FRE6083: Quantitative Methods in Finance
Lecture notes

Agnès Tourin

June 2020

Contents

Chapter 1

Module 1: Sequences and sums of random variables

This week we review sequences of random variables and sums of random variables in probability theory because this is part of the basic material that you need to know in order to be able to understand the models in Finance and Insurance that we will present in this course. A good reference for this chapter are the textbooks, or else the book by Jean Jacod and Philip Protter, Probability Essentials, Universitext, Second Printing, Second Edition, 2004, Springer.

1.1 Sequence of random variables

We are given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where Ω denotes the *state space*, $\mathcal{A}$, the *space of events*, is a σ-Algebra, and $\mathbb{P}$ is a probability measure.

We recall that a random variable is a deterministic function of an outcome $\omega \in \Omega$:

$$X : \omega \in \Omega \to X(\omega) \in \mathbb{R}$$

We consider in this section a sequence $(X_n)_{n \in \mathbb{N}}$ of random variables. The random variable depends on the index n, where n typically represents the time.

1.1.1 First example of sequence of random variables: The Bernouilli distribution

The Bernouilli distribution is our first example. We first describe the underlying experiment performed and measured on the probability space $(\Omega, \mathbb{P})$, called a *Bernouilli trial*: We roll a die infinitely many times and we define *success* as rolling a six. The outcomes of these trials are mutually independent. To each Bernouilli trial is associated a Bernouilli random variable X_k defined by

$$X_k = \begin{cases} 1 & \text{if } k\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

The probability mass function of X_k is given by

$$p(x) = p^x(1-p)^{1-x}, \text{ for } x \in \{0, 1\}, p = 1/6.$$

We also compute its expectation

$$\mathbb{E}[X_k] = 0 \cdot (1-p) + 1 \cdot p = p.$$

1.2 Sums of random variables

We can define for instance the sum

$$Y_n = \Sigma_{k=1}^n X_k,$$

where X_k is the Bernouilli random variable defined earlier. Then $(Y_n)_{n\in\mathbb{N}}$ is itself a sequence of random variables. It describes the number of successes after n trials. The distribution of Y_n is Binomial with parameter n and p, that is, its probability mass is given by

$$p(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x(1-p)^{n-x}, \text{ for } x \in \{0, 1, 2, 3, \cdots, n\}.$$

1.2.1 Expectation of a sum of random variables

In general, the expectation of a sum $Y_n = \Sigma_{k=1}^n X_k$ of random variables is given by

$$\mathbb{E}[Y_n] = \Sigma_{k=1}^n \mathbb{E}[X_k].$$

In the case of the Bernouilli trials, we obtain

$$\mathbb{E}[Y_n] = \Sigma_{k=1}^n \mathbb{E}[X_k] = n \ \cdot p.$$

We also recall the general property:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

1.2.2 Covariance of a sum of random variables

Let X, Y, Z be 3 random variables. Then, we have

$$cov(X, Y + Z) = cov(X, Y) + cov(X, Z).$$

We now generalize to sums of n and m random variables:

$$cov(\Sigma_{i=1}^n X_i, \Sigma_{i=1}^m Y_i) = \Sigma_{i=1}^n \Sigma_{j=1}^m cov(X_i, Y_j).$$

1.2.3 Variance of a sum of random variables

We now compute the variance of a sum of random variable by using the previous result

$$\begin{aligned} var(\Sigma_{i=1}^n X_i) &= cov(\Sigma_{i=1}^n X_i, \Sigma_{j=1}^n X_j) \\ &= \Sigma_{i=1}^n \Sigma_{j=1}^n cov(X_i, X_j) \\ &= \Sigma_{i=1}^n var(X_i) + 2\Sigma_{i=1}^n \Sigma_{j<i} cov(X_i, X_j) \end{aligned}$$

Furthermore, if $X_1, X_2, \cdots, X_n$ are pairwise independent random variables, then

$$var(\Sigma_{i=1}^n X_i) = \Sigma_{i=1}^n var(X_i).$$

We apply this result to the Binomial variable $Y_n = \Sigma_{k=1}^n = X_k$ where X_k are the Bernouilli variables. Since

$$var(X_k) = \mathbb{E}[X_k^2] - (\mathbb{E}[X_k])^2 = p - p^2 = p(1-p),$$

and the Bernouilli variables are independent, we have

$$\begin{aligned} var(Y_n) &= \Sigma_{k=1}^n var(X_k) \\ &= n \cdot p(1-p). \end{aligned}$$

We recall a few additional properties satisfied by the variance and covariance:

- $\mathrm{var}[X] = 0$ if and only if X is deterministic.
- For every pair of constants (a, b),
  $\mathrm{var}[aX + bY] = a^2\mathrm{var}[X] + b^2\mathrm{var}[Y] + 2ab\mathrm{cov}(X, Y)$.
- Cauchy-Schwarz inequality:

  $$|cov(X, Y)|^2 \leq \mathrm{var}[X]\mathrm{var}[Y],$$

  which can also be written as

  $$|\mathbb{E}[XY]|^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

  Consequently, if either X or Y is deterministic, then $\mathrm{cov}(X, Y) = 0$.

- The Cauchy-Schwarz inequality is an equality if and only if X is a multiple of Y or Y is a multiple of X.
- If X and Y are independent, then $\mathrm{cov}(X, Y) = 0$. Note that the converse is not true.

We also recall the definition of the correlation coefficient which measures linear dependence

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}[X]\mathrm{var}[Y]}}.$$

It has the following properties:

- By Cauchy-Schwarz, $-1 \leq \rho_{X,Y} \leq 1$.
- $\rho_{X,Y} = 0$ if and only if $\mathrm{cov}(X, Y) = 0$. We say that X and Y are uncorrelated. Note that this does not imply independence of X and Y!
- $\rho_{X,Y} = \pm 1$ if and only if X and Y are multiple of one another.

1.2.4 Example of financial application to the return and variance of a portfolio

Consider a portfolio of N assets. The asset relative returns over a given time period are denoted by $X_1, X_2, \cdots, X_N$.

Next, we consider the expected returns and variances of all the assets returns as well as their pairwise linear correlations:

$$\begin{aligned}
&\mathbb{E}[X_1], \cdots, \mathbb{E}[X_N], \\
&\mathrm{var}[X_1], \cdots, \mathrm{var}[X_N], \\
&\rho_{X_i, X_j}, \text{ for } i, j = 1 \cdots N.
\end{aligned}$$

We also introduce $\pi_1, \pi_2, \cdots, \pi_N$, the portfolio weights, which represent the fraction of wealth invested in each asset. The total return of the portfolio is then given by

$$R = \pi_1 X_1 + \pi_2 X_2 + \cdots + \pi_N X_N.$$

We compute the expected return of the portfolio

$$\mathbb{E}[R] = \pi_1 \mathbb{E}[X_1] + \cdots + \pi_N \mathbb{E}[X_N],$$

and the variance of the portfolio

$$\mathrm{var}[R] = \mathrm{var}[\pi_1 X_1 + \pi_2 X_2 + \cdots + \pi_N X_N] =$$
$$\sum_{i=1}^{N} \pi_i^2 \mathrm{var}[X_i] + \sum_i \sum_{j \neq i} \pi_i \pi_j \mathrm{cov}(X_i, X_j).$$

1.2.5 Distribution of a sum of random variables

It is not always easy to determine the distribution of a sum of independent random variables in closed form, even if the distribution of each variable is known explicitly.

Let us take the example of 2 independent random variables X and Y.

We denote by F_{X+Y} the distribution of $X+Y$, by F_X, F_Y the respective distributions of X, Y and by f_Y the density of Y. By definition, we have

$$\begin{aligned} F_{X+Y}(z) &= \mathbb{P}[X + Y \leq z] \\ &= \mathbb{P}[X \leq z - Y] \\ &= \int_{-\infty}^{+\infty} \Big(\int_{-\infty}^{z-y} f_X(x)dx \Big) f_Y(y)dy \\ &= \int_{-\infty}^{+\infty} F_X(z-y) f_Y(y)dy. \end{aligned}$$

Hence the distribution of $X+Y$ is the convolution of the distributions F_X and F_Y.

This result can be generalized by induction to compute the distribution of the sum of n independent and identically distributed random variables. If we denote by F the common distribution of each variable, the distribution of the sum is the n-fold convolution of F with itself.

