defined by (14)

Define $S_{0,0} = S_0$ and $S_{j,i} = S_{0,0} u^j d^{i-j}$, $i = 1, 2, \dots, M-1$, $j = 0, \dots, i$

Set $V_{j,M}$ from the payoff

For $i = M-1, M-2, \dots, 0$ calculate

$$\begin{cases} V_{j,i} = \max \left\{ (S_{j,i} - K)^+, e^{-rh_t} (pV_{j+1,i+1} + (1-p)V_{j,i+1}) \right\} & (\text{call}) \text{ or} \\ V_{j,i} = \max \left\{ (K - S_{j,i})^+, e^{-rh_t} (pV_{j+1,i+1} + (1-p)V_{j,i+1}) \right\} & (\text{put}). \end{cases}$$

Output $V_{0,0}$ which is the approximation for $V(S_0, 0)$.

Figure 47 shows the tree of the binomial method obtained for an American put option with the parameters $r = 0.1$, $\sigma = 0.3$, $S_0 = 8$, $T = 1$, $K = 10$ and $M = 50$ and the corresponding value $V(S_0, 0)$.

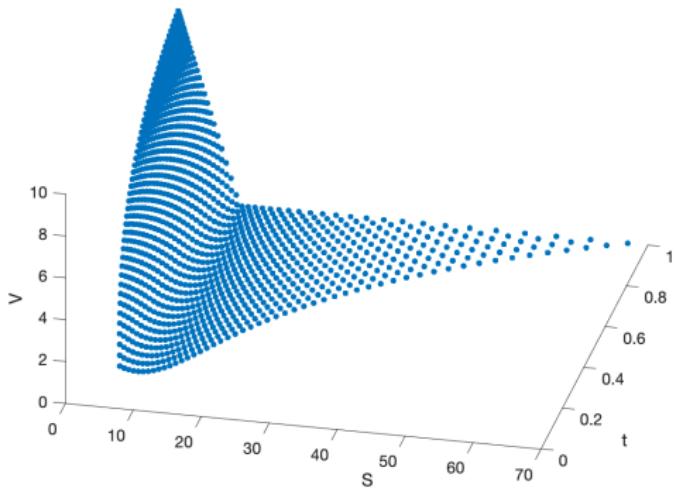


Table 48 shows the approximations obtained by the binomial method for the value $V(S_0, 0)$ of an American put option with the parameters $r = 0.1$, $\sigma = 0.3$, $S_0 = 8$, $T = 1$, $K = 10$ and different values of M .

M	$V_{0,0}$
10	2.0175022
20	2.0264991
50	2.0279317
100	2.0262876
200	2.0269894
1000	2.0269173

- ① Implement the binomial method for pricing European and American options.
- ② Test it with the parameters considered in the examples presented in the notes/slides of the course.

We will see in next chapters that, for instance in the context of the numerical solutions of stochastic differential equations we need to consider methods for generating pseudo-random numbers. We say “pseudo-random” numbers (and not “random numbers”) because the computational generation of the numbers is defined by a deterministic and predictable way. However, it is convenient that they mimic the properties of true random numbers as much as possible.

Computational simulation of random variables

We will assume that x_i , $i = 1, \dots, N$ are independent and identically distributed random variables following a general distribution with mean μ and variance σ^2 . We will estimate the mean and variance based on this sampling. The estimator for the mean μ is

$$\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (17)$$

Exercise

Verify that

$$E(\hat{\mu}) = \mu \quad (18)$$

and that

$$Var(\hat{\mu}) = \frac{\sigma^2}{N}. \quad (19)$$

Note that (18) establishes that the expected value of the estimator $\hat{\mu}$ coincides with μ , the parameter that we would like to estimate and we say that the estimator is unbiased. In general, an estimator with the property that the expectation of the estimator is equal to the parameter that we are estimating is said to be unbiased, otherwise we say that it is biased.

The estimator for the variance is the sample variance

$$\hat{s}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2. \quad (20)$$

$$\hat{s}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Exercise

Verify that the estimator \hat{s}^2 is an unbiased estimator of σ^2 while the estimator

$$\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

is a biased estimator of σ^2 .

The linear congruential generator

In this section we introduce some computational methods for generating pseudo-random numbers.

We start presenting an algorithm for the computational simulation of random variables with uniform distribution known as linear congruential generator. It allows to calculate samples of numbers in the interval $[0, 1]$ that simulate realizations of random variables with uniform distribution over the interval $[0, 1]$.

Recall that for a random variable $X \sim \mathcal{U}([0, 1])$, the density function is $f \equiv 1$ and the distribution function is

$$F(x) = P(X \leq x) = \int_0^x f(t)dt = x, \quad x \in [0, 1]. \quad (21)$$

The linear congruential generator

The computational simulation of X shall generate a sequence of numbers $x_1, \dots, x_n \in [0, 1]$ such that for any a and b satisfying $0 \leq a < b < 1$ we have

$$\lim_{n \rightarrow +\infty} \frac{\#\{x_j : x_j \in [a, b]\}}{n} = \int_0^1 \chi_{[a,b]}(t) dt = P(a \leq X \leq b),$$

where

$$\chi_{[a,b]}(t) = \begin{cases} 1, & \text{if } t \in [a, b] \\ 0, & \text{otherwise.} \end{cases}$$

The linear congruential generator

We know that given $n \in \mathbb{Z}$ and $M \in \mathbb{N}$ there exist unique $q, r \in \mathbb{Z}$ such that

$$n = qM + r, \quad \text{with} \quad 0 \leq r < M$$

and we write

$$n = r \bmod M.$$

We say that r is the residue modulo M of n .

Example

We have $32 = 6 \times 5 + 2$. Since $2 < 5$, we write $32 = 2 \bmod 5$.

The linear congruential generator

The linear congruential generator is defined by

Given integers M, a and b

$\left\{ \begin{array}{l} \text{Choose } m_1 \in [0, 1] \\ \text{For } i = 2, 3, \dots \text{ calculate} \\ m_i = (am_{i-1} + b) \bmod M. \end{array} \right.$

Define $x_i = \frac{m_i}{M}, i = 2, 3, \dots$

The number m_1 is called the seed of the generator.

The linear congruential generator

Note that

- $m_i \in \{0, 1, \dots, M - 1\}$
- The sequence of numbers $(m_i)_i$ is periodic with period $\leq M$ because all the m_i belong to $\{0, 1, \dots, M - 1\}$, therefore, there are at most M different numbers m_i . Thus, at least two numbers of the set $\{m_1, \dots, m_{M+1}\}$ must be equal and we have $m_i = m_{i+p}$, for a $p \leq M$.

Example

Consider the sequence defined by

$$m_i = (3m_{i-1} + 1) \bmod 5, \quad i = 2, 3, \dots$$

Considering the seed $m_1 = 0$, calculate m_i , $i=2,\dots,5$

The linear congruential generator

The integers M , a and b shall be chosen in a convenient way. For instance, if we take $b = 0$ and $m_1 = 0$, we would have

$$m_2 = (am_1) \bmod M = 0,$$

therefore we would have $m_i = 0$, $i = 1, 2, \dots$. Some criteria for the choices of M , a and b have been proposed but a common choice is $M = 2^{31} - 1$, $a = 16807$ and $b = 0$.

The linear congruential generator

Note that the sequence of pseudo-random numbers shall mimic the properties of true random numbers as much as possible. For instance, for the distribution $\mathcal{U}([0, 1])$ we know that the mean is $\mu = \frac{1}{2}$ and the variance is $\sigma^2 = \frac{1}{12}$ and we can use the sample mean

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

as an estimator for the mean μ and the sample variance

$$\hat{s}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

as an estimator for the variance in order to check if the samples produced by the linear congruential generator are close from the mean and variance of the uniform distribution.

The linear congruential generator

Figure 5 shows distribution of 1000 pseudo-random numbers obtained with the linear congruential generator, taking $m_1 = 0.1$

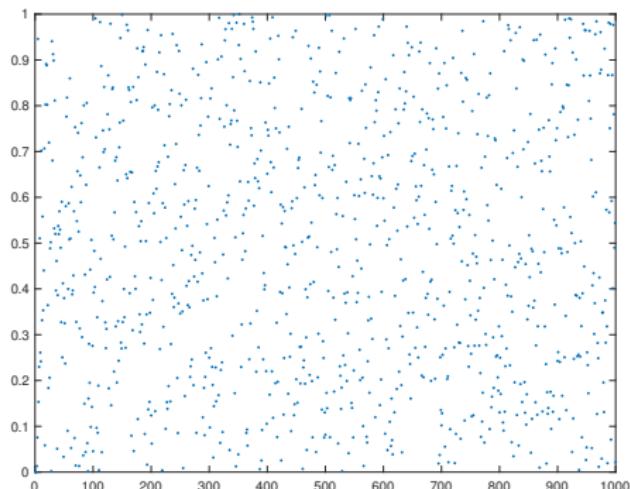


Figure: Distribution of 1000 pseudo-random numbers obtained with the linear congruential generator, taking $m_1 = 0.1$.

The linear congruential generator

Table 1 shows the mean and variance estimators from samples of N numbers obtained with the linear congruential generator, taking $m_1 = 0.1$. Note that for the distribution $\mathcal{U}([0, 1])$ we have $\mu = 0.5$ and $\sigma^2 = \frac{1}{12} \approx 0.083333$.

N	$\hat{\mu}$	\hat{s}^2
1000	0.487499	0.081543
10000	0.496996	0.083830
1000000	0.500252	0.083244

Table: Mean and variance estimators from samples of N numbers obtained with the linear congruential generator, taking $m_1 = 0.1$. For the distribution $\mathcal{U}([0, 1])$ we have $\mu = 0.5$ and $\sigma^2 = \frac{1}{12} \approx 0.083333$.

The linear congruential generator

Figure 6 shows the histogram of the pseudo-random numbers obtained with the linear congruential generator, taking $m_1 = 0.1$ obtained for $N = 1000$ (left plot) and $N = 100000$ (right plot), where we divided the interval $[0, 1]$ into 10 sub-intervals of the same length.

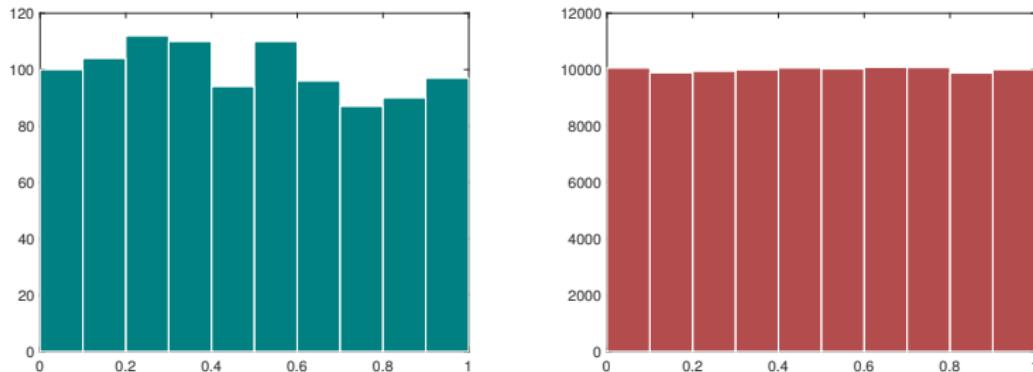


Figure: Histogram of the pseudo-random numbers obtained with the linear congruential generator, taking $m_1 = 0.1$ obtained for $N = 1000$ (left plot) and $N = 100000$ (right plot), where we divided the interval $[0, 1]$ in 10 sub-intervals of the same length.

Inversion method

In this section, we consider the inversion method which allows to generate pseudo-random numbers of an arbitrary distribution from the inversion of the distribution function. Let X be a continuous random variable over the interval $[a, b]$ (the analysis is valid also for unbounded intervals) and f be the density function. The distribution function is given by

$$F(x) = \int_a^x f(t)dt, \quad x \in [a, b].$$

We know that F is a continuous function and that $F(a) = 0$, $F(b) = 1$. Therefore, $F([a, b]) = \{F(x) : x \in [a, b]\} = [0, 1]$ that follows from the continuity of F and from the fact that $F \leq 1$. Moreover, for $a \leq x < y \leq b$ we have

$$F(y) - F(x) = \int_a^y f(t)dt - \int_a^x f(t)dt = \int_x^y f(t)dt \geq 0$$

because f is nonnegative since it is a density function.

Therefore, F is monotonically increasing function and we can define

$$F^{-1}(u) = \inf \{x : F(x) \geq u\}. \quad (22)$$

If the function F is strictly monotonically increasing, for all $y_0 \in [0, 1]$ there is a unique $x_0 \in [a, b]$ such that $F(x_0) = y_0$. Thus, $x_0 = F^{-1}(y_0)$ as illustrated in the following Figure.

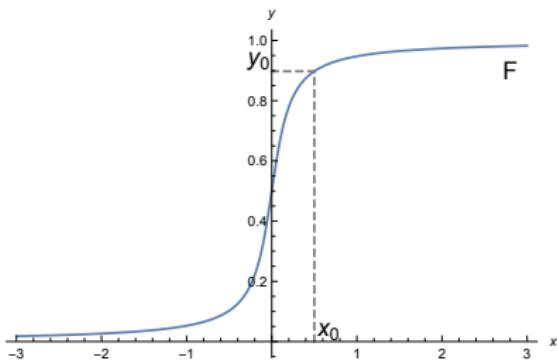


Figure: Illustration of the definition of F^{-1} .

However, note that F^{-1} defined in (22) is more general than the inverse of a(n invertible) function and it is defined also for noninvertible functions F , as illustrated in the following Figure.

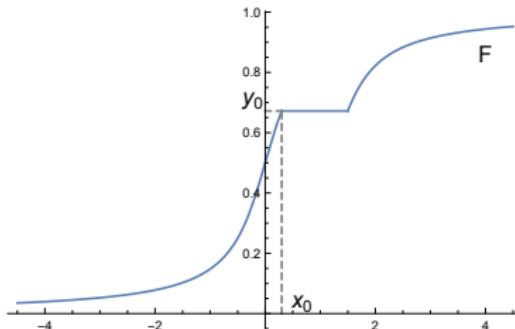


Figure: Illustration of the definition of F^{-1} . Note that F^{-1} is also defined for noninvertible functions F .

In this case, for a given $y_0 \in [0, 1]$ there are distinct values x_0 and z_0 for which

$$F(x_0) = F(z_0) = y_0.$$

In this case, we define $F^{-1}(y_0)$ to be the infimum among all the values x for which $F(x) = y_0$.

Now we prove a theorem about properties of the function F^{-1}

Theorem

The function F^{-1} defined in (22) satisfies the following properties:

- ① F^{-1} is monotonically increasing,
- ② $F^{-1}(F(x)) \leq x$,
- ③ $F(F^{-1}(u)) \geq u$,
- ④ $F^{-1}(u) \leq x \Leftrightarrow u \leq F(x)$.

Using (iv) of previous theorem we know that, for any random variable U , we have

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)).$$

Now, assume that $U \sim U([0, 1])$. By (21), we know that the distribution function of U verifies

$$P(U \leq u) = u.$$

Thus,

$$P(U \leq F(x)) = F(x)$$

which implies (from (iv) of the previous theorem) that

$$P(F^{-1}(U) \leq x) = F(x) \quad (= P(X \leq x)),$$

that is,

$$X \sim F^{-1}(U).$$

In previous section we have already presented an effective algorithm for generating pseudo-random numbers simulating a uniform distribution over the interval $[0, 1]$.

Thus, we will be able to simulate a random variable with an arbitrary distribution function, provided for that distribution we can calculate F^{-1} .

This suggests the following algorithm for simulating a random variable X with distribution function F

Given u_1, u_2, \dots, u_n independent realizations from $U([0, 1])$,

Calculate $x_i = F^{-1}(u_i)$, which are realizations of the random variable X .

Inversion method - example

Next, we illustrate the application of the inversion algorithm.