More precisely, let F_{n-1}, f_{n-1} be respectively the distribution and density of $X_1 + \cdots + X_{n-1}$, let F_n be the distribution of $X_1 + \cdots + X_n$. Then we have

$$F_n(x) = \int_{-\infty}^{+\infty} F(x-y) f_{n-1}(y) dy.$$

We can compute the density of the sum $X_1 + ... + X_n$ by differentiating

$$f_n(x) = \int_{-\infty}^{+\infty} f(x-y) f_{n-1}(y) dy.$$

Exercise : Let $X_1, .., X_n$ be i.i.d random variables uniformly distributed on the interval $[0, 1]$. Compute $F_n(1) = \mathbb{P}[\sum_{i=1}^n X_i \leq 1]$.

Hint : Find a recursive relationship on $F_n(x)$ for $x \in [0, 1]$, and show that $F_n(x) = \frac{x^n}{n!}$.

1.2.6 Another example of sum: the sample mean

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables. Then

$$Y_n = \frac{1}{n} \Sigma_{i=1}^n X_i$$

is called the sample mean. If the variables have a common mean μ and variance σ^2, we have the results

1. $\mathbb{E}[Y_n] = \mu$.

   To see this, we compute

$$\mathbb{E}[Y_n] = \frac{1}{n} \Sigma_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} n \cdot \mu = \mu.$$

2. $var[Y_n] = \frac{\sigma^2}{n}$.

   To see this,

$$\begin{aligned} var[Y_n] &= \frac{1}{n^2} var(\Sigma_{i=1}^n X_i) \\ &= \frac{1}{n^2} \Sigma_{i=1}^n var[X_i] \\ &= \frac{\sigma^2}{n} \end{aligned}$$

3. $cov(Y_n, X_i - Y_n) = 0$, for $i = 1, \cdots, n$.

   To see this, we compute

$$\begin{aligned} cov(Y_n, X_i - Y_n) &= cov(Y_n, X_i) - cov(Y_n, Y_n) \\ &= \frac{1}{n} cov(X_i + \Sigma_{j \neq i} X_j, X_i) - var(Y_n) \\ &= \frac{1}{n} cov(X_i, X_i) + \frac{1}{n} \Sigma_{j \neq i} cov(X_j, X_i) - \frac{\sigma^2}{n} \\ &= \frac{\sigma^2}{n} + 0 - \frac{\sigma^2}{n} \\ &= 0 \end{aligned}$$

1.2.7 Next example: the sample variance

Let again $X_1, \cdots, X_n$ be n independent and identically distributed random variables with common mean μ and common variance σ^2. We denote the sample variance

$$S_n^2 = \frac{1}{n} \Sigma_{i=1}^n (X_i - Y_n)^2.$$

We have the following result

$$\mathbb{E}[S_n^2] = \frac{n-1}{n} \sigma^2.$$

To see this, we compute

$$\begin{aligned}
\mathbb{E}[S_n^2] &= \frac{1}{n}\mathbb{E}[\Sigma_{i=1}^n (X_i - Y_n)^2] \\
&= \frac{1}{n}\mathbb{E}[\Sigma_{i=1}^n X_i^2] - \frac{2}{n}\mathbb{E}[\Sigma_{i=1}^n X_i Y_n] + \frac{1}{n}\mathbb{E}[\Sigma_{i=1}^n Y_n^2] \\
&= \frac{1}{n}\Sigma_{i=1}^n \mathbb{E}[X_i^2] - \frac{2}{n}\Sigma_{i=1}^n \mathbb{E}[Y_n X_i] + \frac{1}{n} n\mathbb{E}[Y_n^2] \\
&= \frac{1}{n}\Sigma_{i=1}^n (var[X_i] + (\mathbb{E}[X_i])^2) - \frac{2}{n^2}\Sigma_{i=1}^n \mathbb{E}[X_i \Sigma_{j=1}^n X_j] + var[Y_n] + (\mathbb{E}[Y_n])^2 \\
&= \frac{1}{n} n(\sigma^2 + \mu^2) - \frac{2}{n^2}(\Sigma_{i=1}^n \mathbb{E}[X_i^2] + \Sigma_{i=1}^n \Sigma_{j \neq i} \mathbb{E}[X_i X_j]) + \frac{\sigma^2}{n} + \mu^2 \\
&= \frac{1}{n} n(\sigma^2 + \mu^2) - \frac{2}{n^2}(n\sigma^2 + n\mu^2 + n(n-1)\mu^2) + \frac{\sigma^2}{n} + \mu^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

The sample variance always underestimates the true variance in average; it is a *biased* estimator. This is because of the error in estimating the mean of the sample, since in general $Y_n \neq \mu$.
It is standard for statisticians to define the sample variance as $S_n^2 = \frac{1}{n-1}\Sigma_{i=1}^n (X_i - Y_n)^2$. It is easy to check that this definition gives an *unbiased* estimator.

1.2.8 The distribution of the sample mean and variance in the normal case

We now assume that the variables $X_1, \cdots, X_n$ are independent and have a normal distribution with common mean μ and common variance σ^2. The following facts hold (we refer to the book by Ross in the list of references in the syllabus for the proofs).

1. Y_n, the sample mean is also distributed normally, with mean μ and variance $\frac{\sigma^2}{n}$.

2. Y_n is independent of S_n^2.

3. S_n^2 has a chi-square distribution with $n-1$ degrees of freedom.

1.2.9 Another aspect of the Bernouilli trials: the Poisson paradigm

This example is also an opportunity to review the moment generating function and to gain experience manipulating sums of random variables.

We consider the Bernouilli trials again but this time, the probability of success at the kth trial is set to p_k and we assume that all the p_k, for $k = 1, \cdots, n$ are small. We still define the number of successes as $Y_n = \Sigma_{k=1}^n X_k$.

Now, we compute the moment generating function of X_k

$$\Phi(t) = \mathbb{E}[e^{tX_k}] = p_k e^t + 1 - p_k = 1 + p_k(e^t - 1).$$

Furthermore, for p_k small, $p_k(e^t - 1)$ is also small and we can use the linear approximation $e^x \approx 1 + x$ to write

$$\Phi(t) \approx exp(p_k(e^t - 1)).$$

We use this result to compute the moment generating function of the variable Y_n

$$\begin{aligned}
\mathbb{E}[e^{tY_n}] &= \mathbb{E}[e^{t\Sigma_{k=1}^n X_k}] \\
&= \mathbb{E}[\Pi_{k=1}^n e^{tX_k}] \\
&= \Pi_{k=1}^n \mathbb{E}[e^{tX_k}] \\
&\approx \Pi_{k=1}^n exp(p_k(e^t - 1)) \\
&= exp\{\Sigma_{k=1}^n p_k(e^t - 1)\}
\end{aligned}$$

and this shows that, when all the p_k are small, Y_n has approximately a Poisson distribution with mean $\Sigma_{k=1}^n p_k$.

To see this, we recall the moment generating function of a Poisson distribution

$$\mathbb{E}[e^{tX}] = e^{\lambda(e^t - 1)},$$

where λ is the parameter of the Poisson distribution.

1.2.10 Another example: the elementary arithmetic random walk

We start with the description of an experiment: one tosses a coin infinitely many times and either gets a head with probability p or a tail with probability $1-p$. The tosses are independent. If the coin is fair, then we have $p = 1 - p = 0.5$. We define the sequence of independent random variables

$$X_k = \begin{cases} +1 & \text{if a head is obtained at the } k\text{th trial} \\ -1 & \text{if a tail is obtained at the } k\text{th trial} \end{cases}$$

Next, we define the random walk by setting

$$Y_0 = 0, Y_n = \Sigma_{k=1}^{n} X_k.$$

We can easily derive the following results

1. $\mathbb{E}[X_k] = p - (1-p) = 2p - 1.$
   In addition, in the case of a fair coin, i.e. $p = 0.5$, $\mathbb{E}[X_k] = 0$.
2. $\mathbb{E}[Y_n | Y_0 = 0] = \Sigma_{k=1}^{n} \mathbb{E}[X_k] = n(2p-1).$
3. $var[X_k] = \mathbb{E}[X_k^2] - (\mathbb{E}[X_k])^2 = 1 - (2p-1)^2 = 4p(1-p).$
4. $var[Y_n] = \Sigma_{k=1}^{n} var[X_k] = 4pn(1-p)$

This is one of the most important examples for the financial applications and we will come back to the random walk many times throughout this course.

1.2.11 Generalization

One can generalize the previous example by allowing different distributions for the increments. One can write the model as

$$Y_{n+1} = Y_n + \epsilon_{n+1},$$

where the random variables $\epsilon_1, \epsilon_2, \cdots \epsilon_{n+1}, \cdots$ are independent and identically distributed.

Note that, by using successive iterations, one can rewrite this model as

$$Y_{n+1} = Y_0 + \sum_{i=1}^{n+1} \epsilon_i,$$

where Y_0 denotes the initial condition, at time $n = 0$.

We will see later that this model constitutes a particular case of the well-known Auto Regressive model (abbreviated AR), in Econometric.

The elementary example of the previous section corresponds to the choice

$$\epsilon_{n+1} = \begin{cases} -1 & \text{with probability } 1-p \\ +1 & \text{with probability } p \end{cases}$$

As we saw earlier, in this case, the residual ϵ_{n+1} has a Bernouilli distribution and, given an initial condition Y_0, the variable Y_{n+1} has a binomial distribution with parameters $n+1$ and p.

Alternately, we can allow the residual to be normally distributed, i.e.

$$\epsilon_{n+1} \sim N(0, \sigma^2),$$

where σ is a positive constant.

Given Y_0, we can then deduce the distribution of Y_{n+1}:

$$Y_{n+1} \sim N(Y_0, (n+1)\sigma^2).$$

Note that its variance depends on the time n and increases linearly with n. We will come back on this fact later in these notes.

A drawback of using the normal distribution in financial applications is the thinness of its tails, which makes extreme events appear very unlikely. In risk management, one cares about extreme event and wants them to be represented in the models. For this purpose, distributions with fatter tails may be preferred.

For instance, we can consider the heavy tail random walk associated with the density function of the Pareto distribution

$$f(x) = \begin{cases} \frac{\alpha}{2}\frac{1}{|x|^{\alpha+1}}, & \text{for } |x| \geq 1 \\ 0, & \text{otherwise} \end{cases},$$

with $0 < \alpha < 2$.

The normal distribution has exponentially fast decreasing tails whereas the above distribution has tails that decrease with a power law. Another density that decreases at the same rate as the one above above when $\alpha = 1$ is the Cauchy density function which gives rise to the *Cauchy random walk*

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

The downside of distribution with thick tails is the lack of well defined moments. In particular, the Cauchy distribution does not have a well defined mean and variance.

The above Cauchy distribution can be generalized to the so-called Student's t-distribution with density function

$$g(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\ \Gamma(\frac{\nu}{2})}(1+\frac{x^2}{\nu})^{-\frac{\nu+1}{2}},\ \text{for } x \in \mathbb{R},$$

where $\nu > 0$ represents the degrees of freedom and $\Gamma(.)$ is the well-known gamma function. The median and mode are 0 but the mean is only defined (and equal to 0) when $\nu > 1$. Finally, the case $\nu = 1$ corresponds to the Cauchy distribution.

1.2.12 A financial application

We consider a stock price, at time t, X_t and the associated log return $Y_t = \log X_{t+1} - \log X_t$. We can model its evolution over time by using a random walk process, i.e.

$$Y_{t+1} = Y_t + \epsilon_{t+1},$$

where the random variables ϵ_{t+1} are independent and identically distributed. An essential feature of this model is that the noise term is evaluated at time $t+1$. Suppose one observes the stock price as time passes or equivalently, the noise term, and that, at time t, Y_t is known, the above equation attempts to predict the log return at time $t+1$. The noise term is not know yet and remains uncertain until time $t+1$.

If we assume that the noise term is normally distributed,the likelihood of extreme price swings will be underestimated whereas a distribution with fatter tails can capture these unlikely events. In practice, the normal distribution may be appropriate at low frequency as stock data will look more and more *normal* as the frequency decreases. However, intraday data typically has fatter tails. For instance, the tails of the distribution of the S&P 500 index futures log returns, sampled at a 5 minute frequency, do not decrease exponentially and behave more like a power low. It is worth noting that tails are notoriously hard to estimate accurately.

1.2.13 Characteristic function of the sum of 2 independent random variables

First, we recall the definition of the characteristic function of a real-valued random variable X

$$\phi_X(u) = \mathbb{E}[e^{iuX}], \text{ for all } u \in \mathbb{R}.$$

Note, that the characteristic function of X always exists and that we also have

$$\phi_X(u) = \int e^{iux} dF_X(x),$$

where F_X denotes the cumulative distribution function of X, or, when X has a density function f_X,

$$\phi_X(u) = \int e^{iux} f_X(x) dx.$$

Now, let X, Y be independent real-valued random variables. Then the characteristic function ϕ_{X+Y} is given by

$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u).$$

1.3 A financial application of sums: an aggregate loss model in Insurance

N is the given number of losses. Here we assume at first that N is known. We will present later in this course a more elaborate model where N itself is random. Furthermore, each loss is an unknown amount X. Usually X is called the *severity*.

The total loss is given by

$$S_N = \Sigma_{n=1}^N X_n,$$

where the variables X_n are independent and identically distributed with common mean μ and common standard deviation σ. This is a collective risk model.

We can easily compute the expected loss

$$\begin{aligned}
\mathbb{E}[S_N] &= \mathbb{E}[\sum_{n=1}^N X_n] \\
&= \sum_{n=1}^N \mathbb{E}[X_n] \\
&= N\mu,
\end{aligned}$$

as well as the variance of the loss

$$\begin{aligned}
\mathrm{var}[S_N] &= \mathbb{E}[(\sum_{n=1}^N X_n)^2] - (\mathbb{E}[S_N])^2 \\
&= \mathbb{E}[\sum_{n=1}^N X_n^2 + \sum_i \sum_{j\neq i} X_i X_j] - N^2\mu^2 \\
&= \sum_{n=1}^N \mathbb{E}[X_n^2] + N(N-1)\mu^2 - N^2\mu^2 \\
&= N\sigma^2 + N\mu^2 + N(N-1)\mu^2 - N^2\mu^2 \\
&= N\sigma^2.
\end{aligned}$$

1.4 Exercises

Problem 1: In this exercise, we consider a continuous random variable X which represents the profit and loss of a portfolio and define the Value at Risk at level q for this portfolio over a fixed time frame as

$$VaR_q(X) = x_q,$$

where x_q denotes the q-quantile of the distribution of X.

Next, denoting by F the cumulative distribution function of the loss, we simply define the q-quantile as

$$x_q = F^{-1}(q),$$

where F^{-1} denotes the inverse function of F. Note that, in the case of a discrete probability model, the q-quantile must be defined differently as F is no longer invertible.

Furthermore, VaR_q can be interpreted as follows: there is only a $q\%$ probability of incurring a loss of at least $-VaR_q$ during the time period.

1. First of all, we assume the distribution of X standard normal and we take $q = 5\%$. Compute VaR_q under these assumptions. Answer: -1.6449.

2. Secondly, we assume that X is Cauchy distributed. We take $q = 5\%$. Compute VaR_q in under these assumptions. Hint: compute and invert the Cauchy cumulative distribution function before computing VaR. Answer: -6.314.

3. Compare the above results and conclude.

Chapter 2

Module 2: Convergence concepts, Law of large numbers, Central Limit Theorem, Markov sequences, the Martingale property

For the concepts of convergence, the Law of large numbers and the central limit Theorem, we draw heavily from the book: "Probability Essentials" by Jean Jacod and P. Protter.

2.1 Convergence of random variables

A random variable is a function of the outcome $\omega \in \Omega$. Therefore the simplest notion of convergence to consider is the so-called *pointwise* convergence, i.e. the sequence X_n converges to the limit X as $n \to +\infty$ if and only if

$$\text{For all } \omega \in \Omega, X_n(\omega) \to X(\omega) \text{ as } n \to +\infty.$$

Convergence for all ω is actually almost never used in probability because it typically does not work (see for instance the example in the book by

Protter and Jacod) and needs to be replaced by a less strict concept of convergence, convergence for *almost all* ω.

For instance, we consider the experiment of the infinite coin toss. We assume that the coin has a probability p of falling on heads and define the variable

$$X_n = \begin{cases} 1 & \text{if a head is obtained at the } n\text{th toss} \\ 0 & \text{if a tail is obtained at the } n\text{th toss} \end{cases}$$

Next, we study the limit

$$\lim_{n \to +\infty} \frac{X_1(\omega) + X_2(\omega) + \cdots X_n(\omega)}{n}.$$

When the outcome of the coin toss is an infinite sequence of Tails, i.e. $\omega = (T, T, \cdots, T)$, the above limit is equal to 0.

However, when it is an infinite sequence of Heads, $\omega = (H, H, \cdots, H)$, the above limit is equal to 1.

Consequently, the above sequence does not converge *pointwise*. As we will see later, as an application of the strong law of large numbers, it does converge for *almost all* ω to p.