Realizations of a uniform distribution over $[a, b]$

If $X \sim \mathcal{U}([a, b])$, we have

$$P(X \leq x) = F(x) = \frac{1}{b-a} \int_a^x 1 dt = \frac{x-a}{b-a}, \quad x \in [a, b]$$

and immediately we get

$$y = F(x) = \frac{x-a}{b-a} \Rightarrow F^{-1}(y) = (b-a)y + a.$$

Therefore, if $U \sim \mathcal{U}([0, 1])$ then,

$$X = (b-a)U + a \sim \mathcal{U}([a, b]).$$

Exponential distribution

For a random variable X following an exponential distribution with parameter θ ($X \sim \text{Exp}(\theta)$) the density function is

$$f(x) = \theta e^{-\theta x}, \quad x \in [0, +\infty[.$$

Therefore, the distribution function is

$$F(x) = \int_0^x f(t)dt = \left(1 - e^{-\theta x}\right), \quad x \geq 0.$$

Thus, F is invertible and we get

$$y = F(x) = \left(1 - e^{-\theta x}\right) \Rightarrow F^{-1}(y) = -\frac{1}{\theta} \log(1 - y)$$

Therefore, we can simulate the random variable X using the random variable $-\frac{1}{\theta} \log(1 - U)$, where $U \sim \mathcal{U}([0, 1])$.

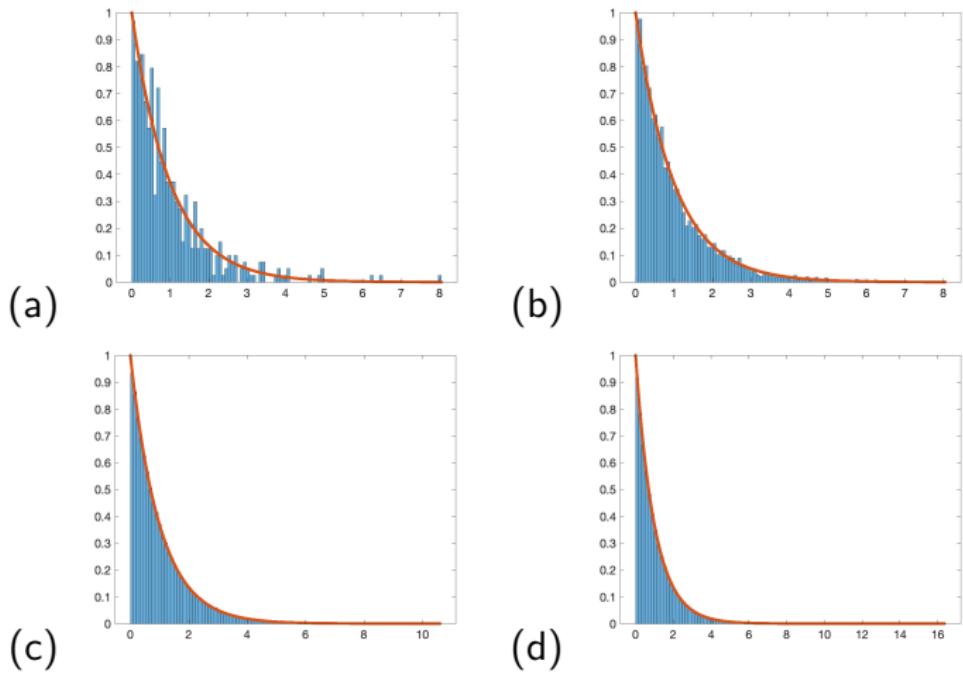


Figure: Histogram (normalized to represent a density function) of samples of N pseudo-random numbers obtained for a random variable following an exponential distribution with parameter $\theta = 1$. We considered (a) $N=500$, (b) $N=5000$, (c) $N=50000$ and $N=500000$. We show also (red line) the plot of the density function $f(x) = e^{-x}$.

Rayleigh distribution

The density function of a random variable X following a Rayleigh distribution with parameter θ is

$$f(x) = \frac{1}{\theta^2} x e^{-\frac{x^2}{2\theta^2}}, \quad x \in [0, +\infty[.$$

Thus, the distribution function is

$$F(x) = \int_0^x f(t) dt = 1 - e^{-\frac{x^2}{2\theta^2}}, \quad x \geq 0$$

and we get

$$y = F(x) = 1 - e^{-\frac{x^2}{2\theta^2}} \Rightarrow F^{-1}(y) = \theta \sqrt{-2 \log(1 - y)}.$$

Therefore,

$$X \sim \theta \sqrt{-2 \log(1 - U)}, \quad \text{where } U \sim \mathcal{U}([0, 1]),$$

or simply

$$X \sim \theta \sqrt{-2 \log(U)}, \quad \text{where } U \sim \mathcal{U}([0, 1]).$$

- ① Implement the linear congruential generator.
- ② Implement the inversion method for random variables $X \sim Exp(\theta)$ and $Y \sim Ray(\theta)$.
- ③ Build histograms together with the plot of the density functions for the distributions considered in (2).
- ④ Estimate the mean and variance in both cases and compare with the theoretical mean and variance of each one the distributions.

$$E(X) = \theta, \quad Var(X) = \theta^2$$

$$E(Y) = \sqrt{\frac{\pi}{2}}\theta, \quad Var(Y) = \frac{(4 - \pi)}{2}\theta^2.$$

Acceptance-rejection method

In the previous section we studied the inversion method for generating pseudo-random numbers. However, that method requires the calculation of F^{-1} and this task is difficult for some distributions. For instance, the distribution function of a random variable $X \sim \mathcal{N}(0, 1)$ is defined by

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

but we have not a closed form for calculating this integral and the calculation of F^{-1} shall be done numerically, for example using Newton's method or any other numerical method for solving nonlinear equations.

Next, we describe a new method to generate pseudo-random numbers obtained from a random variable following a general distribution. We will denote by \mathcal{R}_h the region between the graph of a general density function h and the x -axis.

We assume the random variables $U \sim \mathcal{U}([0, 1])$ and Y , independent of each other. Y has density function g defined over the interval $[a, b]$. We assume also another random variable X , with density function f defined over the interval $[a, b]$.

We assume that we are able to simulate the random variable Y and want to develop a numerical technique to simulate the random variable X . The idea of the method is that the points defined by

$$(x, y) = (Y, U.h(Y)),$$

for a general distribution function h , are uniformly distributed on \mathcal{R}_h , as illustrated in next Figure.

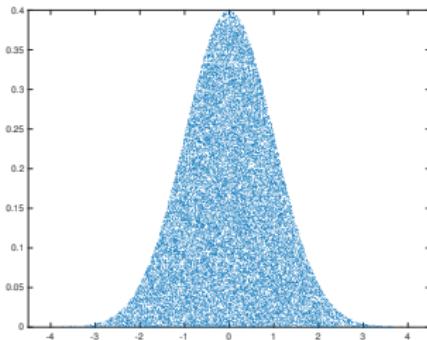


Figure: Plot of 30 000 points $(x, y) = (Y, U.h(Y))$, where $Y \sim \mathcal{N}(0, 1)$.

Moreover, if we take just a sub-region of \mathcal{R}_h , the points are distributed uniformly too.

Assume that there is a constant $M > 0$ for which

$$f(x) \leq Mg(x), \quad \forall x \in [a, b], \quad (23)$$

as illustrated in next Figure.

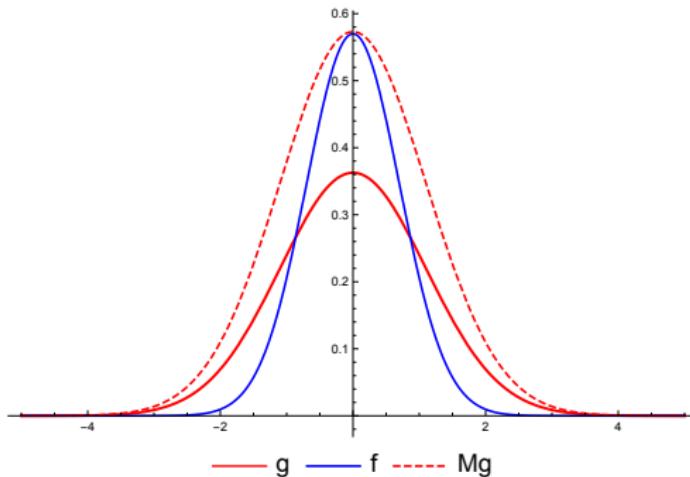


Figure: Plots of density distributions f and g and plot of Mg (dashed line).

Note that the property (23) actually implies that $M \geq 1$, because

$$f(x) \leq Mg(x) \Rightarrow \underbrace{\int_a^b f(x)dx}_{=1} \leq M \underbrace{\int_a^b g(x)dx}_{=1} \Rightarrow M \geq 1,$$

because f and g are density functions.

The idea is that, since we have (23), then $\mathcal{R}_f \subset \mathcal{R}_{Mg}$. We assumed that we are able to simulate the random variable Y with density function g - and so can simulate the random variable MY with density function Mg - and want to develop a numerical technique to simulate the random variable X with density function f , by keeping just the points (obtained from the simulation of Mg) that belong to \mathcal{R}_f , that is, we consider the following algorithm

Repeat
generate y from a random variable with density function g
generate u from $U \sim \mathcal{U}([0, 1])$
until $uMg(y) \leq f(y)$.

Return $x=y$

Example

We will simulate a random variable $X \sim \mathcal{N}(0, 1)$ by using the acceptance-rejection method, comparing with a random variable Y with probability density function $g(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ that we are able to simulate. We note that

$$\frac{f(x)}{g(x)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{\frac{1}{\pi} \frac{1}{1+x^2}} = \sqrt{\frac{\pi}{2}} (1 + x^2) e^{-\frac{x^2}{2}}$$

and the derivative of this function is given by

$$\sqrt{\frac{\pi}{2}} x (1 - x^2) e^{-\frac{x^2}{2}}$$

that vanishes for $x \in \{-1, 0, 1\}$.

Example

It is straightforward to verify that

$$M := \frac{f(1)}{g(1)} = \sqrt{\frac{2\pi}{e}}$$

is the absolute maximum of $\frac{f(x)}{g(x)}$. Now, we remark that we can simulate Y by using the inversion method. Indeed, we have

$$G(y) = \int_{-\infty}^y g(t)dt = \frac{1}{\pi} \int_{-\infty}^y \frac{1}{1+t^2} dt = \frac{1}{\pi} \arctan(y) + \frac{1}{2}.$$

Therefore,

$$y = G^{-1}(u) = \tan\left(u\pi - \frac{\pi}{2}\right).$$

Example

Therefore, we can simulate the random variable Y by taking pseudo-random points defined by $y = \tan(u_0\pi - \frac{\pi}{2})$, where u_0 are realizations of a $\mathcal{U}([0, 1])$ distribution.

Now we define $u_1 \sim \mathcal{U}([0, 1])$ (independent from u_0) and accept just the points that satisfy $u_1 Mg(y) \leq f(y)$. Next Figure shows plots of f (red line), Mg (blue line) and the pseudo-random points obtained for the simulation of X using the acceptance-rejection method (in red) and the points of the simulation of $(Y, UMg(Y))$ for which the acceptance condition is not satisfied (in blue).

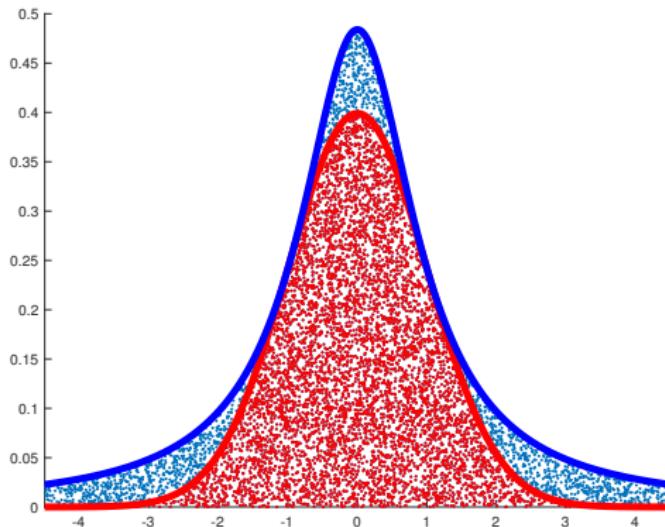


Figure: Plots of f (red line), Mg (blue line) and the pseudo-random points obtained for the simulation of X using the acceptance-rejection method (in red) and the points of the simulation of $(Y, UMg(Y))$ for which the acceptance condition is not satisfied (in blue).

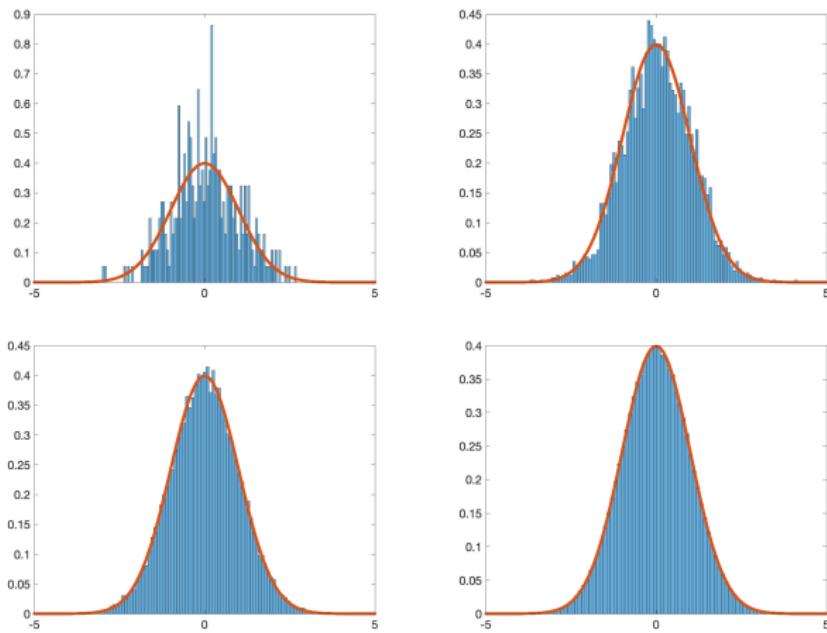


Figure: Histogram of samples defined by the accepted points of the acceptance-rejection method by taking N points for the simulation of Y . We considered (a) $N=500$, (b) $N=5000$, (c) $N=50000$ and $N=500000$. We show also (red line) the plot of the density function of a $\mathcal{N}(0, 1)$ distribution.

We can prove that the probability of acceptance of the method is equal to $\frac{1}{M} = \sqrt{\frac{e}{2\pi}} \approx 0.65774462$. Table 2 shows some estimates for the acceptance probability of the pseudo-random numbers obtained in the simulations presented in Figure 13.

N	$\#\{x_i\}$	$\frac{\#\{x_i\}}{N}$
500	326	0.652
5000	3280	0.656
50000	32767	0.65534
500000	328871	0.657742

Table: Numerical estimates of the probability of acceptance of the acceptance-rejection method. The theoretical probability is equal to $\sqrt{\frac{e}{2\pi}} \approx 0.65774462$.

In the previous section, as an example of the application of the acceptance-rejection method, we discussed an algorithm to generate samples of realizations from the standard normal distribution. Next, we introduce another method which is more efficient for calculating samples from the standard normal distribution which is known as the Box-Muller method.

We start mentioning a classical result for the change of variables in the context of random variables.

Lemma

Let X_1 and X_2 be jointly continuous random variables with joint probability density function f_{X_1, X_2} . Let $Y_1 := g_1(X_1, X_2)$ and $Y_2 := g_2(X_1, X_2)$, for some functions g_1 and g_2 that satisfy the following conditions:

- the equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by, say, $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.
- the functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2) and are such that the following 2×2 determinant

$$J(x_1, x_2) := \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix} \neq 0$$

at all points (x_1, x_2) .

Then, the random variables Y_1 and Y_2 are jointly continuous with joint density function given by

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(x_1, x_2) |J(x_1, x_2)|^{-1},$$

where $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.

Box-Muller method

The Box-Muller method relies on the following result.

Theorem

The random variables X and Y are independent and follow a standard normal distribution if and only if the random variables

$$R := \sqrt{X^2 + Y^2} \quad \text{and} \quad \Theta := \arctan\left(\frac{Y}{X}\right)$$

are such that R^2 follows an exponential distribution with parameter $\frac{1}{2}$ and Θ follows a uniform distribution over the interval $[0, 2\pi]$ and R and Θ are independent.

Box-Muller method

Note that r and θ are the polar coordinates of the point (x, y) , which means that we can write

$$x = r \cos(\theta) \quad \text{and} \quad y = r \sin(\theta).$$

The Box-Muller method can be summarized as follows

Sample d from the exponential distribution with parameter $\frac{1}{2}$,

Sample θ , from the uniform distribution over $[0, 2\pi]$,

Let $r = \sqrt{d}$ and $x = r \cos(\theta); y = r \sin(\theta)$,

Return x and y

Note that Box-Muller method algorithm generates a pair of independent samples from two independent standard normal random variables X and Y .

- ① Implement the congruential generator.
- ② Implement the inversion method for a random variable with probability density function given by

$$g(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

- ③ Implement the acceptance-rejection method for simulating a random variable $X \sim \mathcal{N}(0, 1)$.
- ④ Build a histogram together with the plot of the density function for the distributions considered in (2) and (3).

Monte Carlo method

We start this section with a motivation. Let

$R = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d] \subset \mathbb{R}^d$ be a d -dimensional rectangle and consider the numerical calculation of

$$\int_R g(x) dx,$$

for some function g assumed to be integrable in R . This can be done with classical numerical quadratures. For instance, in the case of having $d = 1$, the simplest quadrature is the midpoint rule which corresponds to the approximation

$$\int_a^b g(x) dx \approx (b - a) g\left(\frac{a + b}{2}\right).$$

Monte Carlo method

If the interval $[a, b]$ is divided into n subintervals of width $h = \frac{b-a}{n}$, and we apply the midpoint rule to each one of the intervals we get the approximation

$$\int_a^b g(x)dx = \sum_{i=1}^n \int_{a+(i-1)h}^{a+ih} g(x)dx \approx Q_n(g) := h \sum_{i=1}^n g(x_i^*),$$

where $x_i^* = a + (i - \frac{1}{2})h$ is the midpoint of the i -th interval.

Monte Carlo method

This procedure can be extended to higher dimensions. For instance, if we consider a partition of R into cells obtained from the division of the intervals $[a_1, b_1], [a_2, b_2], \dots, [a_d, b_d]$ respectively into n_1, n_2, \dots, n_d subintervals having widths equal to $h_1 = \frac{b_1 - a_1}{n_1}, h_2 = \frac{b_2 - a_2}{n_2} \dots h_d = \frac{b_d - a_d}{n_d}$, respectively, we get

$$\int_R g(x) dx = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_d=1}^{n_d} \int_{a_1 + (i_1-1)h_1}^{a_1 + i_1 h_1} \int_{a_2 + (i_2-1)h_2}^{a_2 + i_2 h_2} \dots \int_{a_d + (i_d-1)h_d}^{a_d + i_d h_d} g(x_1, x_2, \dots, x_d) dx$$
$$\approx Q_n(g) := h_1 h_2 \dots h_d \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_d=1}^{n_d} g\left(a_1 + \left(i_1 - \frac{1}{2}\right) h_1, a_2 + \left(i_2 - \frac{1}{2}\right) h_2, \dots, a_d + \left(i_d - \frac{1}{2}\right) h_d\right)$$

Monte Carlo method

We will not discuss in detail this topic but the classical quadrature rules are very effective in low dimensions (1D, 2D or 3D) but they are not a convenient tool for numerical integration in high dimensions.

We will just discuss a simple example for which we are able to compute all the approximations. Assume that we would like to approximate the integral of the function

$g(x_1, x_2, \dots, x_d) = (x_1)^2 + (x_2)^2 + \dots + (x_d)^2$ in the hypercube $[0, 1]^d$, that is, we would like to approximate

$$I_d = \int_{[0,1]^d} \sum_{j=1}^d (x_j)^2 dx.$$

Note that this integral can be calculated in a straightforward way and we have $I_d = \frac{d}{3}$

Monte Carlo method

Now, as an exercise, we will calculate the numerical approximations for this integral obtained by the midpoint rule with n integration points in each dimension. It is easy to prove that

$$\sum_{i=1}^n \left[\left(i - \frac{1}{2} \right) \frac{1}{n} \right]^2 = \frac{4n^2 - 1}{12n}, n \in \mathbb{N}$$

and, as a consequence, in the simplest case $d = 1$ we have $h = \frac{1}{n}$ and the quadrature is given by

$$Q_n(g) = \frac{1}{n} \sum_{i=1}^n \left[\left(i - \frac{1}{2} \right) \frac{1}{n} \right]^2 = \frac{4n^2 - 1}{12n^2} = \frac{1}{3} - \frac{1}{12n^2} = I_1 - \frac{1}{12n^2}.$$

In the d -dimensional case, to simplify the computations, we assume that $n_1 = n_2 = \dots = n_d := n$ and we obtain

$$Q_n(g) = \frac{d}{3} - \frac{d}{12n^2} = I_d - \frac{d}{12n^2}.$$

Therefore, we conclude that the absolute error of the approximation given by the quadrature (in dimension d) is equal to $\frac{d}{12n^2}$.

This implies that we can determine the number of sub-intervals (n) that allows to calculate I_d with an accuracy equal to ϵ ,

$$\frac{d}{12n^2} \approx \epsilon \Rightarrow n = \left\lceil \sqrt{\frac{d}{12\epsilon}} \right\rceil.$$

This means that, for instance when $\epsilon = 10^{-5}$ and $d = 1$, we should consider $n = \left\lceil \sqrt{\frac{1}{12 \times 10^{-5}}} \right\rceil = 92$ sub-intervals and since we are using the midpoint rule, which implies the use of one point for each sub-interval, then we need to consider 92 integration nodes.

Monte Carlo method

However, if we increase the dimension d , the number of integration nodes that are needed for keeping the same accuracy increases very quickly.

For instance, for $d = 5$, we should have $n = 205$ sub-intervals for each dimension, which imply that we should have $205^5 = 362050628125$ integration nodes.

If $d = 10$, we should have $n = 289$ which means that we should consider $289^{10} \approx 4.06 \times 10^{24}$ integration nodes which is prohibitive to be considered in a standard laptop, even using the simplest quadrature which is the midpoint rule, that uses just one integration node for each sub-interval in each dimension.

The situation is even worse if we consider higher order quadratures such the trapezoidal or Simpson rules.

Monte Carlo method

This example illustrates that the classical quadratures are not effective to tackle integrations in high dimensional regions, because the number of integration points needed for keeping a fixed accuracy grows exponentially which implies that very quickly the number of points that are required is prohibitive to be considered. This is known as dimensionality curse.

Moreover, the classical quadratures based on the division of the interval of integration in each dimension by sub-intervals having the same width is not adequate because in high dimensions, roughly speaking, most of the volume of a hypercube is concentrated close to the boundary of the hypercube.

Monte Carlo method

To illustrate this fact, consider the hypercube $[0, 1]^d$, which has volume equal to 1. Now consider a smaller hypercube $H_d := [\delta, 1 - \delta]^d$, for which the volume is equal to

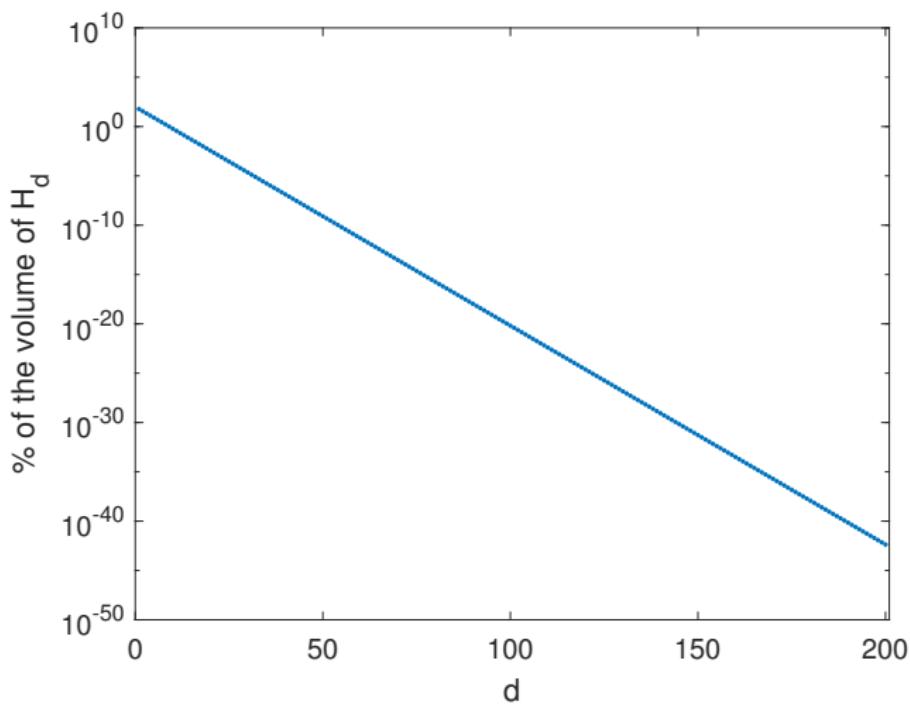
$$|H_d| = (1 - 2\delta)^d.$$

This means that, for instance taking $\delta = 0.1$, we have $|H_1| = 0.8$, which means that the volume of H_1 corresponds to 80% of the volume of the interval $[0, 1]$.

However, the volume of H_d decreases when we increase d . For instance $|H_5| \approx 0.32768$, which means that the volume of H_5 corresponds to just 32.768% of the volume of the hypercube $[0, 1]^5$ and

$|H_{20}| \approx 0.01153$ which means that the volume of H_{20} corresponds to just 1.153% of the volume of the hypercube.

Monte Carlo method



Also it is interesting enough to remark that the Euclidean distance from the “corners” of the hypercube to the “midpoint” of the hypercube can be arbitrarily large, when we increase the dimension d .

Indeed, for the hypercube $[0, 1]^d$, for instance the Euclidean distance between the points $(1, 1, \dots, 1)$ and $(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$ is given by

$$\begin{aligned}\sqrt{\left(1 - \frac{1}{2}\right)^2 + \left(1 - \frac{1}{2}\right)^2 + \dots + \left(1 - \frac{1}{2}\right)^2} &= \sqrt{\frac{d}{4}} \\ &= \frac{\sqrt{d}}{2} \rightarrow \infty, \text{ as } d \rightarrow \infty.\end{aligned}$$

Monte Carlo method

A much convenient approach for the numerical calculation of integrals defined in high dimensional regions is the Monte Carlo integration. For instance, consider a random variable X following a uniform distribution on an interval $[a, b]$. Thus, we have the density function

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

and for a function g that we assume to be integrable in $[a, b]$ we have

$$E(g(X)) = \int_a^b g(x)f(x)dx = \int_a^b \frac{g(x)}{b-a} dx = \frac{1}{b-a} \int_a^b g(x)dx.$$

Therefore, we have

$$\int_a^b g(x)dx = (b-a)E(g(X))$$

Monte Carlo method

$$\int_a^b g(x)dx = (b-a)E(g(X))$$

which means that we can calculate/estimate the integral $\int_a^b g(x)dx$ assuming that we are able to calculate/estimate $E(g(X))$.

We know that an estimator for $E(g(X))$ is

$$\frac{1}{N} \sum_{i=1}^N g(x_i), \tag{24}$$

for independent and identically distributed random variables x_i , $i = 1, \dots, N$.

For Monte Carlo method we can simply consider $x_i \sim \mathcal{U}([a, b])$.

Monte Carlo method

Another possibility of estimating the integral by (24) but choosing the points x_i in such a way that they are almost "equally distributed" in the region of integration, rather than placing them randomly.

Indeed, we can choose those points such that, for instance in the case of hypercube $[0, 1]^d$ and an arbitrary axially parallel d -dimensional rectangle Q - which means that we can define Q as a product of intervals - then the location of the points x_i is such that

$$\frac{\#\{x_i : x_i \in Q\}}{\#\{x_i\}} \approx |Q|.$$

This leads to the definition of discrepancy and points of low discrepancy, such as Halton nodes.

Monte Carlo method

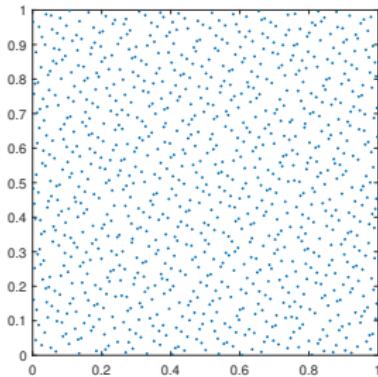
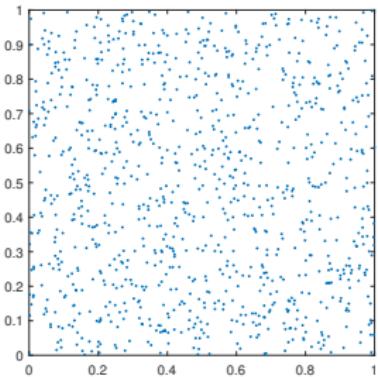


Figure: Plot of 1000 nodes in the square $[0, 1] \times [0, 1]$ obtained from the simulation of the distribution $\mathcal{U}([0, 1])$ to define each of the two components of each point (left plot) and 1000 points obtained by Halton distribution.

Halton nodes provide a better distribution in terms of equidistributedness.

Monte Carlo method

Before defining the sequence of Halton nodes we recall that we can represent an arbitrary integer number \mathcal{N} in a base $b = 2, 3, 4, \dots$ as

$$\mathcal{N} = d_k b^k + d_{k-1} b^{k-1} + \dots + d_1 b^1 + d_0 b^0,$$

where $d_i = 0, 1, \dots, b - 1$, $i = 0, 1, \dots, k$ and we write

$$\mathcal{N} = (d_k d_{k-1} \dots d_1 d_0)_b.$$

Example

- We have $13 = (1101)_2$.
- We have $32 = (1012)_3$.

Definition

For $\mathcal{N} = 1, 2, \dots$ and $b = 2, 3, \dots$ let

$$\mathcal{N} = (d_k d_{k-1} \dots d_1 d_0)_b$$

be the representation of \mathcal{N} in the base b . Then, the radical inverse function is defined by

$$\phi_b(\mathcal{N}) = \sum_{j=0}^k d_j b^{-j-1}.$$

and we can define the Halton nodes

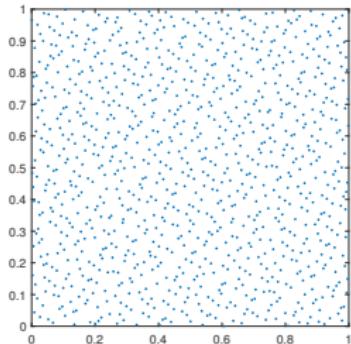
Monte Carlo method

Definition

Let p_1, p_2, \dots, p_m be prime integers. The Halton sequence is defined as the sequence of points

$$x_i := (\phi_{p_1}(\mathcal{N}), \phi_{p_2}(\mathcal{N}), \dots, \phi_{p_m}(\mathcal{N})), \quad \mathcal{N} = 1, 2, 3, \dots$$

The points plotted in the previous Figure were obtained for $p_1 = 2$ and $p_2 = 3$.



Monte Carlo method

A quasi-Monte Carlo method is a numerical technique similar to the Monte Carlo method, but instead of using random samples, we use low-discrepancy sequences.

Thus, the estimate for $E(g(X))$ is given by (24), where the points x_i in (24) are defined by a low-discrepancy sequence such as the Halton sequence.

Monte Carlo method

Example

Consider the calculation of $\int_0^1 e^x dx$ for which we know

$$I = \int_0^1 e^x dx = [e^x]_{x=0}^{x=1} = e^1 - e^0 = e - 1.$$

In order to estimate this integral by Monte Carlo and quasi-Monte Carlo integration methods, we define $g(x) = e^x$ and calculate realizations of N independent samples $x_i \sim \mathcal{U}([a, b])$. The Monte Carlo estimate for $E(e^X)$ is given by

$$I_N = \frac{1}{N} \sum_{i=1}^N g(x_i) = \frac{1}{N} \sum_{i=1}^N e^{x_i}.$$

Monte Carlo method

Example

The estimate from quasi-Monte Carlo method - and we will denote it by \tilde{I}_N is given by

$$\tilde{I}_N = \frac{1}{N} \sum_{i=1}^N g(z_i) = \frac{1}{N} \sum_{i=1}^N e^{z_i},$$

where z_i are Halton nodes.