2.1.1 Almost sure convergence

This notion of convergence is called *almost sure* convergence and abbreviated *a.s.* or alternately *with probability 1.* We state the definition below

Definition:

$X_n \to X$ as $n \to +\infty$ *almost surely* if and only if

$$\mathbb{P}\{\omega : \lim_{n \to +\infty} X_n(\omega) = X(\omega)\} = 1.$$

This can be rephrased as

$X_n \to X$ as $n \to +\infty$ *almost surely* if and only if

$$\text{the set } N = \{\omega : \lim_{n\to+\infty} X_n(\omega) \neq X(\omega)\} \text{ has probability } 0.$$

2.1.2 Other notions of convergence

We also need some weaker notions of convergence that can be used when the a.s. convergence fails. We present here briefly the following concepts of convergence: *in probability, in mean squares* and *in distribution.* In these notes, we only glide over these concepts and if you wish, you will find a more rigorous and thorough presentation in the book by Jacod and Protter.

- Convergence *in probability (i.p.)*:

  A sequence of random variables X_n converges *in probability* to X as $n \to +\infty$ if and only if

  $$\lim_{n\to+\infty} \mathbb{P}\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\} = 0, \text{ for every } \epsilon > 0.$$

- Convergence *in mean squares (m.s.):*

  A sequence of random variables X_n converges *in mean squares* to X as $n \to +\infty$ if and only if

  If $\mathbb{E}\{|X_n|^2\} < +\infty$ for all n, $\mathbb{E}\{|X|^2\} < +\infty$ and

  $$lim_{n\to+\infty}\mathbb{E}\{|X_n - X|^2\} = 0.$$

- Convergence *in distribution*:

  This notion of convergence is the weakest of all and is known as *weak convergence*. Roughly speaking, it is not the sequence of random variables X_n that converges here but rather the sequence of its cumulative distribution functions F_n. In reality, this concept of convergence is a little more subtle and richer mathematically than what we describe here and we refer to the book by Jacod and Protter for a precise definition. Here, we denote by F the distribution function obtained at the limit. We denote by C, the set of points of continuity of F, i.e., the set containing all the points x, such that

$F(x-) = F(x)$ and we assume that the set C is *dense* in $\mathbb{R}$, i.e. , every real number can be approximated by a sequence of points of continuity of F. Now, we state our definition in this slightly restrictive setting:

The sequence of random variables X_n converges *in distribution* to X as $n \to +\infty$ if and only if

$$\lim_{n \to +\infty} F_n(x) = F(x) \text{ at each point of continuity of } F.$$

2.1.3 Relations between the different notions of convergence

Finally, we summarize in the diagram below how these notions of conference relate to each other.

$$\begin{array}{ccc} \boxed{\text{a.s.}} & & \\ \Downarrow & & \\ \boxed{\text{in prob.}} & \Rightarrow & \boxed{\text{in dis.}} \\ \Uparrow & & \\ \boxed{\text{m.s.}} & & \end{array}$$

It is important to note that none of the arrows above can be reversed in general; for example, it is not hard to find a sequence that converges in distribution but does not converge in mean square.

2.2 Law of large numbers and Central Limit Theorem

Now, we are going to apply these concepts of convergence to some fundamental and very useful results, namely the Law of large numbers and the Central Limit Theorem that have many applications in Quantitative

Finance. As you will see, the Central Limit Theorem explains the prevalence of the normal distribution in many practical applications.

Let $X_1, X_2, ...$ be an infinite sequence of independent and identically distributed random variables (i.i.d. r.v.s in short) and let $S_n = X_1 + X_2 + ... + X_n$. We have the following convergence results. Let $\mu = \mathbb{E}[X_i] < +\infty$ and $\sigma^2 = var[X_i] < +\infty$. We begin by stating the *Strong Law of larges numbers*

Strong Law of large numbers

$$\lim_{n \to +\infty} \frac{S_n}{n} = \mu \text{ a.s and in mean squares.}$$

Central limit Theorem: If in addition, $\sigma^2 > 0$, then,

as $n \to +\infty$, $\frac{S_n - n\mu}{\sqrt{n}\sigma}$ converges in distribution to a standard normal distribution.

2.2.1 Elementary Examples

1. Going back to the infinite coin toss example, by the strong law of large numbers, the limit

$$\lim_{n \to +\infty} \frac{X_1(\omega) + X_2(\omega) + \cdots X_n(\omega)}{n}$$

converges to p almost surely.

2. The central limit Theorem may be applied in the context of the Bernouilli trials to the sequence of Bernouilli random variables $X_1, X_2, ...$ and the number of successes after n trials, S_n. Furthermore, as we saw last week, the expectation of each Bernouilli variable X_k is $\mu = p$ and its variance is

$$Var[X_k] = \mathbb{E}[X_k^2] - (\mathbb{E}[X_k])^2 = p \times 1 - p^2 = p(1-p).$$

The Central limit Theorem tells us in this case that the distribution of the rescaled number of successes

$$\frac{S_n - np}{\sqrt{np(1-p)}}$$

converges to the standard normal distribution. Furthermore, we recall that S_n has a Binomial distribution with parameters n and p.

More generally, you can approximate a Binomial distribution by a normal distribution for large values of n. More precisely, if X is a random variable with a Binomial distribution with parameters n and p, then by the Central Limit Theorem, the quantity

$$\frac{X - np}{\sqrt{np(1-p)}}$$

approaches the standard normal distribution as n approaches $+\infty$. In practice, such an approximation is fairly satisfactory as soon as $np(1-p) \geq 10$.

Here is a numerical application which is drawn from the book by S. Ross, *Introduction to Probability Models*, 11th edition, Academic Press, 2014: take $n = 40, p = 0.5$ and compute $\mathbb{P}[X = 20]$ both exactly and approximately by using the Central Limit Theorem.

Answer:

On the one hand, the exact result is given by

$$\mathbb{P}[X = 20] = \begin{pmatrix} 40 \\ 20 \end{pmatrix} 0.5^{40} = 0.1268.$$

On the other hand, we compute an approximation

$$\begin{aligned}
\mathbb{P}[X = 20] &= \mathbb{P}[19.5 < X < 20.5] \\
&= \mathbb{P}[\frac{19.5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}] \\
&= \mathbb{P}[-0.16 < \frac{X - 20}{\sqrt{10}} < 0.16] \\
&\approx \Phi(0.16) - \Phi(-0.16) \\
&= 2\Phi(0.16) - 1 \approx 0.1272
\end{aligned}$$

where Φ denotes the standard normal cumulative distribution function.

3. An elementary example:

   Let $X_1, X_2, \cdots X_{10}$ be independent random variables with a common uniform distribution over the interval $(0, 1)$. Estimate

$$\mathbb{P}[\Sigma_{i=1}^{10} X_i \leq 6].$$

   Answer: First of all, we have $\mathbb{E}[X_i] = 0.5$ and $var[X_i] = \frac{1}{12}$. By the Central Limit Theorem,

$$\begin{aligned}\mathbb{P}[\Sigma_{i=1}^{10} X_i \leq 6] &= \mathbb{P}[\frac{\Sigma_{i=1}^{10} X_i - 5}{\sqrt{10/12}} \leq \frac{6-5}{\sqrt{10/12}}] \\ &\approx \Phi(1.1) \\ &= 0.8643\end{aligned}$$

4. An application to Statistics:

   let $X_1, \cdots, X_n$ be n i.i.d. random variables and consider the following estimator of their common mean μ:

$$\hat{\theta}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

   By the strong law of large number, the estimator $\hat{\theta}_n$ is *consistent*, i.e

$$\hat{\theta}_n \to \mu \text{ a.s. as } n \to +\infty.$$

   Note that $\hat{\theta}_n$ is also *unbiased*, that is,

$$\mathbb{E}[\hat{\theta}_n] = \mu.$$

5. Adapted from *Heard on the street*

   You are offered two games: in the first game, you roll a die once and you are paid 1 million dollars times the number you obtain on the upturned face of the die. In the second game, you roll a die one million times and for each roll, you are paid 1 dollar times the number of dots on the upturned face of the die. You are *risk averse.* Which game do you prefer?

Answer First of all, the expectation of the gain for each game is 3.5 millions dollars.

To see this, denote by X the gain in dollars for the first game and compute its expectation

$$\mathbb{E}[X] = 10^6 \times 1 \times \frac{1}{6} + 10^6 \times 2 \times \frac{1}{6} + \cdots + 10^6 \times 6 \times \frac{1}{6} = 3.5 \text{ Millions.}$$

Next, denote by Y the gain in dollars for the second game. Then,

$$\mathbb{E}[Y] = 10^6 \times (1 \times 1 \times \frac{1}{6} + 1 \times 2 \times \frac{1}{6} + \cdots + 1 \times 6 \times \frac{1}{6}) = 3.5 \text{ Millions.}$$

Finally, the law of large numbers tells you that, with the second game, your actual payoff will be much closer to the expected payoff than in the first game.