Monte Carlo method

Example

N	$I_N = \frac{\sum_{k=1}^N g(x_i)}{N}$	$ I - I_N $	$\tilde{I}_N = \frac{\sum_{k=1}^N g(z_i)}{N}$	$ I - \tilde{I}_N $
100	1.592697	1.2558×10^{-1}	1.700212	1.8069×10^{-2}
1000	1.695913	2.2368×10^{-2}	1.716266	2.0155×10^{-3}
10000	1.713663	4.6185×10^{-3}	1.717992	2.8969×10^{-4}
100000	1.718482	2.0053×10^{-4}	1.718245	3.5894×10^{-5}

Table: Numerical estimates for $\int_0^1 e^x dx$ obtained by Monte Carlo and quasi-Monte Carlo integration methods, for different values of N . The exact value of the integral is $I = e - 1 \approx 1.7182818285$

Monte Carlo method

The idea of Monte Carlo and quasi-Monte Carlo integration methods can also be applied in higher dimensions. Let $R \subset \mathbb{R}^d$ be a d -dimensional rectangle and consider the numerical calculation of

$$\int_R g(x)dx,$$

for some function g assumed to be integrable in R .

In this case we consider a random variable X following a uniform distribution on R with the density function

$$f(x) = \frac{1}{|R|}$$

where $|R|$ denotes the volume of R .

Monte Carlo method

Then,

$$E(g(X)) = \int_R g(x)f(x)dx = \int_R g(x)\frac{1}{|R|}dx = \frac{1}{|R|} \int_R g(x)dx$$

which implies that

$$\int_R g(x)dx = |R| \cdot E(g(X)) \quad (25)$$

and again the integral $\int_R g(x)dx$ can be estimated once we have an estimate for $E(g(X))$, that can be calculated by

$$\frac{1}{N} \sum_{i=1}^N g(x_i),$$

where x_1, \dots, x_N are independent and uniformly distributed random samples in R in the case of Monte Carlo method and Halton nodes in the case of quasi-Monte Carlo method.

Monte Carlo method

In general, we will assume that we have a random variable X for which the density function is f , and for a given function g we would like to compute

$$I = E[g(X)] = \int_R g(x)f(x)dx. \quad (26)$$

The Monte Carlo estimator for I is

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^N g(X_i), \quad (27)$$

where X_1, \dots, X_N are independent and identically distributed random variables, for which the density function is f .

We can also define the variance of $g(X)$,

$$Var(g(X)) = E[(g(X) - E[g(X)])^2] := \sigma^2.$$

Lemma

Let $\hat{\theta}$ be an estimator of a parameter θ . Then,

$$E \left[(\hat{\theta} - \theta)^2 \right] = Var(\hat{\theta}) + \left(E(\hat{\theta}) - \theta \right)^2$$

Now we prove that the convergence rate of Monte Carlo is $\mathcal{O} \frac{1}{\sqrt{N}}$ which typically is established in terms of the root-mean-square-error of the estimator \hat{I}_N ,

$$\sqrt{E \left[(\hat{I}_N - I)^2 \right]}.$$

Theorem

Let X be a random variable with density probability function f , g be a given function, and I and its Monte Carlo estimator \hat{I}_N be defined, respectively, by (26) and (27). Then,

$$\sqrt{E[(\hat{I}_N - I)^2]} = \frac{\sigma}{\sqrt{N}}.$$

Note that in Theorem 0.10 the order of convergence of Monte Carlo method is independent of the dimension d which makes the method very appealing for integrations defined in high dimensional regions.

Monte Carlo method

As an example of the application of the Monte Carlo integration method, we apply it to the numerical approximation of π , which is the area of the unitary disk. By the symmetries of the disk, we can focus just on the approximation of the area of the region

$$D = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, 0 < y < 1, x^2 + y^2 \leq 1\}$$

and we have $\pi = 4|D|$.

Taking $R = [0, 1] \times [0, 1]$ and $g(x) = \chi_D(x)$ in (25) where χ is the characteristic function of the region D defined by

$$\chi_D(x) = \begin{cases} 1 & x \in D \\ 0 & \text{otherwise.} \end{cases}$$

we get

$$\int_R \chi_D(x) dx = |R| \cdot E(\chi_D(X)).$$

Monte Carlo method

Now, we note that

$$\int_R \chi_D(x) dx = \int_D 1 dx = |D|$$

and $|R| = 1$, which implies that

$$|D| = E(\chi_D(X))$$

and this expected value can be estimated by

$$\frac{1}{N} \sum_{i=1}^N \chi_D(x_i),$$

where x_1, \dots, x_N are independent and uniformly distributed random samples in R .

Monte Carlo method

Note that

$$\chi_D(x_i) = \begin{cases} 1 & x_i \in D \\ 0 & \text{otherwise.} \end{cases}$$

which implies that

$$\sum_{i=1}^N \chi_D(x_i) = \#\{x_i : x_i \in D\}.$$

Therefore, we have the approximation given by Monte Carlo integration method,

$$|D| \approx \frac{\#\{x_i : x_i \in D\}}{N}, \quad (28)$$

which suggest that in order to estimate $|D|$ we generate a sample of N independent and uniformly distributed random samples in $[0, 1] \times [0, 1]$, x_1, \dots, x_N then count the number of points x_i are in D , and divide that number by N .

Monte Carlo method

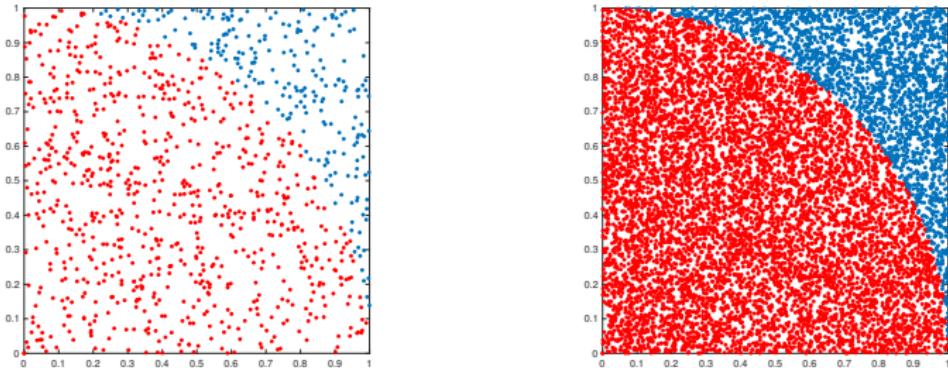


Figure: Left plot shows a sample of $N = 1000$ points generated from a uniform distribution in $[0, 1] \times [0, 1]$. The red points correspond to points that are in D , while the blue points are at the exterior of D . The right plot shows similar results for $N = 10000$.

Monte Carlo method

Then, using (28) we can estimate the area of D and the estimate for π follows immediately by multiplying the previous estimate by four.

N	$\#\{x_i : x_i \in D\}$	$\tilde{\pi} := 4 \frac{\#\{x_i : x_i \in D\}}{N}$	$ \pi - \tilde{\pi} $
100	86	3.44	2.9841×10^{-1}
1000	779	3.116	2.5593×10^{-2}
2000	1570	3.14	1.5927×10^{-3}
3000	2360	3.1466666	5.0740×10^{-3}
1000000	785304	3.141216	3.7665×10^{-4}

Table: Numerical approximations for π , obtained using the Monte Carlo integration method for different values of N and the corresponding absolute error.

Note that the sequence of the absolute error obtained for different values of N is not necessarily a decreasing sequence, as illustrated in Table 4, for $N = 2000$ and $N = 3000$.

Monte Carlo method

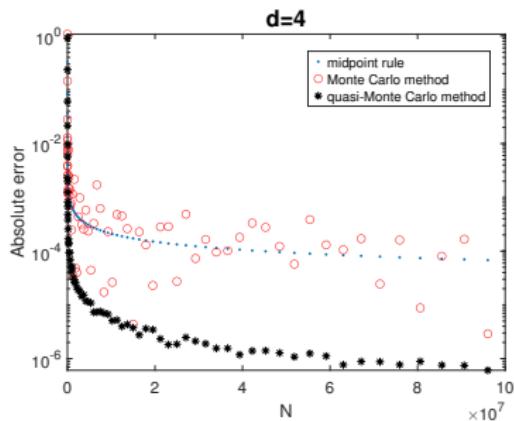
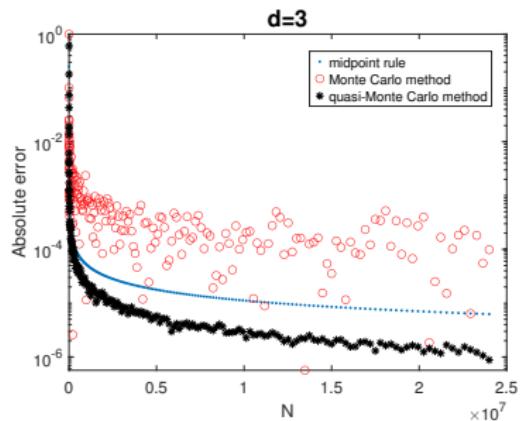
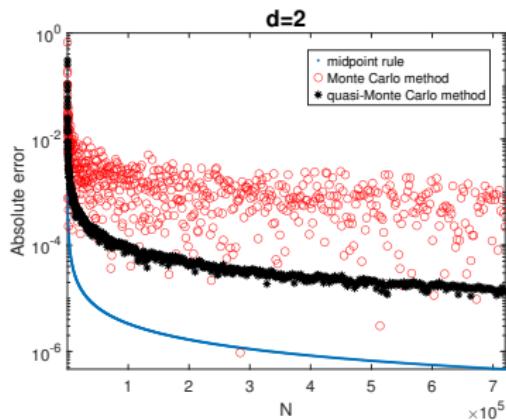
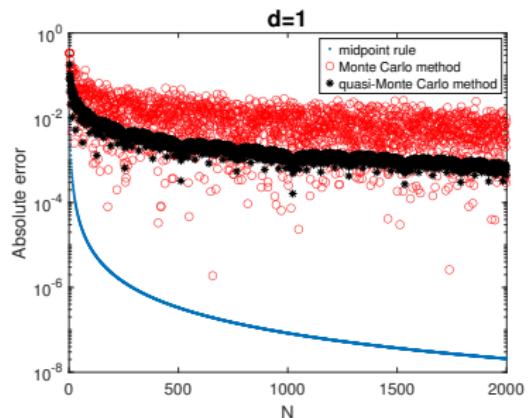
Table 5 shows similar results obtained with quasi-Monte Carlo method.

N	$\#\{z_i : z_i \in D\}$	$\tilde{\pi} := 4 \frac{\#\{z_i : z_i \in D\}}{N}$	$ \pi - \tilde{\pi} $
100	81	3.24	9.8407×10^{-2}
1000	787	3.148	6.4073×10^{-3}
2000	1575	3.15	8.4073×10^{-3}
3000	2363	3.150667	9.0740×10^{-3}
1000000	785393	3.141572	2.0654×10^{-5}

Table: Numerical approximations for π , obtained using the quasi-Monte Carlo integration method for different values of N and the corresponding absolute error.

The last example corresponds to the integration of the function $g(x_1, x_2, \dots, x_d) = \sum_{i=1}^d x_i^2$ in the hypercube $[0, 1]^d$, already considered as a motivation for Monte Carlo method. We consider the approximation for the integral given by the midpoint rule, the Monte Carlo method and quasi-Monte Carlo method, for several choices of the dimension d .

Monte Carlo method



Monte Carlo method

The quasi-Monte Carlo method is also very effective for higher dimensions. Note that the rate at which the error decreases seem to be independent of the dimension.

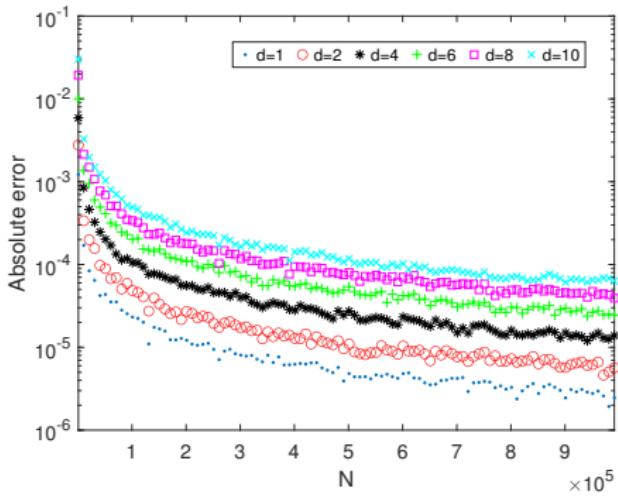


Figure: Plot of the absolute errors of the approximations for the integral $g(x_1, x_2, \dots, x_d) = \sum_{i=1}^d x_i^2$ in the hypercube $[0, 1]^d$, obtained using the quasi-Monte Carlo methods, for $d = 1, 2, 4, 6, 8, 10$.

- ① Implement a routine for generating Halton nodes in an arbitrary dimension. You can use the routines in the zip folder "*example_my_dec2base*" available at the webpage of the course.
- ② Write Matlab code for reproducing the results presented in the slides for the approximation of π with Monte Carlo and quasi-Monte Carlo methods.
- ③ Write Matlab code for reproducing the results presented in the slides for the integration of $g(x_1, x_2, \dots, x_d) = \sum_{i=1}^d x_i^2$ in the hypercube $[0, 1]^d$.

Random walk and Brownian motion

Next, we introduce the discrete-time random walk. Suppose we study a stochastic process corresponding to repeating the process of tossing a (fair) coin. Every time we toss the coin we obtain one of the possibilities “head” or “tail”. Let’s use the notation “H” for denoting a head and “T” for a tail. Then, we define the stochastic process

$$\omega = \omega_1 \omega_2 \omega_3 \dots$$

where each ω_j is either equal to “H” or “T”. For instance two possible realizations of our process could be

$$HHTHTHH\dots \tag{29}$$

or

$$THTTHHT\dots$$

We define Ω to be the probability sample space of all possible outcomes (or realizations) of ω .

We can construct a random variable X_j which takes the values 1 and -1 depending (respectively) whether the j -th toss of the coin produces a “head” or a “tail”, that is

$$X_j = \begin{cases} 1 & \omega_j = H \\ -1 & \omega_j = T. \end{cases}$$

Next, we can define a new stochastic process $\{S_k\}_{k=0}^{\infty}$, given by

$$S_0 = 0, \quad S_k = \sum_{j=1}^k X_j.$$

This stochastic process is called a simple random walk.

For instance, for the realization (29), we have

$$\omega_1 = H, \omega_2 = H, \omega_3 = T, \omega_4 = H, \omega_5 = T, \omega_6 = H, \omega_7 = H, \dots$$

which implies that

$$X_1 = 1, X_2 = 1, X_3 = -1, X_4 = 1, X_5 = -1, X_6 = 1, X_7 = 1, \dots$$

and so

$$S_0 = 0, S_1 = 1, S_2 = 2, S_3 = 1, S_4 = 2, S_5 = 1, S_6 = 2, S_7 = 3, \dots$$

which is represented in Figure 17.

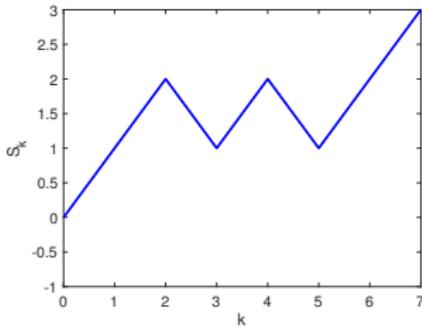


Figure: Illustration of a random walk.

Of course, different realizations of the repeated toss coin process will generate different random walks. In Figure 18 we show three different realizations of random walks.

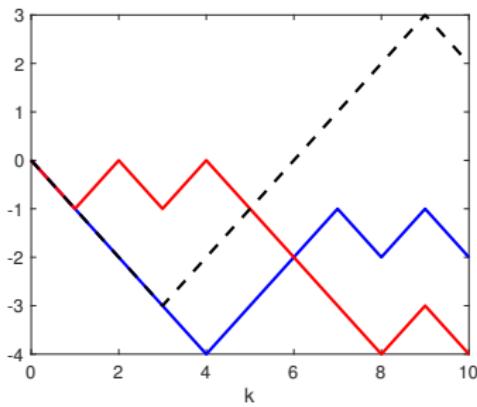


Figure: Three different realizations of random walks.

Note that since $X_j \sim Ber\left(\frac{1}{2}\right)$, with probability mass function

$$p_{X_j}(x) = \begin{cases} \frac{1}{2} & x = 1 \\ \frac{1}{2} & x = -1, \end{cases}$$

we have

$$E(X_j) = 1 \times \frac{1}{2} + (-1) \times \frac{1}{2} = 0 \quad (30)$$

and

$$Var(X_j) = E(X_j^2) - (E(X_j))^2 = E(X_j^2) = 1^2 \times \frac{1}{2} + (-1)^2 \times \frac{1}{2} = 1. \quad (31)$$

Therefore,

$$E(S_k) = E \left(\sum_{j=1}^k X_j \right) = \sum_{j=1}^k E(X_j) = \sum_{j=1}^k 0 = 0 \quad (32)$$

and

$$\text{Var}(S_k) = \text{Var} \left(\sum_{j=1}^k X_j \right) \stackrel{(2)}{=} \sum_{j=1}^k \text{Var}(X_j) = \sum_{j=1}^k 1 = k \quad (33)$$

because the random variables X_j , $j = 1, \dots, k$ are independent.

Moreover, for nonnegative integers

$$0 = k_0 < k_1 < k_2 < \dots < k_m,$$

we define the random variables

$$S_{k_{i+1}} - S_{k_i}$$

which are called increments of the random walk and correspond to the change in the position of the random walk between times k_i and k_{i+1} .

By definition of S_k , we have

$$S_{k_{i+1}} - S_{k_i} = \sum_{j=1}^{k_{i+1}} X_j - \sum_{j=1}^{k_i} X_j = \cancel{\sum_{j=1}^{k_i} X_j} + \sum_{j=k_i+1}^{k_{i+1}} X_j - \cancel{\sum_{j=1}^{k_i} X_j} = \sum_{j=k_i+1}^{k_{i+1}} X_j. \quad (34)$$

Therefore, increments over nonoverlapping time intervals are independent because they depend on different coin tosses, that is,

$$(S_{k_1} - S_{k_0}), (S_{k_2} - S_{k_1}), \dots, (S_{k_m} - S_{k_{m-1}})$$

are independent.

Using (34), we obtain

$$E(S_{k_{i+1}} - S_{k_i}) = E \left(\sum_{j=k_i+1}^{k_{i+1}} X_j \right) = \sum_{j=k_i+1}^{k_{i+1}} E(X_j) = \sum_{j=k_i+1}^{k_{i+1}} 0 = 0$$

and since the different X_j are independent,

$$\begin{aligned} \text{Var}(S_{k_{i+1}} - S_{k_i}) &= \text{Var} \left(\sum_{j=k_i+1}^{k_{i+1}} X_j \right) = \sum_{j=k_i+1}^{k_{i+1}} \text{Var}(X_j) \\ &= \sum_{j=k_i+1}^{k_{i+1}} 1 = k_{i+1} - k_i. \end{aligned}$$

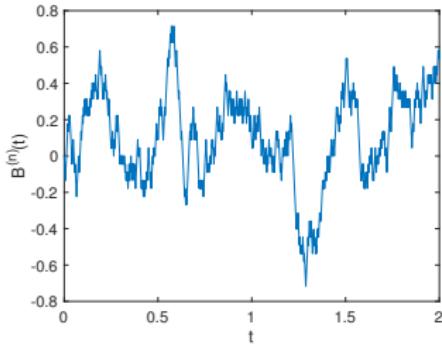
Next, we would like to construct a model of the process of tossing the coin which can occur continuously in time and we will try to define a continuous limit of the simple random walk. For that, we will speed up the time and reduce the time step size of the random walk.

For a given $n \in \mathbb{N}$, we define the scaled random walk

$$B^{(n)}(t) = \frac{S_{nt}}{\sqrt{n}}, \quad (35)$$

provided that $nt \in \mathbb{N}$, otherwise, we define $B^{(n)}(t)$ by linear interpolation between its values at the nearest points l and u to the left and right of t for which nl and nu are integers.

Next figure shows the plot of a scaled random walk $B^{(500)}(t)$, $t \in [0, 2]$.



The procedure starts with the calculation of a realization of a simple random walk and we need to know how many tosses of the coin we must simulate. We note that the time step size of 1 of the simple random walk will correspond to a new time step size of $\frac{1}{500}$, because we took $n = 500$. Then, in order to plot $B^{(500)}(t)$, for $t \in [0, T]$, where $T = 2$ we need to simulate $N = n \times T = 500 \times 2 = 1000$ coin tosses.

It is straightforward to verify that assuming that t is such that the product (nt) is an integer, then

$$E(B^{(n)}(t)) = E\left(\frac{S_{nt}}{\sqrt{n}}\right) = \frac{E(S_{nt})}{\sqrt{n}} \stackrel{(32)}{=} 0 \quad (36)$$

and

$$\begin{aligned} Var(B^{(n)}(t)) &= Var\left(\frac{S_{nt}}{\sqrt{n}}\right) \\ &= \frac{Var(S_{nt})}{n} \stackrel{(33)}{=} \frac{nt}{n} = t \end{aligned} \quad (37)$$

because the random variables X_j , $j = 1, \dots, nt$ are independent.

Moreover, like the simple random walk, the scaled random walk has independent increments. If $0 = t_0 < t_1 < \dots < t_m$ are such that each product (nt_j) is an integer, then

$$\left(B^{(n)}(t_1) - B^{(n)}(t_0) \right), \left(B^{(n)}(t_2) - B^{(n)}(t_1) \right), \dots, \\ \left(B^{(n)}(t_m) - B^{(n)}(t_{m-1}) \right)$$

are independent because they depend on different coin tosses. For instance, the increment $B^{(500)}(0.1) - B^{(500)}(0)$ depends on the first $0.1 \times 500 = 50$ coin tosses, while $B^{(500)}(0.5) - B^{(500)}(0.1)$ depends on the next $(0.5 - 0.1) \times 500 = 200$ coin tosses.

Moreover, for $0 \leq s \leq t$ chosen in such a way that the products (ns) and (nt) are integers, we have

$$\begin{aligned} E\left(B^{(n)}(t) - B^{(n)}(s)\right) &= E\left(\frac{S_{nt}}{\sqrt{n}} - \frac{S_{ns}}{\sqrt{n}}\right) = E\left(\frac{S_{nt}}{\sqrt{n}}\right) - E\left(\frac{S_{ns}}{\sqrt{n}}\right) \\ &= \frac{E(S_{nt})}{\sqrt{n}} - \frac{E(S_{ns})}{\sqrt{n}} \stackrel{(32)}{=} 0 \end{aligned}$$

and

$$\begin{aligned} Var\left(B^{(n)}(t) - B^{(n)}(s)\right) &= Var\left(\frac{S_{nt}}{\sqrt{n}} - \frac{S_{ns}}{\sqrt{n}}\right) \\ &= Var\left(\frac{\sum_{j=1}^{nt} X_j}{\sqrt{n}} - \frac{\sum_{j=1}^{ns} X_j}{\sqrt{n}}\right) \\ &= Var\left(\frac{\sum_{j=ns+1}^{nt} X_j}{\sqrt{n}}\right) \stackrel{(2)}{=} \frac{\sum_{j=ns+1}^{nt} Var(X_j)}{n} \\ &\stackrel{(31)}{=} \frac{\sum_{j=ns+1}^{nt} 1}{n} = \frac{nt - ns}{n} = t - s. \end{aligned}$$

Moreover, if $t \geq 0$ is such that (nt) is an integer, we have

$$\begin{aligned} \sum_{j=1}^{nt} \left[B^{(n)} \left(\frac{j}{n} \right) - B^{(n)} \left(\frac{j-1}{n} \right) \right]^2 &= \sum_{j=1}^{nt} \left[\frac{S_{n \frac{j}{n}}}{\sqrt{n}} - \frac{S_{n \frac{j-1}{n}}}{\sqrt{n}} \right]^2 \\ &= \sum_{j=1}^{nt} \left[\frac{S_j}{\sqrt{n}} - \frac{S_{j-1}}{\sqrt{n}} \right]^2 \\ &= \sum_{j=1}^{nt} \left[\frac{X_j}{\sqrt{n}} \right]^2 = \sum_{j=1}^{nt} \frac{1}{n} = t, \end{aligned}$$

that is, if we move along the path of scaled random walk from time 0 to time t and calculate the sum of the squares of the increments over each time step, we obtain t , the length of the time interval over which we do the calculation. This value is independent of the particular path along which we perform the computation.

In Figure 140 we plotted the scaled random walk $B^{(500)}(t)$, for $t \in [0, 2]$. For instance, for time $t = 0.1$, we have

$$B^{(500)}(0.1) = \frac{S_{50}}{\sqrt{500}}.$$

In particular, we can determine which values can be obtained in the simulations of $B^{(500)}(0.1)$. We start by determining the possible values for S_{50} depending on the number of heads and tails obtained in 50 coin tosses.

The values obtained are presented in next Table.

N. of tails	50	49	48	47	...	26	25	24	...	3	2	1	0
N. of heads	0	1	2	3	...	24	25	26	...	47	48	49	50
Value of S_{50}	-50	-48	-46	-44	...	-2	0	2	...	44	46	48	50

Table: Possible values for S_{50} .

Then, the possible values for $B^{(500)}(0.1)$ are obtained by dividing the values presented in Table 6 by $\sqrt{500}$, that is, the possible values for $B^{(500)}(0.1)$ are

$$-\frac{50}{\sqrt{500}}, -\frac{48}{\sqrt{500}}, -\frac{46}{\sqrt{500}}, \dots, 0, \dots, \frac{46}{\sqrt{500}}, \frac{48}{\sqrt{500}}, \frac{50}{\sqrt{500}}.$$

We can calculate the probability of obtaining each of these values.

Let X be random variable defining the number of heads that we obtain in 50 coin tosses of a (fair) coin. We know that $X \sim Bin(50, \frac{1}{2})$ and the probability mass function is

$$p_X(k) = \frac{50!}{k!(50-k)!} \left(\frac{1}{2}\right)^{50}, \text{ for } k = 0, 1, \dots, 50.$$

In particular, for instance, we get

$$p_X(0) = \frac{50!}{0!(50)!} \left(\frac{1}{2}\right)^{50} = \left(\frac{1}{2}\right)^{50}$$

$$p_X(1) = \frac{50!}{1!(49)!} \left(\frac{1}{2}\right)^{50} = 50 \times \left(\frac{1}{2}\right)^{50}$$

⋮

$$p_X(25) = \frac{50!}{25!(25)!} \left(\frac{1}{2}\right)^{50} \approx 0.112275$$

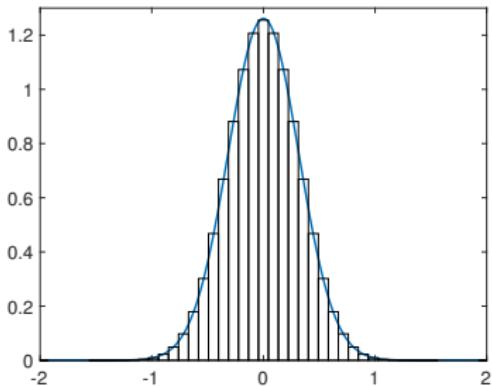
⋮

and after performing all the computations for $k = 0, 1, \dots, 50$ we can plot an histogram. For instance, we concluded that

$$\begin{aligned} P\left(B^{(500)}(0.1) = 0\right) &= P(S_{50} = 0) = p_X(25) \\ &= \frac{50!}{25!(25)!} \left(\frac{1}{2}\right)^{50} \approx 0.112275 \end{aligned}$$

and we will plot a histogram bar centered at 0 with area approximately equal to 0.112275 and since the bars have width equal to $\frac{2}{\sqrt{500}} \approx 0.0894427$ the corresponding bar height shall be equal to $\frac{0.112275}{0.0894427} \approx 1.25527$. Applying the same procedure to the other bars, we obtain the histogram plotted in Figure 149. By (36) and (37) we know that $E(B^{(500)}(0.1)) = 0$ and $V(B^{(500)}(0.1)) = 0.1$.

We plot in the same Figure the plot of a probability density function of a distribution $\mathcal{N}(0, 0.1)$.



We can observe that the distribution of $B^{(500)}(0.1)$ is approximately normal. Indeed, by the central limit theorem it is possible to prove that for a fixed value of $t \geq 0$, the distribution of the scaled random walk $B^{(n)}(t)$ converges to the normal distribution with mean zero and variance t .

It can be proven that when $N \rightarrow \infty$, the distribution of the process $\{B^{(N)}(t), 0 \leq t \leq 1\}$ converges to a process $\{B(t), 0 \leq t \leq 1\}$, that is called Brownian motion, that satisfies the following properties:

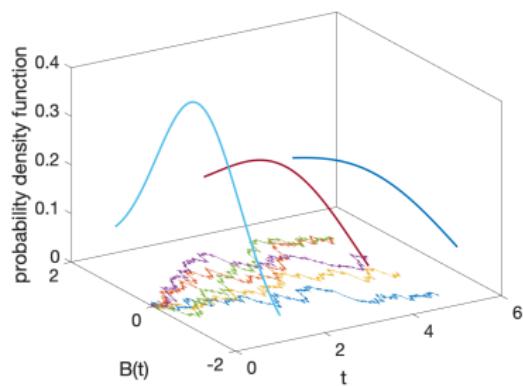
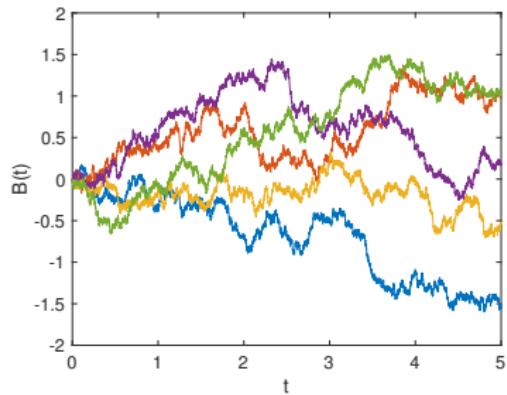
- ① $B(0) = 0$.
- ② for any $0 \leq s \leq t \leq 1$, the random variable $B(t) - B(s)$ is normally distributed with expected value equal to 0 and variance equal to $t - s$, that is $B(t) - B(s) \sim \mathcal{N}(0, t - s)$.
- ③ for any integer k and any partition $0 \leq t_1 \leq t_2 \leq \dots \leq t_k \leq 1$, the random variables $B(t_1)$, $B(t_2) - B(t_1)$, ... and $B(t_k) - B(t_{k-1})$ are independent.
- ④ the trajectory $t \mapsto B(t)$ is continuous.

Note that (ii) implies that

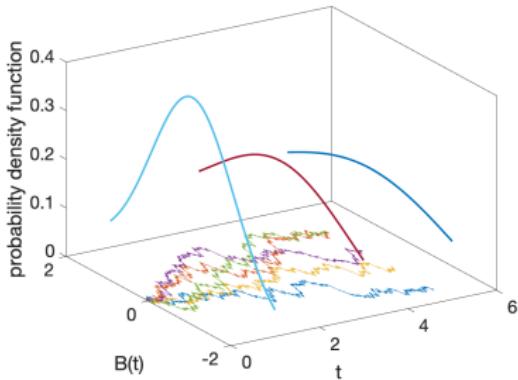
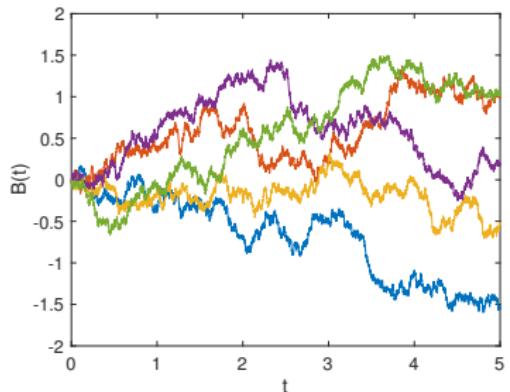
$$B(t) - B(0) = B(t) \sim \mathcal{N}(0, t), \quad \forall t \geq 0, \quad (38)$$

that is, for each fixed value of $t > 0$, the Brownian motion is a random variable normally distributed with expected value equal to 0. The variance is equal to t , which means that the larger the value of t , the larger the variance.

Next figure-left shows five realizations of a Brownian motion, for $t \in [0, 5]$.



The right plot shows the same five realizations of the Brownian motion together with the plot of the probability density functions of distributions $\mathcal{N}(0, 1)$, $\mathcal{N}(0, 3)$ and $\mathcal{N}(0, 5)$, respectively for $t = 1$, $t = 3$ and $t = 5$.



For small values of t , the variance is also small which means that all the values that are obtained are relatively close to the expected value (which is equal to zero) and the plot of the probability density function has a narrow bell shape. For larger values of t the values obtained for the different realizations of the Brownian motion are spread out from the expected value and the probability density function shows a wide bell shape.