Another way of looking at it is by saying that, since you are risk averse, you prefer the game with the lowest variance. It turns out that the variance of the first game is $1,000,000$ times bigger than the variance of the second game. So, you definitely pick the second game over the first game if you are risk averse.

Finally, one could also look at the problem using Jensen's inequality. Letting ξ_i denote the value each independent role of the dice, and choosing utility function $U(x) = \log(x/10^6)$, Jensen's inequality helps us to see that Y is better than X

$$U(Y) = \log\left(\frac{1}{10^6}\sum_{i=1}^{10^6}\xi_i\right) > \frac{1}{10^6}\sum_{i=1}^{10^6}\log\left(\xi_i\right) = \frac{1}{10^6}\sum_{i=1}^{10^6}U\left(10^6\xi_i\right) \ ,$$

and with law of large numbers we can see further that

$$U(Y) \approx U(\mathbb{E}X) > \mathbb{E}U(X) \ .$$

2.2.2 Application to historical estimations

Consider a sequence $X_1, X_2, \cdots, X_i, \cdots$ of i.i.d random variables with a finite and positive variance. We denote by F their common cumulative distribution function. Next, we define

$$Y_i(x) = \mathbb{I}_{\{X_i \leq x\}},$$

where $\mathbb{I}_{\{X_i \le x\}}$ is the indicator function of the set $\{X_i \le x\}$. It can be shown that

1. the variables $Y_i(x)$ are i.i.d.

2. the common mean of $Y_i(x)$ is $F(x)$.

3. the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i(x)$$

   satisfies

$$\lim_{n \to +\infty} F_n(x) = F(x) \text{ a.s } , \forall x.$$

4. the common variance of $Y_i(x)$ is given by $F(x)(1 - F(x))$.

5. the distribution of the error for the approximation of F by F_n as $n \to +\infty$, is normal and centered. The rate of convergence is $\frac{\sqrt{F(x)(1-F(x))}}{\sqrt{n}}$.

The so-called Glivenko-Cantelli Theorem gives us a stronger result. It tells us that the convergence is uniform in the variable x, i.e.

$$\sup_x |F_n(x) - F(x)| \to 0 \text{ a.s.} \quad \text{as } n \to +\infty.$$

Note that there are roughly three approaches to estimations:

1. Historical estimations (non-parametric): this method is appropriate when the number of samples is large and the expert knowledge is absent.

2. Parametric (maximum likelihood or generalized method of moments for instance): appropriate in the presence of expert knowledge and a relatively small number of samples.

3. Baysian approach: appropriate when the number of samples is very small (and in principle when you can also incorporate substantial expert knowledge). This part is considered a branch of machine learning and is not covered in FRE-GY 6083.

2.2.3 Application to Monte Carlo simulations

Consider the integral

$$a = \int_0^1 f(x)dx.$$

It can be interpreted as the following expectation

$$\mathbb{E}[f(U)],$$

where U is a random variable that is uniformly distributed on the interval $(0,1)$. Consider next a sequence $U_1, U_2, \cdots, U_n$ of independent and uniformly distributed random variables in the interval $(0,1)$.

1. It can be shown by using the strong law of large numbers that
$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} f(U_i) = a, \text{ almost surely.}$$
2. Once can also determine the distribution of the error
$$\frac{1}{n}\sum_{i=1}^{n} f(U_i) - a$$
by applying the Central Limit Theorem.

So, in principle, one can approximate a single integral by using Monte-Carlo simulations. However, approximating a single integral by using the Trapezoidal rule would be more precise than the Monte Carlo simulations technique. The Monte Carlo simulations method becomes competitive as the number of dimensions increases and it is often used in practice for the approximation of multiple integrals when the number of dimensions is high.

2.3 Markov sequences

Some random sequences are said to have *no memory.* Roughly speaking, this means that the future distribution depends only on the current state and not on the whole history. We start with an easy example before giving the formal definition of a Markov sequence.

2.3.1 The Markov property of the elementary arithmetic random walk

The general idea is that the symmetric random walk seen in the previous class does not have a memory since at each coin toss, you restart *from scratch* (each coin toss is independent from the previous one). In other words, the future depends only on the present, not the past. The future value of the random walk is only determined by its present value and the outcomes of the next coin tosses. It does not depend on the whole path that the random walk took, starting at X_0. It is actually possible to construct a different type of random walk which incorporates short-term memory. I will present an example later, after we study Markov Chains.

2.3.2 Definition of a Markov sequence

More generally, we suppose that part of the history of a random sequence is known. Then, we consider the future distribution conditionally to this past history. It turns out that, the future distribution depends exclusively on the most recent value in the history.

A random sequence $(X_n)_{n\in\mathbb{N}}$ is said to be Markovian if for all $m < n$, for all x in the state space,

$$\mathbb{P}[X_n \leq x | \{X_k, k \leq m\}] = \mathbb{P}[X_n \leq x | X_m].$$

A Markov sequence for which all X_n take values in the same space which is discrete is called a Markov Chain; $X_n = 1, 2, ..., N$ or $X_n \in \mathbb{N}$, $\forall n$. These Markov chains are described by their transition matrix P, where P_{ij} represents the probability of moving to state j, given that we are currently at state i: $P_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$. Usually Markov Chains are assumed to be time-homogeneous so that the latter quantity does not depend on n. We will review all these notions in the next chapter.

2.3.3 A random walk with short memory

On can introduce short memory in a random walk, by constructing it in the following way. At each round, one tosses a coin and obtains either a head with probability p, or a tail with probability $1-p$. If a head is obtained, the random walk will keep going in the same direction as in the previous step (if the random walk just went up by 1 at the previous step, it will keep going up and if it just went down, it will keep going down). Conversely, if a tail is obtained, the random walk will reverse its direction (it will go down if it previously went up and will go up if it went down at the previous step). In this situation, the process $(Y_n)_n$ is not Markovian since it depends on two of its own lags (current position and position at the previous step). One can say that $(Y_n)_n$ is a process with short memory. It can be described by writing the equation

$$Y_{n+1} = Y_n + (Y_n - Y_{n-1})\ \epsilon_{n+1},$$

where the variables ϵ_n are the independent and identically distributed one-step increments defined in the previous chapter, in the context of the elementary random walk, i.e.

$$\epsilon_{n+1} = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } 1-p \end{cases}$$

2.3.4 Financial application

Here is an example of a financial application of a Markovian process: consider a sequence of end-of-trading day stock prices. Then, assuming that this sequence is Markov means that the stock price at the end of the nth trading day depends on the previous end-of-day prices only through the price at the end of day m.

The example presented above is a sequence of random variables, with a discrete state space, that satisfy the Markov property. Such processes are called Markov chains and we will study them next week.

2.4 The discrete-time Martingale Property

2.4.1 The martingale property, example of the symmetric random walk

Clearly, the symmetric random walk we defined last week satisfies the following property

$$\mathbb{E}[X_n | X_0 = 0] = 0.$$

More generally, a sequence of random variables is a discrete-time martingale if for all integers k, n, we have

$$\mathbb{E}[X_{k+n} | \{X_m, m = 0 \cdots k\}] = X_k.$$

The interpretation of this property is as follows: a Martingale is expected to perform consistently on average. It does not have a tendency to rise, nor to fall.

We will see later, that in the celebrated Black-Scholes-Merton model, the present (discounted) value of the stock price and of the hedging portfolio are martingales under the risk-neutral probability measure. Similarly, in portfolio management, the optimal wealth is a martingale.

The problem we have in mind can be expressed as the following optimization problem: we denote by π the portfolio allocation strategy and by $X^\pi(t)$ the value at time t of the portfolio for the investment policy π. Then, we seek the portfolio allocation π maximizing the expected utility derived from terminal wealth, i.e.

$$\max_\pi \mathbb{E}[U(X^\pi(T)) | X(t) = x],$$

where $X(t) = x$ is the initial wealth and U is a concave and increasing utility function. The formulation above is somewhat vague and we will make it more precise later. In particular, we need to include in the condition above the information generated by observing the portfolio's individual asset prices up to time t.

Note that this is not a general definition of a martingale. In particular, we can observe another process $(Y_n)_n$ instead of $(X_n)_n$, and compute the conditional expectation with respect to the process (Y_n) instead of the process X_n itself. This would lead to the definition. If

$$\mathbb{E}[X_{k+n}|\{Y_m, m = 0 \cdots k\}] = Y_k,$$

$(X_n)_n$ is said to be a discrete-time martingale with respect to the process $(Y_n)_n$.