Note that the scaled random walk $B^{(n)}(t)$ has a natural time step $\frac{1}{n}$ and is linear between these time steps (it is a linear spline) but the Brownian motion has no linear pieces.

Moreover, the scaled random walk is only approximately normal for each value of t , as illustrated in Figure 149 while the Brownian motion follows exactly a normal distribution and the increments $B(t) - B(s)$, for $0 \leq s \leq t \leq 1$ follow also a normal distribution.

Also from (ii), we have that

$$E(B(t) - B(s)) = 0$$

and

$$\text{Var}(B(t) - B(s)) = t - s.$$

In particular, for a small positive parameter h_t we get

$$\text{Var}(B(t + h_t) - B(t)) = t + h_t - t = h_t$$

and

$$\text{Var} \left(\frac{B(t + h_t) - B(t)}{h_t} \right) = \frac{\text{Var}(B(t + h_t) - B(t))}{h_t^2} = \frac{h_t}{h_t^2} = \frac{1}{h_t}$$

which means that the variance will become arbitrarily large, by taking $h_t \rightarrow 0$ and the Brownian motion is not differentiable. Indeed, the Brownian motion is continuous everywhere but nowhere differentiable.

Stochastic integral

In this section we will define the (Itô) stochastic integral and start recalling the definition of the Riemann integral.

Riemann integral

Let $g : [a, b] \rightarrow \mathbb{R}$ be a function, fix $n \in \mathbb{N}$ and let

$$\Pi_n = \{a = t_0 < t_1 < \dots < t_{n-1} < t_n = b\}$$

be a partition of the interval $[a, b]$ and take a point in each of the sub-intervals $[t_{i-1}, t_i]$, say $t_i^* \in [t_{i-1}, t_i]$, $i = 1, 2, \dots, n$ defining $\gamma_n^* = \{t_i^*, i = 1, 2, \dots, n\}$ and define $h_i := t_i - t_{i-1}$. We say that

$$S(g; \Pi_n; \gamma_n^*) := \sum_{i=1}^n g(t_i^*) h_i$$

is the Riemann sum of g corresponding to the partition Π_n and set of points γ_n^* . We say that $\Pi = \{\Pi_n, n = 1, 2, \dots\}$ is a refinement of the interval $[a, b]$ if Π is a sequence of partitions of $[a, b]$ such that $\Pi_n \subset \Pi_{n+1}$, for all n .

Then, we can define the Riemann integral.

Definition

(Riemann integral) We say that g is Riemann integrable over the interval $[a, b]$ if for any $\epsilon > 0$ and for every refinement

$\Pi = \{\Pi_n, n = 1, 2, \dots\}$ such that $\max_{i=1,2,\dots,n} h_i \rightarrow 0$, as $n \rightarrow \infty$, there exist $N \in \mathbb{N}$ and $I \in \mathbb{R}$ such that

$$|S(g; \Pi_m; \gamma_m^*) - I| < \epsilon$$

for all choices of γ_m^* and for all $m \geq N$. In that case, we say that I is the Riemann integral of g over the interval $[a, b]$ and use the notation

$$I = \int_a^b g(s) ds.$$

There are other (equivalent) definitions of Riemann integral, but this variant will be the most useful in the context of the construction of a definition of a stochastic integral.

Remark

We can verify that if $g : [0, 1] \rightarrow \mathbb{R}$ is Riemann integrable on $[0, 1]$, then the limit of the left-hand Riemann sum

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g\left(\frac{i-1}{n}\right)$$

and limit of right-hand Riemann sum

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g\left(\frac{i}{n}\right)$$

both exist. So, we have

$$I = \int_0^1 g(t) dt = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g\left(\frac{i-1}{n}\right) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g\left(\frac{i}{n}\right).$$

We recall a well known result that allows to define the class of Riemann integrable functions.

Theorem

Let $g : [a, b] \rightarrow \mathbb{R}$ be a function that is bounded. Then, the function g is Riemann integrable on $[a, b]$ if and only if the set of discontinuities of g has zero Lebesgue measure.

Next, we will construct the definition of the Itô integral. Before introducing the integral with respect to Brownian motion we start by discussing how to manage to integrate the Brownian motion itself. Since the Brownian motion is continuous, by Theorem 0.13, it shall be Riemann integrable. We will use the notation $B_\omega(t)$, $0 \leq t \leq 1$ to denote a realization of Brownian motion which (for a fixed realization) is a function of t . Let

$$I = \int_0^1 B_\omega(t) dt \tag{39}$$

denote the Riemann integral of the function B_ω on $[0, 1]$. Some naïve comments about what I should verify:

- We know from elementary calculus that the Riemann integral may be interpreted as the area under the graph of the integrand.
- Note that the value of I depends on the particular realization of the Brownian motion that was considered and different realizations can lead to different values of I . Thus, I can be seen as a function of ω , that is

$$I(\omega) = \int_0^1 B_\omega(t) dt.$$

Therefore, I is a random variable and we shall identify its distribution.

Since the Riemann integral of Brownian motion (39) exist, by Remark 0.12 we have

$$I = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n B\left(\frac{i}{n}\right)$$

and we start by calculating the distribution of

$$I^{(n)} := \frac{1}{n} \sum_{i=1}^n B\left(\frac{i}{n}\right). \quad (40)$$

We know that Brownian motion increments are independent and we can write

$$\begin{aligned}\sum_{i=1}^n B\left(\frac{i}{n}\right) &= nB\left(\frac{1}{n}\right) + (n-1)\left(B\left(\frac{2}{n}\right) - B\left(\frac{1}{n}\right)\right) \\ &\quad + (n-2)\left(B\left(\frac{3}{n}\right) - B\left(\frac{2}{n}\right)\right) \\ &\quad + \dots + 2\left(B\left(\frac{n-1}{n}\right) - B\left(\frac{n-2}{n}\right)\right) \\ &\quad + \left(B\left(\frac{n}{n}\right) - B\left(\frac{n-1}{n}\right)\right).\end{aligned}$$

Moreover, we note that by the properties of Brownian motion, we have

$$B\left(\frac{i}{n}\right) - B\left(\frac{i-1}{n}\right) \sim \mathcal{N}\left(0, \frac{1}{n}\right), \quad i = 1, 2, \dots, n$$

and by (38) we have

$$B\left(\frac{i}{n}\right) \sim \mathcal{N}\left(0, \frac{i}{n}\right), \quad i = 1, 2, \dots, n.$$

The sum of independent normal distributions is also a normal distribution and, by independence of Brownian motion increments we have

$$\begin{aligned} E \left(\sum_{i=1}^n B \left(\frac{i}{n} \right) \right) &= E \left(nB \left(\frac{1}{n} \right) \right) \\ &\quad + E \left((n-1) \left(B \left(\frac{2}{n} \right) - B \left(\frac{1}{n} \right) \right) \right) \\ &\quad + E \left((n-2) \left(B \left(\frac{3}{n} \right) - B \left(\frac{2}{n} \right) \right) \right) + \dots \\ &\quad + E \left(2 \left(B \left(\frac{n-1}{n} \right) - B \left(\frac{n-2}{n} \right) \right) \right) \\ &\quad + E \left(B \left(\frac{n}{n} \right) - B \left(\frac{n-1}{n} \right) \right) \\ &= nE \left(B \left(\frac{1}{n} \right) \right) + (n-1)E \left(B \left(\frac{2}{n} \right) - B \left(\frac{1}{n} \right) \right) \\ &\quad + (n-2)E \left(B \left(\frac{3}{n} \right) - B \left(\frac{2}{n} \right) \right) + \dots \\ &\quad + 2E \left(B \left(\frac{n-1}{n} \right) - B \left(\frac{n-2}{n} \right) \right) \\ &\quad + E \left(B \left(\frac{n}{n} \right) - B \left(\frac{n-1}{n} \right) \right) \\ &= 0 \end{aligned}$$

which implies that

$$E(I^{(n)}) = 0$$

and

$$\begin{aligned} Var \left(\sum_{i=1}^n B \left(\frac{i}{n} \right) \right) &= n^2 Var \left(B \left(\frac{1}{n} \right) \right) \\ &\quad + (n-1)^2 Var \left(B \left(\frac{2}{n} \right) - B \left(\frac{1}{n} \right) \right) \\ &\quad + (n-2)^2 Var \left(B \left(\frac{3}{n} \right) - B \left(\frac{2}{n} \right) \right) + \dots \\ &\quad + 2^2 Var \left(B \left(\frac{n-1}{n} \right) - B \left(\frac{n-2}{n} \right) \right) \\ &\quad + Var \left(B \left(\frac{n}{n} \right) - B \left(\frac{n-1}{n} \right) \right) \\ &= [n^2 + (n-1)^2 + \dots + 2^2 + 1] \frac{1}{n}. \end{aligned}$$

By induction, we can prove that

$$n^2 + (n-1)^2 + \dots + 2^2 + 1 = \frac{2n^3 + 3n^2 + n}{6}$$

which implies that

$$\text{Var} \left(\sum_{i=1}^n B \left(\frac{i}{n} \right) \right) = \frac{2n^2 + 3n + 1}{6}$$

and

$$\text{Var} \left(I^{(n)} \right) = \frac{2n^2 + 3n + 1}{6n^2} = \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}.$$

Therefore,

$$I^{(n)} \sim \mathcal{N} \left(0, \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} \right).$$

Now, we will make use of the following theorem

Theorem

Suppose that X_1, X_2, \dots is a sequence of random variables with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots$. If the limits

$$\lim_{n \rightarrow \infty} \mu_n \quad \text{and} \quad \lim_{n \rightarrow \infty} \sigma_n^2$$

exist and are finite, then the sequence $\{X_n, n = 1, 2, \dots\}$ converges in distribution to a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, where

$$\mu = \lim_{n \rightarrow \infty} \mu_n \quad \text{and} \quad \lim_{n \rightarrow \infty} \sigma_n^2.$$

Thus, since

$$\lim_{n \rightarrow \infty} \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} = \frac{1}{3}$$

we conclude that

$$I \sim \mathcal{N}\left(0, \frac{1}{3}\right).$$

Roughly speaking, this result states that if we consider realizations of Brownian motion up to time 1, then the area under the graph of the path is normally distributed with mean 0 and variance $\frac{1}{3}$.

- ① Implement a routine for simulating a Bernoulli trial. It can call another routine that simulates the distribution $\mathcal{U}([0, 1])$.
- ② Implement a routine for simulating a simple random walk.
- ③ Use the routine of (2) to generate realizations of the simple random walk. Plot them and estimate the mean and the variance.
- ④ Implement a routine for simulating a scaled random walk.
- ⑤ Use the routine of (4) to generate realizations of the scaled random walk. Plot them and estimate the mean and the variance.
- ⑥ Implement a routine to calculate $\int_0^1 B^{(n)}(t)dt$, for a given realization of the scaled random walk. Estimate the mean and the variance.
- ⑦ Plot a histogram of the values obtained for $\int_0^1 B^{(n)}(t)dt$, together with the density probability function of a $\mathcal{N}(0, \frac{1}{3})$.

Next, we construct the Itô stochastic integral the integral with respect to the Brownian motion. We start with the case where the integrand is a step function. We consider a partition of the interval $[0, T]$,

$$0 = t_0 < t_1 < \dots < t_n = T$$

and assume that

$$f(t) = \sum_{i=1}^n a_i \chi_{[t_{i-1}, t_i]}(t), \quad a_i \in \mathbb{R}, \quad i = 1, 2, \dots, n, \quad (41)$$

where $\chi_A(t) = \begin{cases} 1, & t \in A \\ 0 & \text{otherwise} \end{cases}$ is the characteristic function of the set A .

For this type of functions, the Riemann integral of f is given by

$$\int_0^T f(t) dt = \sum_{i=1}^n a_i (t_i - t_{i-1})$$

that can be interpreted as the area below the graph of f . The stochastic integral is interpreted as the sum of profits and losses

$a_i (B(t_i) - B(t_{i-1}))$, $i = 1, 2, \dots, n$ in a portfolio of a quantity of a_i risky asset whose price variation is $B(t_i) - B(t_{i-1})$ at time $i = 1, 2, \dots, n$.

Definition

The stochastic integral of a step function f of type (41) with respect to the Brownian motion $B(t)$, $t \in [0, T]$ is defined by

$$\int_0^T f(t) dB(t) = \sum_{i=1}^n a_i (B(t_i) - B(t_{i-1})). \quad (42)$$

Note that the stochastic integral of step functions is a linear operator.

Theorem

Let f be a step function associated to the partition of the interval $[0, T]$ given by

$$0 = t_0^{(f)} < t_1^{(f)} < \dots < t_p^{(f)} = T,$$

$$f(t) = \sum_{i=1}^p a_i \chi_{[t_{i-1}^{(f)}, t_i^{(f)}]}(t), \quad a_i \in \mathbb{R}, \quad i = 1, 2, \dots, p$$

and g be a step function associated to the partition

$$0 = t_0^{(g)} < t_1^{(g)} < \dots < t_m^{(g)} = T,$$

$$g(t) = \sum_{i=1}^m b_i \chi_{[t_{i-1}^{(g)}, t_i^{(g)}]}(t), \quad b_i \in \mathbb{R}, \quad i = 1, 2, \dots, m.$$

Then, we have

$$\int_0^T (\alpha f(t) + \beta g(t)) dB(t) = \alpha \int_0^T f(t) dB(t) + \beta \int_0^T g(t) dB(t).$$

Next, we define a convenient functional space in this context.

Definition

The space $L^2([0, T])$ is defined as the functional space of measurable functions $f : [0, T] \rightarrow \mathbb{R}$ such that

$$\|f\|_{L^2([0, T])} := \sqrt{\int_0^T |f(t)|^2 dt} < \infty.$$

Note that $\|\cdot\|$ is a norm, thus it induces a notion of distance between any two functions f and g belonging to $L^2([0, T])$ by

$$\|f - g\|_{L^2([0, T])} = \sqrt{\int_0^T |f(t) - g(t)|^2 dt}.$$

We can also define convergence in $L^2([0, T])$.

Definition

We say that a sequence $(f_n)_{n \in \mathbb{N}}$ of functions in $L^2([0, T])$ converges in $L^2([0, T])$ to another function $f \in L^2([0, T])$ if

$$\lim_{n \rightarrow \infty} \|f - f_n\|_{L^2([0, T])} = 0.$$

Next, we determine the distribution of $\int_0^T f(t)dB(t)$ when f is a step function.

Theorem

Let f be a step function of type (41). Then,

$$\int_0^T f(t)dB(t) \sim \mathcal{N} \left(0, \int_0^T |f(t)|^2 dt \right).$$

Proof: Makes use of Theorem 0.2

Therefore,

$$\int_0^T f(t) dB(t) \sim \mathcal{N} \left(0, \int_0^T |f(t)|^2 dt \right). \quad \square$$

The last theorem can be extended to an arbitrary function $f \in L^2([0, T])$, not necessarily a step function, which can be proven by using the fact that any function $f \in L^2([0, T])$ can be arbitrarily approximated by step functions, that is, there exists a sequence $(f_n)_{n \in \mathbb{N}}$ of step functions converging to f in $L^2([0, T])$.

Theorem

Let $f \in L^2([0, T])$. Then,

$$\int_0^T f(t) dB(t) \sim \mathcal{N} \left(0, \int_0^T |f(t)|^2 dt \right).$$

Example

We can determine the distribution of $\int_0^T e^{-\frac{t}{2}} dB(t)$. Since, for $f(x) = e^{-\frac{x}{2}}$ we have

$$\int_0^T |f(t)|^2 dt = \int_0^T e^{-t} dt = [-e^{-t}]_{t=0}^{t=T} = -e^{-T} + 1$$

The stochastic integral of f verifies

$$\int_0^T f(t) dB(t) \sim \mathcal{N}\left(0, -e^{-T} + 1\right).$$

Example

Figure-left shows the histogram of 1000 realizations of $\int_0^1 f(t) dB(t)$ together with the plot of the density function of a $\mathcal{N}(0, -e^{-1} + 1)$ distribution. The right plot of the same Figure corresponds to similar results obtained for 1 000 000 realizations.