In order to be able to give a more general definition, we would also need to introduce the mathematical concept of filtration to model the flow of information. I won't introduce the concept of filtration in this course and we will use instead the simple definitions that are given above. The concept of filtration is taught in the course on Stochastic Calculus and option pricing.

2.4.2 What if the arithmetic random walk is not symmetric?

Consider now an elementary random walk which is not necessarily symmetric. The probability of getting a Head is $p \in (0,1)$ and the probability of getting a Tail is $q = 1 - p$. This corresponds to flipping a biased coin. We can compute the expectation of a one-step increment of this random walk

$$\mathbb{E}[Y_i] = p \times 1 + (1-p) \times (-1) = 2p - 1.$$

Next, we can calculate the expectation of the random walk itself, assuming that it starts at $X_0 = 0$

$$\mathbb{E}[X_n|X_0 = 0] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = n(2p-1).$$

Note that it is equal to 0 if and only if $p = 0.5$. Hence, $p = 0.5$ is the only value of p for which the random walk is a martingale. For other values of p, the random walk is not a martingale! You will see later that the values $p = 1/2, q = 1/2$ correspond to the so-called *risk-neutral* probability measure.

On the one hand, for $p < 0.5$, the random walk tends to fall and is called a *discrete-time supermartingale*. On the other hand, for $p > 0.5$, the random walk tends to rise and is therefore called a *discrete-time submartingale*.

Note that the arguments above are not proofs since we used the restrictive martingale property instead of the more general definition given at the end of last subsection. However, indeed the above arguments can be made rigorous and these statements are true.

2.4.3 Markov and martingale property

As we saw earlier, the elementary symmetric random walk satisfies both the Markov and the martingale property. We also exhibited a particular process which is Markovian and is not a martingale. The remaining question is whether there exists a martingale that is not Markovian. Consider the process defined by using the recursive equation

$$X_{n+1} = X_n + X_{n-1}\ \epsilon_{n+1},$$

where the random variables $\epsilon_0, \epsilon_1, \cdots \epsilon_{n+1}$ are centered, independent, identically distributed, and X_0, X_1 are both given. On the one hand, the above process is not Markovian because its value X_{n+1} at time $n+1$ depends not only on its value at the previous time X_n but also on X_{n-1}. On the other hand, one can show that it is a discrete-time martingale with respect to itself, i.e.

$$\mathbb{E}[X_{n+1}|\{X_0, X_1, \cdots, X_{n-1}, X_n\}] = X_n + X_{n-1}\mathbb{E}[\epsilon_{n+1}] = X_n.$$

More generally, for any $m < n$

$$\begin{aligned} & \mathbb{E}[X_{n+1}|\{X_0, X_1, \cdots, X_{m-1}, X_m\}] \\ = \ & X_m + X_{m-1}\mathbb{E}[\epsilon_{m+1}] + X_m\mathbb{E}[\epsilon_{m+2}] + \mathbb{E}[X_{m+1}]\mathbb{E}\epsilon_{m+3}] + \cdots + \mathbb{E}[X_{n-1}]\mathbb{E}[\epsilon_{n+1}] \\ = \ & X_m, \end{aligned}$$

which proves the martingale property.

2.4.4 Basic definitions of a discrete-time submartingale and a discrete-time supermartingale

A sequence of random variables is a discrete-time submartingale if for all integers k, n, we have

$$\mathbb{E}[X_{k+n} | \{X_m, m = 0 \cdots k\}] \geq X_k.$$

A sequence of random variables is a discrete-time supermartingale if for all integers k, n, we have

$$\mathbb{E}[X_{k+n} | \{X_m, m = 0 \cdots k\}] \leq X_k.$$

2.5 Some empirical facts about data

Generally, as we will see later in this course, it is actually very difficult to come up with a quantitative model that possesses all the known realistic features of financial data.

Whether the Markov property is satisfied by asset prices is a very important question since most of continuous-time Finance relies on it; unfortunately, it is also a very difficult question to answer.

It is not until the last few years that some authors have started proposing some statistical tests for the Markov property. This material is very complex and cannot be presented in this course. The sparse evidence that is available today suggests that, for most data, the Markov property does not hold.

Clearly, the martingale property is not satisfied by market prices in the sense that they often exhibit a trend. The question is whether the price becomes a martingale, after removing its deterministic trend. Some statistical methods have been proposed to test the martingale assumption and there is some evidence that the martingale property holds, at least locally for some assets prices. Besides, some researchers have studied the formation of price bubbles, which clearly do not behave like martingales and have modeled them with so-called *local martingales*. Roughly, local

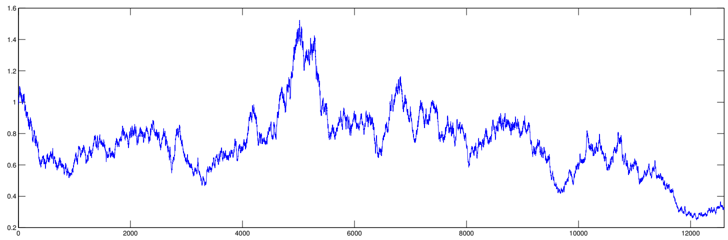


Figure 2.1: The exponential martingale with $\sigma = 0.2$, for a time horizon of 50 years and a daily time step

martingales are martingales on small fixed time intervals but in the long run, they blow up.

We show in the first figure a simulated martingale and in the second figure, a housing price bubble which do not behave like a martingale.

2.6 Exercises

Problem 1 (15 points)

Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with common mean 0 and variance $\sigma^2 < +\infty$. Consider $S_n = \sum_{i=1}^n X_i$. Show that

$$\lim_{n \to +\infty} \mathbb{E}\left[\frac{|S_n|}{\sqrt{n}}\right] = \sqrt{\frac{2}{\pi}}\sigma.$$

Problem 2 (25 points)

We consider a sequence of independent and identically distributed random variables $X_1, X_2, \cdots, X_n, \cdots$ with common mean μ and variance $\sigma^2 > 0$. We also assume that the variables X_i have finite third and fourth moments. Next, we consider the sum

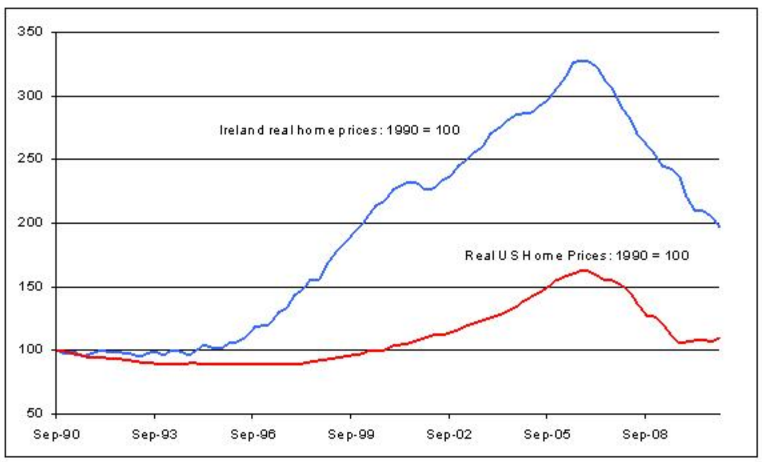


Figure 2.2: This picture is available on the website of the European Central bank, as part of a speech given by Gertrude Tumpel-Gugerell, Member of the Executive Board of the ECB, at alumni event of the Faculty of Economics at University of Vienna, Vienna, 3 May 2011.

$$S_N = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2.$$

1. (4 points) Are the variables $(X_i - \mu)^2$ independent and identically distributed?
2. (4 points) What is their common mean? Their common variance?
3. (5 points) Does S_N have a limit in the almost surely sense? What is it?
4. (4 points) Compute $\mathbb{E}[S_N]$.
5. (8 points) Give the approximate distribution of the error $S_N - \sigma^2$ in terms of N and the moments of X_i. Specify the mean and variance.

Problem 3 (10 points)

Consider the independent and identically distributed random variables $X_1, X_2, \cdots, X_{100}$, with a common uniform distribution on $(0,1)$. Use the Central Limit Theorem to compute an approximation of

$$\mathbb{P}[X_1 + \cdots + X_{50} < X_{51} + \cdots + X_{100}].$$

Problem 4 (10 points)

Consider the independent and identically distributed random variables $X_1, X_2, \cdots, X_{100}$, with common density function

$$f(x) = \frac{1}{x}, \text{ for all } 1 \leq x \leq e.$$

Use the Central Limit Theorem to deduce an approximate density for the product $X_1 X_2 \cdots ...X_{100}$.