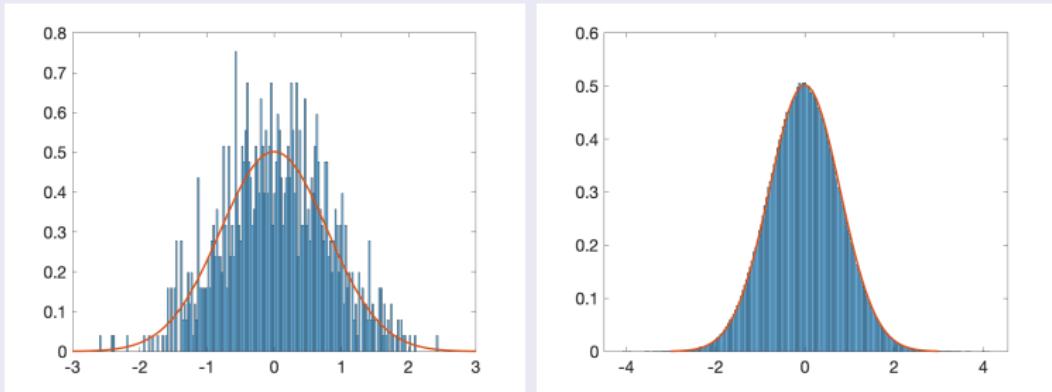


Figure: Histogram of 1000 realizations of $\int_0^1 f(t) dB(t)$, for $f(x) = e^{-\frac{x}{2}}$ together with the plot of the density function of a $\mathcal{N}(0, -e^{-1} + 1)$ distribution. The right plot corresponds to similar results obtained for 1 000 000 realizations.

Theorem 0.16 can immediately be extended to general $L^2([0, T])$ functions,

Theorem

Let $f, g \in L^2([0, T])$ and $\alpha, \beta \in \mathbb{R}$. Then,

$$\int_0^T (\alpha f(t) + \beta g(t)) dB(t) = \alpha \int_0^T f(t) dB(t) + \beta \int_0^T g(t) dB(t).$$

Now we prove that the stochastic integral satisfies an isometry property

Theorem

Let $f, g \in L^2([0, T])$. Then, we have

$$E \left[\int_0^T f(t) dB(t) \int_0^T g(t) dB(t) \right] = \int_0^T f(t) g(t) dt.$$

Remark

In example 10 we concluded that if $f(x) = e^{-\frac{x}{2}}$ we have

$$\int_0^T |f(t)|^2 dt = \int_0^T e^{-t} dt = [-e^{-t}]_{t=0}^{t=T} = -e^{-T} + 1.$$

The same result could be obtained using Theorem 0.22. Indeed, since

$$E \left(\int_0^T f(t) dB(t) \right) = 0,$$

then

$$\begin{aligned} \text{Var} \left(\int_0^T f(t) dB(t) \right) &= E \left[\left(\int_0^T f(t) dB(t) \right)^2 \right] \\ &= E \left[\int_0^T f(t) dB(t) \int_0^T f(t) dB(t) \right] \\ &= \int_0^T (f(t))^2 dt = -e^{-T} + 1. \end{aligned}$$

Next, we extend the stochastic integral from deterministic functions to stochastic processes. In the context of finance we will consider simple predictable stochastic processes. These processes depend just on the information known up to time t which implies that an investor can not “look into the future”. Again, we will consider a partition

$$\Pi_n = \{t_0 = 0 < t_1 < \dots < t_{n-1} < t_n = T\}$$

and introduce the following definition.

Definition

A simple predictable process is a stochastic process of the form

$$u_\omega(t) = \sum_{i=1}^n F_i \chi_{[t_{i-1}, t_i]}(t), \quad t \geq 0, \tag{43}$$

where F_i , $i = 1, \dots, n$ are random variables allocated at time t_{i-1} and kept constant over the whole trading interval $[t_{i-1}, t_i]$.

For instance, a simple predictable process is

$$u_\omega(t) = \sum_{i=1}^n F_i \chi_{[t_{i-1}, t_i[}(t) := \sum_{i=1}^n B(t_{i-1}) \chi_{[t_{i-1}, t_i[}(t), \quad t \geq 0.$$

Sometimes, for simplicity of notation we will drop the random outcome ω and write simply $u(t)$ but we should keep in mind that $u(t)$ is a stochastic process. Now we can define the stochastic integral of $u(t)$.

Definition

The stochastic integral of a simple predictable process $u(t)$ with respect to the Brownian motion $B(t)$, $t \in [0, T]$ is defined by

$$\int_0^T u(t) dB(t) = \sum_{i=1}^n F_i (B(t_i) - B(t_{i-1})). \quad (44)$$

Now, we can extend the construction of the stochastic integral from simple predictable processes to square-integrable processes for which the value at time t depends only on the information contained in the Brownian motion realization up to time t . We start by defining the space of square-integrable stochastic processes.

Definition

The space $L^2(\Omega \times [0, T])$ is defined as the functional space of stochastic processes

$$\begin{aligned} u : \Omega \times [0, T] &\rightarrow \mathbb{R} \\ (\omega, t) &\mapsto u_\omega(t) \end{aligned}$$

such that

$$\|u\|_{L^2(\Omega \times [0, T])} := \sqrt{E \left[\int_0^T |u_\omega(t)|^2 dt \right]} < \infty.$$

Note that $\|\cdot\|_{L^2(\Omega \times [0, T])}$ is a norm, and so, it induces a notion of distance between any two stochastic processes u and v in $L^2(\Omega \times [0, T])$ by

$$\|u - v\|_{L^2(\Omega \times [0, T])} = \sqrt{E \left[\int_0^T |u(t) - v(t)|^2 \right]}.$$

We can also define convergence in $L^2(\Omega \times [0, T])$.

Definition

We say that a sequence $(u^{(n)})_{n \in \mathbb{N}}$ of stochastic processes in $L^2(\Omega \times [0, T])$ converges in $L^2(\Omega \times [0, T])$ to another stochastic process $u \in L^2(\Omega \times [0, T])$ if

$$\lim_{n \rightarrow \infty} \|u - u^{(n)}\|_{L^2(\Omega \times [0, T])} = 0.$$

It can be proven that any stochastic process in $u_\omega(t) \in L^2(\Omega \times [0, T])$ that depend only on the information contained in the Brownian motion realization up to time t can be arbitrarily approximated by simple predictable processes. Therefore, we can extend the definition of stochastic integral to that case. Moreover, we have

$$E \left[\int_0^T u(t) dB(t) \right] = 0 \quad (45)$$

and

$$E \left[\int_0^T u(t) dB(t) \int_0^T v(t) dB(t) \right] = E \left[\int_0^T u(t)v(t) dt \right], \quad (46)$$

for all stochastic processes u and v in $L^2(\Omega \times [0, T])$ that depend only on the information contained in the Brownian motion realization up to time t .

Moreover, in that case we can also define the definite stochastic integral over an interval $[a, b] \subset [0, T]$ simply by

$$\int_a^b u(t) dB(t) = \int_0^T (u(t) \cdot \chi_{[a,b]}(t)) dB(t)$$

from which we conclude that

$$\int_a^c u(t) dB(t) = \int_a^b u(t) dB(t) + \int_b^c u(t) dB(t),$$

for $0 \leq a < b < c \leq T$.

Stochastic differential equations

In this section we will study stochastic differential equations (SDE) which can be written using integral representation that we have already used. For instance, assume that for a certain stochastic process $X(t)$ we have

$$X(t) - X(0) = \int_0^t a(s, X(s))ds + \int_0^t b(s, X(s))dB(s), \quad (47)$$

for some functions a and b . The equality (47) can be written using notation of differentials,

$$dX(t) = a(t, X(t))dt + b(t, X(t))dB(t) \quad (48)$$

which is a SDE that shall be understood as symbolic short form of the stochastic integral equation (47).

A SDE of type (48) is called an Itô SDE and in that case $X(t)$ is said to be an Itô process.

The term $a(t, X(t))$ is the drift and the term $b(t, X(t))$ is the diffusion coefficient.

The Brownian motion is particular case of an Itô process corresponding to the SDE $dX(t) = dB(t)$ which is a particular case of Itô SDE (48) for which $a \equiv 0$ and $b \equiv 1$.

An important mathematical result in this context is the Itô's formula that plays the same role in the context of stochastic integration as the fundamental theorem of calculus does for the integration of deterministic functions.

For a motivation of Itô's formula we start proving the chain rule for deterministic functions.

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ be infinitely differentiable functions. Then, we know that

$$\frac{d [f(g(t))]}{dt} = g'(t) \cdot f'(g(t)) \quad (49)$$

which is the well known chain rule. It is illustrative to prove this result using Taylor theorem.

Since f is infinitely differentiable, then we know that it can be expanded around $a \in \mathbb{R}$ as

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots \quad (50)$$

and taking $a = t$ and $x = t + h$, for some $h > 0$ we get

$$f(t+h) = f(t) + f'(t)h + \frac{f''(t)}{2}h^2 + \frac{f'''(t)}{3!}h^3 + \dots$$

from which we obtain

$$\frac{f(t+h) - f(t)}{h} = f'(t) + \frac{f''(t)}{2}h + \frac{f'''(t)}{3!}h^2 + \dots$$

and taking the limit, when $h \rightarrow 0$,

$$\lim_{h \rightarrow 0} \left[\frac{f(t+h) - f(t)}{h} \right] = \lim_{h \rightarrow 0} \left[f'(t) + \frac{f''(t)}{2}h + \frac{f'''(t)}{3!}h^2 + \dots \right] = f'(t)$$

which is the definition of the derivative of f .

Now, we prove the chain rule, taking $x = g(t + h)$ and $a = g(t)$ in (50). We obtain

$$\begin{aligned} f(g(t + h)) &= f(g(t)) + f'(g(t)).[g(t + h) - g(t)] \\ &\quad + \frac{f''(g(t))}{2}.[g(t + h) - g(t)]^2 \\ &\quad + \frac{f'''(g(t))}{3!}.[g(t + h) - g(t)]^3 + \dots \end{aligned}$$

Dividing by h we get

$$\begin{aligned} \frac{f(g(t + h)) - f(g(t))}{h} &= f'(g(t)).\frac{g(t + h) - g(t)}{h} \\ &\quad + \frac{f''(g(t))}{2}.\frac{[g(t + h) - g(t)]^2}{h} \\ &\quad + \frac{f'''(g(t))}{3!}.\frac{[g(t + h) - g(t)]^3}{h} + \dots \end{aligned} \tag{51}$$

Now we calculate the limit when $h \rightarrow 0$ of each one of the terms in the previous equality,

$$\lim_{h \rightarrow 0} \left(\frac{f(g(t+h)) - f(g(t))}{h} \right) = \frac{d [f(g(t))]}{dt},$$

$$\lim_{h \rightarrow 0} \left(f'(g(t)) \cdot \frac{g(t+h) - g(t)}{h} \right) = f'(g(t)) \cdot g'(t),$$

$$\begin{aligned} & \lim_{h \rightarrow 0} \left(\frac{f''(g(t))}{2} \cdot \frac{[g(t+h) - g(t)]^2}{h} \right) \\ &= \lim_{h \rightarrow 0} \left(\frac{f''(g(t))}{2} \cdot \frac{g(t+h) - g(t)}{h} \cdot (g(t+h) - g(t)) \right) \\ &= \frac{f''(g(t))}{2} \cdot g'(t) \cdot 0 = 0 \end{aligned} \tag{52}$$

because g is differentiable (thus, it is continuous).

We also have

$$\lim_{h \rightarrow 0} \left(\frac{f'''(g(t))}{3!} \cdot \frac{[g(t+h) - g(t)]^3}{h} \right) = 0$$

and we obtain the well know chain rule (49).

Note that a crucial assumption in the proof of (52) is that g is differentiable. Brownian motion is nowhere differentiable, so if we need to prove a similar result to be applied in the context of stochastic processes we should expect to obtain some differences.

Indeed, if we take $g(t) = B(t)$ in (51) we obtain

$$\begin{aligned}\frac{f(B(t+h)) - f(B(t))}{h} &= f'(B(t)). \frac{B(t+h) - B(t)}{h} \\ &\quad + \frac{f''(B(t))}{2} \cdot \frac{[B(t+h) - B(t)]^2}{h} \\ &\quad + \frac{f'''(B(t))}{3!} \cdot \frac{[B(t+h) - B(t)]^3}{h} + \dots\end{aligned}\tag{53}$$

Again taking the limit of each of the terms when $h \rightarrow 0$ we get

$$\lim_{h \rightarrow 0} \left(\frac{f(B(t+h)) - f(B(t))}{h} \right) = \frac{d [f(B(t))]}{dt}.$$

We know that B is nowhere differentiable, so the limit when $h \rightarrow 0$ of $\frac{B(t+h) - B(t)}{h}$ does not exist.

However, we know that

$$B(t + h) - B(t) \sim \mathcal{N}(0, h)$$

which implies that

$$\text{Var}(B(t + h) - B(t)) = E[(B(t + h) - B(t))^2] = h$$

which suggest that we could have the approximation

$$(B(t + h) - B(t))^2 \approx h$$

which might imply that

$$\lim_{h \rightarrow 0} \frac{(B(t + h) - B(t))^2}{h} = 1 \quad (54)$$

and that for $k \geq 3$

$$\lim_{h \rightarrow 0} \frac{(B(t + h) - B(t))^k}{h} = \lim_{h \rightarrow 0} \frac{h^{\frac{k}{2}}}{h} = 0. \quad (55)$$

Indeed, it can be proved that (54) and (55) are true. Therefore (formally), we obtain

$$\frac{d [f(B(t))]}{dt} = f'(B(t)) \cdot \frac{dB(t)}{dt} + \frac{f''(B(t))}{2}$$

and multiplying by dt and integrating from 0 to T we get

$$\int_0^T d [f(B(t))] = \int_0^T f'(B(t)) dB(t) + \frac{1}{2} \int_0^T f''(B(t)) dt.$$

Since

$$\int_0^T d [f(B(t))] = f(B(T)) - f(B(0)) = f(B(T)) - f(0)$$

we motivated the following result which is known as Itô's formula.

Theorem

If $f \in C^2(\mathbb{R})$, then

$$f(B(T)) - f(0) = \int_0^T f'(B(t))dB(t) + \frac{1}{2} \int_0^T f''(B(t))dt \quad (56)$$

or in an equivalent way using notation with differentials,

$$df(B(t)) = f'(B(t))dB(t) + \frac{1}{2}f''(B(t))dt. \quad (57)$$

Note that the first integral in the right-hand side of (56) is an Itô integral, while the second integral is a Riemann integral.

Example

Let $f(x) = x$, which implies that $f'(x) = 1$ and $f''(x) = 0$. Itô's formula allows to conclude that

$$B(T) = \int_0^T 1 dB(t).$$

Example

Let $f(x) = x^2$, so that $f'(x) = 2x$ and $f''(x) = 2$. Applying Itô's formula we obtain

$$(B(T))^2 = \int_0^T 2B(t)dB(t) + \frac{1}{2} \int_0^T 2dt = 2 \int_0^T B(t)dB(t) + T$$

which implies that

$$\int_0^T B(t)dB(t) = \frac{(B(T))^2 - T}{2}.$$

Example

Let $f(x) = x^3$, which implies that $f'(x) = 3x^2$ and $f''(x) = 6x$.

Itô's formula implies that

$$\begin{aligned}(B(T))^3 &= \int_0^T 3(B(t))^2 dB(t) + \frac{1}{2} \int_0^T 6B(t)dt \\&= 3 \int_0^T (B(t))^2 dB(t) + 3 \int_0^T B(t)dt.\end{aligned}$$

Therefore,

$$\int_0^T (B(t))^2 dB(t) = \frac{(B(T))^3}{3} - \int_0^T B(t)dt.$$

The Itô formula can be extended to functions of two variables $f(t, x)$.