Problem 5 (28 points)

Consider a sequence of i.i.d random variables with finite and positive variance. We denote by F their common cumulative distribution function. Next, we define

$$Y_i(x) = \mathbb{I}_{\{X_i \leq x\}},$$

where $\mathbb{I}_{\{X_i \leq x\}}$ is the indicator function of the set $\{X_i \leq x\}$.

1. (5 points) Are the variables $Y_i(x)$ i.i.d? Justify your answer.
2. (4 points) What is the common mean of $Y_i(x)$?
3. (5 points) Consider the empirical distribution function

   $$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i(x).$$

   Show that

   $$\lim_{n \to +\infty} F_n(x) = F(x) \text{ a.s .}$$

4. (4 points) Compute the common variance of $Y_i(x)$.
5. (10 points) Give the distribution of the error for the approximation of F by F_n as $n \to +\infty$. What is the rate of convergence?

Problem 6 (10 points)

Consider $X_1, X_2, \cdots, X_i, \cdots$ a sequence of independent and identically distributed random variables. We define $Y_i = e^{X_i}$.

Show that

$$\left(\prod_{i=1}^{n} Y_i \right)^{\frac{1}{n}}$$

converges to a constant almost surely and give the value of this constant.

Problem 7 (17 points)

Consider a sequence $X_1, X_2, \cdots, X_i, \cdots$ of independent and identically distributed random variables with a continuous distribution with mean 0 and variance σ^2. We define the sum

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

We also define the process

$$M_n = Y_n^2 - n\sigma^2,$$

for all n.

1. (17 points) Prove the martingale property for every pair (n, m) such that $m \leq n$.

$$\mathbb{E}[M_{n+1}|\{X_1, X_2, \cdots, X_m\}] = M_m.$$

Problem 8 (15 points)

Consider $X_1, X_2, \cdots, X_i$, a sequence of independent and Poisson distributed random variables, with mean 1 and variance 1. We know that $\sum_{i=1}^{n} X_i$ is also Poisson distributed with parameter n.

1. (5 points) Show that

$$\mathbb{P}[\sum_{i=1}^{n} X_i \leq n] = \sum_{k=0}^{n} e^{-n}n^k/n!.$$

2. (10 points) Deduce by using the Central Limit Theorem that

$$\sum_{k=0}^{n} e^{-n}n^k/n! = \frac{1}{2}.$$

Problem 9 (25 points) Drawn from *Probability Essentials* by Jacod and Protter

We consider the sequence of independent variables X_j that have a uniform distribution on $(-j, j)$.

1. (5 points) Compute the characteristic function $\phi_{X_j}(u)$ of X_j.
2. (5 points) Compute the characteristic function $\phi_{S_n}(u)$ of the sum $S_n = \sum_{j=1}^{n} X_j$.
3. (5 points) Compute the characteristic function $\phi_{S_n/n^{3/2}}(u)$ of $S_n/n^{3/2}$.
4. (***difficult: 5 points) Show that its limit as $n \to +\infty$ is $e^{-u^2/18}$ by using
$$\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}.$$
5. (5 points) Conclude that the sequence $S_n/n^{3/2}$ converges in distribution as n goes to ∞ to a normal distribution with mean 0 and variance 1/9.

Problem 10 (28 points)

A portfolio of n loans is modeled by using correlated Bernouilli variables as follows. For $i = 1 \cdots N$, we denote by X_i the variable defined for each loan i as
$$X_i = \sqrt{\rho} Y + \sqrt{1-\rho}\epsilon_i,$$
where Y is a factor common to all the loans and has a standard normal distribution, $\rho \in (0,1)$ is the correlation coefficient between two loans, and ϵ_i are independent random variables with a standard normal distribution. Moreover we assume that ϵ_i is independent of Y, for all i. Finally, we define the Bernouilli variable Y_i as
$$Y_i = \mathbb{I}_{\{X_i < x^*\}},$$
where $\mathbb{I}$ is the indicator funtion and x^* is the level below which loan i defaults. In other words, Y_i takes the value 1 if the loan i defaults and a value 0 if it does not.

1. (2 points) Is the distribution of X_i normal? Justify your answer.
2. (2 points) Compute $\mathbb{E}[X_i]$.

3. (3 points) Compute var$[X_i]$.

4. (5 points) Verify that
$$\text{cov}(X_i, X_j) = \rho,$$
for all i, j, such that $i \neq j$.

5. (5 points) Write the probability p_i of default of the ith loan in terms of the standard normal cumulative distribution function and give $\mathbb{E}[Y_i]$. Does p_i depend on i?

6. (3 points) Compute var$[Y_i]$.

7. (5 points) We consider the random variables L representing the number of defaults in the portfolio. Write it in terms of the Bernouilli variables Y_i, for $i = 1 \cdots n$.

8. (3 points) Compute the expected loss.

Problem 11 (14 points)

Consider a sequence of independent and identically distributed random variables $X_1, X_2, \cdots, X_i, \cdots$ such that $X_i \geq 0$, with common mean $\mathbb{E}[X_i] = 1$. Define the sequence $M_n = \prod_{i=1}^{n} X_i$.

1. (4 points) Compute $\mathbb{E}[M_n]$.

2. (6 points) Show that
$$\mathbb{E}[M_n | \{M_k, 1 \leq k \leq n-1\}] = M_{n-1}.$$

3. (4 points) Conclude.

Chapter 3

Module 3: Markov chains

3.1 Markov chains

3.1.1 Introduction

This week, we study Markov chains. This study is motivated by the numerous applications to Insurance and Finance, in particular to credit risk, but also to the modeling of asset prices or equilibrium. It is particularly useful for modeling a volatility coefficient that switches between two levels, a low-volatility level and a high-volatility level. This type of model is called *regime-switching model*. In addition, you will find in the book by Hull, *Options, Futures and other derivatives*, 5th edition, a description of a straightforward application of Markov chains to credit ratings in credit risk. Furthermore, Markov chains can be used to estimate the probability of default of a firm. We will study at the end of this chapter the gambler's ruin problem which constitutes a well-known step in this direction. Most existing models however are slightly too advanced for a basic core course and here, I only present the most basic mathematical concepts and some simple applications because this is not a subject that can be covered thoroughly in one single lecture. You will learn more by reading some of the books quoted in the list of references, in particular, the book by Rick Durrett, *Probability: Theory and Examples, 4th ed., Cambridge University Press, Cambridge, 2010*, or else the book by Ross, *Introduction to Probability Models*, 11th edition, Academic Press, 2014.

Next, we start with a definition of Markov chains.

A Markov chain is a discrete-time Markovian stochastic process. Consider the Markov chain denoted by $\{X_n, n = 0, 1, ...\}$. Here we restrict ourselves to discrete-state Markov chains.

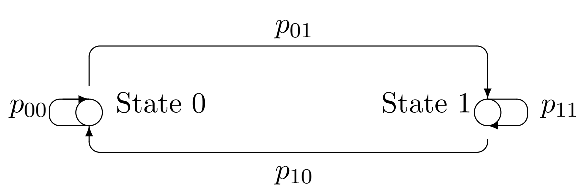
A discrete Markov process is a discrete-time and discrete-state process satisfying the Markov property

$$\mathbb{P}[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0] = \mathbb{P}[X_{n+1} = j | X_n = i].$$

The *transition probability* from state i to state j does not depend on the the path followed by the process from X_0 to X_n, only on the last position X_n.

Note that the random walk is an example of a discrete-time and discrete-state Markov process and hence, it is a Markov chain.

Here is now below a representation of a 2-state Markov chain



3.1.2 Time-homogeneity

A Markov chain X_n is a time-homogeneous Markov chain if its transition probabilities are independent of time,i.e.

$$\mathbb{P}[X_{n+1} = j | X_n = i] = p_{i,j} \text{ for all } n = 0, 1, ...$$

Note that, under the time-homogeneity assumption, this definition of the Markov property is equivalent to the definition seen in the second lecture.

A random walk

The random walk described on the picture is time-homogeneous in the weak sense above: its transition probabilities are independent of time.

$$i \begin{cases} j = i+1 \text{ with proba. } p \\ j = i \text{ with prob. } r \\ j = i-1 \text{ with prob. } q \end{cases} \qquad\qquad p+q+r=1$$

3.1.3 The transition probability matrix

The one-step transition probability matrix P of a time-homogeneous Markov chain is given by

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & p_{0,2} & \cdots \\ p_{1,0} & p_{1,1} & p_{1,2} & \cdots \\ p_{2,0} & p_{2,1} & p_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

with

$$\begin{aligned} p_{i,j} &\geq 0 \text{ for all } i, j \\ \sum_{j=0}^{\infty} p_{i,j} &= 1 \text{ for all } i \end{aligned}$$

The second constraint models the fact the process must be in one and only one state at any given time.