Theorem

If $f(t, x)$ has one continuous derivative in the first component and two continuous derivatives in the second component, then

$$\begin{aligned} f(T, B(T)) - f(0, 0) &= \int_0^T \frac{\partial}{\partial x} f(t, B(t)) dB(t) \\ &\quad + \int_0^T \left[\frac{1}{2} \frac{\partial^2}{\partial x^2} f(t, B(t)) + \frac{\partial}{\partial t} f(t, B(t)) \right] dt \end{aligned}$$

or in an equivalent way using notation with differentials,

$$df(t, B(t)) = \frac{\partial}{\partial x} f(t, B(t)) dB(t) + \left[\frac{1}{2} \frac{\partial^2}{\partial x^2} f(t, B(t)) + \frac{\partial}{\partial t} f(t, B(t)) \right] dt.$$

Example

Let $f(t, x) = tx^2$, which implies that

$$\frac{\partial}{\partial x} f(t, x) = 2xt, \quad \frac{\partial^2}{\partial x^2} f(t, x) = 2t \text{ and } \frac{\partial}{\partial t} f(t, x) = x^2.$$

Itô's formula for functions of two variables implies that

$$\begin{aligned} T(B(T))^2 &= \int_0^T 2tB(t)dB(t) + \frac{1}{2} \int_0^T 2tdt + \int_0^T (B(t))^2 dt \\ &= 2 \int_0^T tB(t)dB(t) + \frac{T^2}{2} + \int_0^T (B(t))^2 dt. \end{aligned}$$

Therefore,

$$\int_0^T 2tB(t)dB(t) = \frac{1}{2} \left(T(B(T))^2 - \frac{T^2}{2} - \int_0^T (B(t))^2 dt \right).$$

Example

The asset price can be modeled by the stochastic differential equation

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t), \quad t \geq 0, \quad (58)$$

and it is straightforward to verify that a solution of this SDE is given by the geometric Brownian motion which is the stochastic process

$$S(t) = S(0) \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B(t) \right). \quad (59)$$

Example

Indeed, applying Itô's formula, taking

$f(t, x) = S(0) \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma x \right)$, which implies that

$$S(t) = f(t, B(t)) \tag{60}$$

and noting that

$$\frac{\partial}{\partial x} f(t, x) = \sigma f(t, x), \quad \frac{\partial^2}{\partial x^2} f(t, x) = \sigma^2 f(t, x)$$

and

$$\frac{\partial}{\partial t} f(t, x) = \left(\mu - \frac{1}{2} \sigma^2 \right) f(t, x)$$

Example

we get

$$\begin{aligned} df(t, B(t)) &= \frac{\partial}{\partial x} f(t, B(t)) dB(t) \\ &\quad + \left[\frac{1}{2} \frac{\partial^2}{\partial x^2} f(t, B(t)) + \frac{\partial}{\partial t} f(t, B(t)) \right] dt \\ &= \sigma f(t, B(t)) dB(t) + \left[\frac{1}{2} \sigma^2 f(t, B(t)) \right. \\ &\quad \left. + \left(\mu - \frac{1}{2} \sigma^2 \right) f(t, B(t)) \right] dt \\ &= \sigma f(t, B(t)) dB(t) + \mu f(t, B(t)) dt, \end{aligned}$$

and using (60) we obtain (65).

Itô's formula can be used for deriving the solution of some SDE, as illustrated in Example 16.

However, this approach depends on some knowledge of a guess for the type of solution and choice of convenient function $f(t, x)$.

In practice, this approach using Itô's formula does not allow to determine the solution for most SDE and it is difficult (or even impossible) to find the closed form of their solutions.

In those cases, we should use numerical methods for solving the SDE and this will be the subject of next section.

Numerical methods for stochastic differential equations

Next, we study numerical methods for the solution of SDE. For instance, assume that we wish to approximate the solution of the SDE (48). We start by discretizing the interval $[0, T]$, for instance, by taking equally spaced points,

$$t_0 = 0, \quad t_1 = h, \quad t_2 = 2h, \dots, \quad t_N = Nh = T, \quad \text{where } h = \frac{T}{N},$$

for some $N \in \mathbb{N}$. Assume that we know the initial condition of the solutions of the SDE,

$$X(t_0) = x_0$$

and, for simplicity we will assume that $x_0 \in \mathbb{R}$ is fixed (it is not a random variable). The aim is to determine an approximation for

$$X(t_1), X(t_2), \dots, X(t_N) \tag{61}$$

that will be denoted by

$$X_1, X_2, \dots, X_N.$$

Note that $X(t_i)$, $i = 1, \dots, N$ are random variables and their values depend on the particular realization of the stochastic process that is considered.

By (47), we know that

$$X(t_{i+1}) = x_0 + \int_0^{t_{i+1}} a(s, X(s))ds + \int_0^{t_{i+1}} b(s, X(s))dB(s) \quad (62)$$

and that

$$X(t_i) = x_0 + \int_0^{t_i} a(s, X(s))ds + \int_0^{t_i} b(s, X(s))dB(s) \quad (63)$$

and subtracting equations (62) and (63), we get

$$X(t_{i+1}) = X(t_i) + \int_{t_i}^{t_{i+1}} a(s, X(s))ds + \int_{t_i}^{t_{i+1}} b(s, X(s))dB(s). \quad (64)$$

Depending on the numerical approximation of the integrals in (64), we define different numerical methods.

Euler-Maruyama method

The Euler-Maruyama method is a generalization of Euler's method for ODE to SDE.

The equation (64) is discretized into

$$X_{i+1} = X_i + a(t_i, X_i)h + b(t_i, X_i)\Delta B_i, \quad i = 0, 1, \dots, N-1,$$

where $\Delta B_i = B(t_{i+1}) - B(t_i) = \sqrt{h}Z$, $Z \sim \mathcal{N}(0, 1)$.

Note that, when $b \equiv 0$ we obtain Euler's method for ODE and in that case, we know that the discretization error is $O(h)$.

Note that both the theoretical solution and the numerical approximation given by Euler's method depend on the realization of Brownian motion that is considered.

If a realization of the Brownian motion is given, we call a solution $X(t)$ of the SDE a strong solution. For instance, when we fix a Brownian motion realization, then the solution obtained in (66) is a strong solution of the SDE (65).

If we are allowed to consider different Brownian motions, then a solution of the SDE is called weak solution. In the case of weak solutions, only the distribution of the solution is of interest, not any of its paths.

Assuming that we consider the same realization of Brownian motion for the theoretical solution and for the numerical solution given by Euler's method, we can compare the trajectories $\{X(t_i), i = 0, 1, \dots, N\}$ and $\{X_i, i = 0, 1, \dots, N\}$ and particular, we can calculate $|X(t_N) - X_N^h|$, where we used the notation X_N^h to denote the approximation given by Euler's method obtained for a certain value of h .

Since the error obtained depends on the realization of Brownian motion, it is natural to average among "all" possible paths of the Brownian motion and the absolute error is defined by

$$e(h) = E(|X(t_N) - X_N^h|).$$

Definition

(Strong convergence) We say that X_N^h converges strongly to $X(t_N)$ if

$$\lim_{h \rightarrow 0} E(|X(t_N) - X_N^h|) = 0.$$

Moreover, we say that X_N^h converges strongly to $X(t_N)$ with order $\gamma > 0$ if

$$e(h) = O(h^\gamma).$$

Next, we consider the SDE (65) with theoretical solution given by (66) with the following parameters

$$x_0 = 10, \sigma = 0.3 \text{ and } \mu = 0.6$$

We start taking $h = 0.001$ and calculating a realization of the Brownian motion, from which we obtained the theoretical solution and the numerical solution that are plotted in Figure 20.

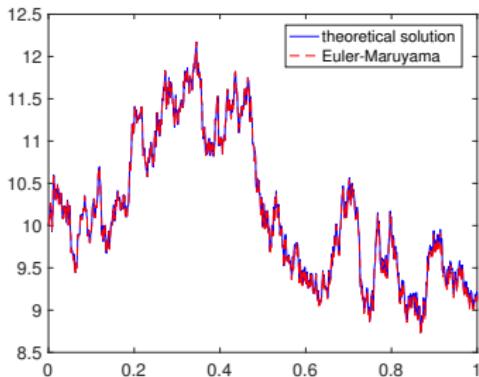


Figure: Plot of the theoretical solution and the numerical solution defined by Euler-Maruyama method for the SDE (65), obtained for the same realization of the Brownian motion.

Then, we took the following values of h ,
0.01, 0.005, 0.0025, 0.00125 and for each one of them we
calculated 1000 realizations of the Brownian motion. Then, in
order to estimate the error, we calculated

$$\hat{e}(h) = \frac{1}{N} \sum_{k=1}^N |X^{(k)}(t_N) - X_N^{(k),h}|,$$

where in each simulation the theoretical solution ($X^{(k)}(t_N)$) and
numerical solution ($X_N^{(k),h}$) were obtained for the same realization
of the Brownian motion. For each value of h we repeated the
procedure three times, where the pseudo-random numbers allowing
to define the realization of the Brownian motion were calculated
starting from different seeds. The results are presented in Table 7.

h	0.01	0.005	0.0025	0.00125
$\hat{e}(h)$	0.099	0.071	0.047	0.032
	0.098	0.067	0.044	0.033
	0.090	0.069	0.046	0.032

Table: Errors obtained with the Euler-Maruyama method applied to the SDE (65), for some values of h .

In order to estimate the strong order of convergence of the Euler-Maruyama method, for each value of h , we average the three values of the errors presented in Table 7 and we obtained the estimate $\gamma \approx 0.53$. Indeed, it can be proven that the Euler-Maruyama method converges strongly with order $\gamma = \frac{1}{2}$.

The strong convergence is appropriate when the trajectory of the solution is of interest. However, sometimes the pointwise approximation of $X(t_N)$ is not relevant and we can be just interested in knowing the probability distribution of $X(t)$. In that case, it is sufficient to know the rate at which the error of the means decrease when the time step size goes to zero.

Definition

(Weak convergence) We say that X_N^h converges weakly to $X(t_N)$ if

$$\lim_{h \rightarrow 0} |E(X(t_N)) - E(X_N^h)| = 0.$$

Moreover, we say that X_N^h converges weakly to $X(t_N)$ with order $\gamma > 0$ if

$$|E(X(t_N)) - E(X_N^h)| = O(h^\gamma).$$

It can be proven that the Euler-Maruyama method converges weakly with order $\gamma = 1$, assuming some regularity of a and b .

The strong order of convergence $\gamma = \frac{1}{2}$ of the Euler-Maruyama method implies the (strong) convergence is very slow and below the order 1 of the Euler's method when applied to ODEs. Using Itô's lemma, we can derive an higher order method for which the strong order of convergence is equal to 1 that using the same notation that was used for the Euler-Maruyama is given by

$$X_{i+1} = X_i + a(t_i, X_i)h + b(t_i, X_i)\Delta B_i + \frac{1}{2}b(t_i, X_i)b'(t_i, X_i)[(\Delta B_i)^2 - h], \\ i = 0, 1, \dots, N - 1,$$

where $b' = \frac{\partial b}{\partial X}$. This method is known as Milstein method and has both strong and weak orders of convergence equal to 1.

Figure 21 shows the plots of the theoretical solution and the numerical solution defined by Milstein method for the SDE (65), obtained for the same realization of the Brownian motion.

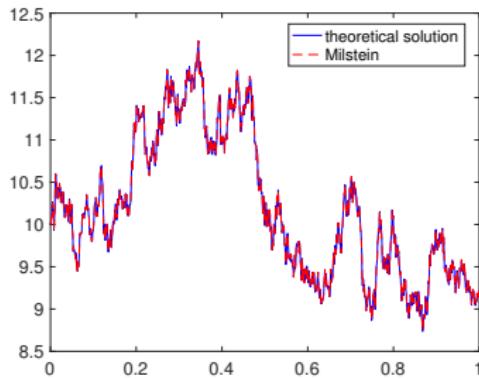


Figure: Plot of the theoretical solution and the numerical solution defined by Milstein method for the SDE (65), obtained for the same realization of the Brownian motion.

Applying the method to the same example that was considered for the Euler-Maruyama method we obtained the results presented in next Table.

h	0.01	0.005	0.0025	0.00125
	0.0382	0.0195	0.0095	0.0049
$\hat{e}(h)$	0.0379	0.0197	0.0093	0.0047
	0.0380	0.0202	0.0096	0.0048

Table: Errors obtained with Milstein method applied to the SDE (65), for some values of h .

Using these results we obtained the estimate for strong order of convergence 1.001 which illustrates the fact the method has strong order of convergence $\gamma = 1$.

- ① Write Matlab routines of Euler-Maruyama and Milstein methods for solving the SDE (taking $\mu = 0.5$ and $\sigma = 0.3$)

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t), \quad 0 < t \leq 1 \quad (65)$$

whose solution is given by the geometric Brownian motion,

$$S(t) = S(0) \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B(t) \right). \quad (66)$$

- ② For a given realization of the Brownian motion, in the same figure plot the exact solution defined by (66) and the numerical solutions obtained by Euler-Maruyama and Milstein methods, taking $h = 0.001$.
- ③ Consider the following values of $h = 0.005 \times \left(\frac{1}{2}\right)^i$, $i = 0, 1, 2, 3$ and for each of the values of h , run 100 000 simulations of Euler-Maruyama and Milstein methods and use them to estimate the order of strong convergence and order of weak convergence of both methods.

Monte Carlo method for pricing a European option

In this section we present a numerical method for pricing an option coupling numerical methods for solving stochastic differential equations and Monte Carlo method.

We assume that we have to price a European option for which the underlying asset price is modeled by a certain stochastic differential equation. Let us denote by $V(S, t)$ the value of the option for a given market price S and instant of time t .

Monte Carlo method for pricing a European option

It can be proven that

$$V(S_0, 0) = e^{-rT} E(V(S_T, T)),$$

which means that we can calculate an approximation for $V(S_0, 0)$ for a given S_0 , provided we can estimate $E(V(S_T, T))$.

We can estimate an expected value by using Monte Carlo method. In practice, for P independent realizations of the Brownian motion we calculate the corresponding numerical solution of the stochastic differential equation obtained for a certain value of h - for instance, obtained through Milstein method - and obtain

$$\bar{S}_1 := X_N^{(1),h}, \bar{S}_2 := X_N^{(2),h}, \dots, \bar{S}_P := X_N^{(P),h}.$$

Monte Carlo method for pricing a European option

Note that we know that at maturity, V must coincide with the payoff defined through function $\phi(S)$, which means that $V(S_T, T) = \phi(S_T)$. Therefore, we consider the following estimate

$$E(V(S_T, T)) \approx \frac{1}{P} \sum_{k=1}^P \phi(\bar{S}_k),$$

and, thus,

$$V(S_0, 0) \approx \frac{e^{-rT}}{P} \sum_{k=1}^P \phi(\bar{S}_k). \quad (67)$$

Monte Carlo method for pricing a European option

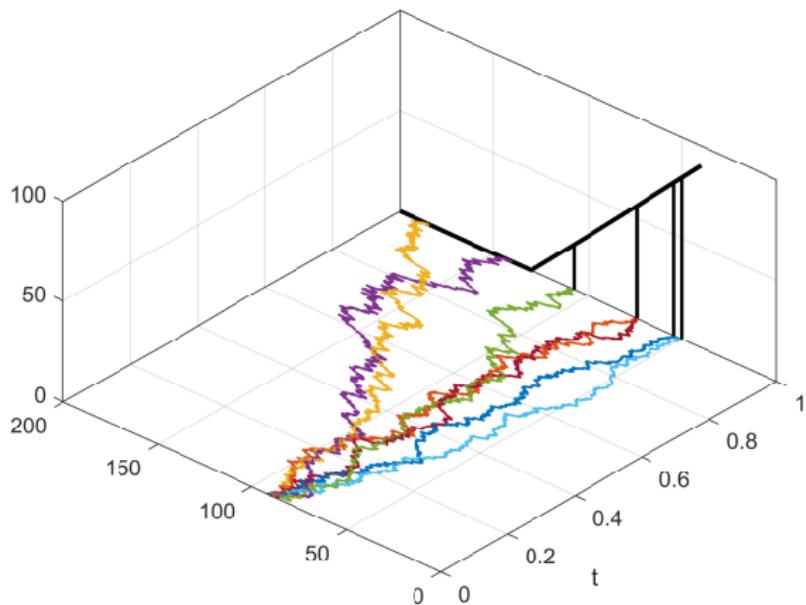


Figure: Plot of seven realizations of the solution of the stochastic differential equation, together with the plot of the payoff at the points \bar{S}_i , $i = 1, 2, \dots, 7$.

Monte Carlo method for pricing a European option

To illustrate the application of this procedure, we consider the stochastic differential equation considered in Example 16,

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t), \quad 0 \leq t \leq T,$$

for which the solution is the geometric Brownian motion

$$S(t) = S(0) \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B(t) \right).$$

Monte Carlo method for pricing a European option

This stochastic differential equation corresponds to the Black-Scholes equation and to test the new approach we consider the case of a put with the following parameters $S_0 = 90$, $K = 130$, $\sigma = 0.3$, $r = 0.5$ and $T = 1$ for which we have
 $V(S_0, 0) = 5.44488\dots$

The approximation given by (67), by taking $P = 1000000$ realizations of the solution of the stochastic differential equation was to $V(S_0, 0) \approx 5.445496$.

The same procedure could be applied to more complex stochastic differential equations for which the exact solution is unknown.