The random walk

We assume that $X_0 = 0$. The random walk's evolution is described by its transition probability matrix P defined by

$$P = \begin{bmatrix} & \ddots & \ddots & \ddots & & & \\ & & q & r & p & & O \\ & O & & q & r & p & \\ & & & & \ddots & \ddots & \ddots \end{bmatrix},$$

with $p + q + r = 1$.

Note: Either the matrix P is of infinite dimension or alternately you may restrict yourselves to a finite number of states $0..N-1$ and impose *boundary conditions* at the edges of the state space 0 and $N-1$.

3.1.4 Transition probability in n steps

We can construct the matrix $P^{(n)}$ of the transition probabilities in n steps $p_{i,j}^{(n)} = \mathbb{P}[X_{m+n} = j | X_m = i]$. It has the same dimensions as the matrix of transition probabilities P. We can obtain $P^{(n)}$ by raising P to the power n, thanks to the Chapman-Kolmogorov equations.

Chapman-Kolmogorov equations

$$p_{i,j}^{(m+n)} = \sum_{k=0}^{\infty} p_{i,k}^{(m)} p_{k,j}^{(n)}, \text{ for all } m, n, i, j \geq 0$$

or in matricial form,

$$P^{(m+n)} = P^{(m)} P^{(n)}.$$

Consequence

$$P^{(n)} = \underbrace{P^{(1)} ... P^{(1)}}_{n \text{ times}}.$$

- Proof of the corollary: Take $m = 1$ and $(n-1)$ for n in the Chapman-Kolmogorov equations.
- Proof of the main result:

$$\begin{aligned} p_{i,j}^{(m+n)} &= \mathbb{P}[X_{m+n} = j | X_0 = i] \\ &= \sum_{k=0}^{\infty} \mathbb{P}[X_{m+n} = j, X_m = k | X_0 = i] \\ &= \sum_{k=0}^{\infty} \mathbb{P}[X_{m+n} = j | X_m = k, X_0 = i] \mathbb{P}[X_m = k | X_0 = i] \\ &= \sum_{k=0}^{\infty} p_{k,j}^{(n)} p_{i,k}^{(m)} = \sum_{k=0}^{\infty} p_{i,k}^{(m)} p_{k,j}^{(n)}. \end{aligned}$$

So we have

$$P^{(m+n)} = P^{(m)} P^{(n)}.$$

3.1.5 First time probability, probability of first return

We now define some useful concepts for the applications, the first time probability and the probability of first return.

- Probability that the chain will move to state j, from $X_0 = i$ for the first time at time n:

$$\rho_{i,j}^{(n)} = \mathbb{P}[X_n = j, X_{n-1} \neq j, ..., X_1 \neq j | X_0 = i] \text{ for all } n \geq 1, i, j$$

- Probability of first return to state i: it is the same as the first time probability for $i = j$:

$$\rho_{i,i}^{(n)} = \mathbb{P}[X_n = i, X_{n-1} \neq i, ..., X_1 \neq i | X_0 = i] \text{ for all } n \geq 1, i$$

- Relation between $p_{i,j}^{(n)}$ and $\rho_{i,j}^{(n)}$:

$$p_{i,j}^{(n)} = \sum_{k=0}^{n} \rho_{i,j}^{(k)} p_{j,j}^{(n-k)}.$$

- How do you compute the first time probability? Use the above formula recursively, assuming that $p_{i,j}^{(n)}$ is known, to compute $\rho_{i,j}^{(n)}$.

3.1.6 More preliminary concepts: accessibility, irreducibility and aperiodicity

- A state j is accessible from the state i if there exists an $n \geq 0$ such that $p_{i,j}^{(n)} > 0$.
- If i is accessible from j and j is accessible from i, we say that the states i and j communicate.
- A Markov chain is said to be irreducible if all the states communicate.
- A Markov chain is said to be aperiodic if for every state i, the greatest common divisor of $\{n | p_{i,i}^{(n)} > 0\}$ is equal to 1.

3.1.7 Ergodic Markov chain

An irreducible Markov chain is said to be ergodic if the limit

$$\pi_j = \lim_{n \to +\infty} p_{i,j}^{(n)}$$

exists, is independent of i, and satisfies $\sum_{j=0}^{+\infty} \pi_j = 1$.

Furthermore, if a finite state Markov chain is irreducible and aperiodic, then it is ergodic.

In general, proving the existence of the limit π_j may be difficult, except in some very particular cases. However, you may be able to determine whether π_j exist by using an iterative algorithm that can be implemented on a computer.

In the case when π_j exist, you can obtain it by solving the system

$$\begin{aligned} \pi &= \pi P \\ \sum_{j=0}^{\infty} \pi_j &= 1 \end{aligned}$$

where $\pi = (\pi_0, \pi_1, ...)$ is a row vector.

Where does the above system come from? Simply apply the Chapman-Kolmogorov equations with $m = 1$:

$$P^{n+1} = P^n P.$$

This is equivalent to

$$p_{i,j}^{n+1} = \sum_{k=0}^{\infty} p_{i,k}^n p_{k,j}.$$

Next, let n tend to $+\infty$ in the above formula which converges to

$$\pi = \pi P.$$

It is worth mentioning that one can also interpret π_j as being the fraction of time spent on average in the long run in the state j and we do not show the proof of this fact here.

If $t_j(N)$ denotes the amount of time spent in the state j during the time periods $1, \cdots, N$, we have that

$$\frac{t_j(N)}{N} \to \pi_j \text{ as } N \to +\infty$$

in the almost surely sense.

You can express the random variable $t_j(N)$ as

$$t_j(N) = \sum_{k \leq N} \mathbb{I}_{\{X_k = j\}},$$

where $\mathbb{I}$ denotes the indicator function.

3.1.8 A few words about initial conditions

A note about the initial distribution of a Markov chain: In many applications, the initial position is *deterministic*. For instance, we supposed earlier that we had $X_0 = 0$.

Sometimes we give instead an initial distribution, i.e. a row vector $\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, ..)$ where

$$\mu_i^{(0)} = \mathbb{P}[X_0 = i].$$

Then the marginal probabilities must be conditioned on the initial state:

$$\mathbb{P}[X_n = j] = \sum_{i=0}^{\infty} \mathbb{P}[X_n = j | X_0 = i]\mathbb{P}[X_0 = i] = \sum_{i=0}^{\infty} p_{i,j}^{(n)} \mathbb{P}[X_0 = i].$$

Or simply rewritten as

$$\mu^{(n)} = \mu^{(0)} P^n$$

A discrete-time Markov chain is fully determined by its initial distribution $\mu^{(0)r}$ and its transition probability matrix P.

Here is now another way of looking at the limit π_j: the limits π_j are often called *stationary* probabilities for the following reason. If I have initially

$$\mathbb{P}[X_0 = j] = \pi_j, \text{ for all} j \geq 0,$$

then

$$\mathbb{P}[X_n = j] = \pi_j, \text{ for all } n, j \geq 0.$$

This fact is easy to prove by induction and we refer to the book by Ross, p. 222, for a proof.

3.1.9 Expectation of a Markov chain

Next, we can compute the expectation of a Markov chain, conditionally to its initial state:

$$\mathbb{E}[X_n | X_0 = i] = \sum_{j=1}^{\infty} j \mathbb{P}(X_n = j | X_0 = i) = \sum_{j=1}^{\infty} j p_{i,j}^{(n)}.$$

3.1.10 A simple example (C. Tapiero)

Two airline companies compete for passengers. The initial number of clients of the first company is $N_0^0 = 500,000$ and the number of clients of the second company is $N_0^1 = 200,000$. In order to gain a bigger share of the market, the second company proposes a price reduction. We assume here that the total number of customer remains constant, equal to $700,000$. What is the expected number of clients of the first and second airlines after a month, 2 months, etc... and in the long run?

This can be modeled using a Markov chain with 2 states:

- State 0: A customer flies with first airline.
- State 1: A customer flies with second airline.
- Time $n = 0, 1, \ldots$ is the number of months.

We denote the transition probabilities by $p_{i,j}$. They represent the probability that a customer flying with company i switches to company j. They are estimated to be

$$\begin{array}{ll} p_{0,0} = 1/6 & p_{0,1} = 5/6 \\ p_{1,0} = 1/3 & p_{1,1} = 2/3 \end{array}.$$

These numbers can be interpreted in the following way: $p_{0,0} = 1/6$ means that 1/6th of the customers who are currently using the first airline will stay loyal and keep using it whereas 5/6th of them will switch to the second airline.

The Markov chain is thus described by the matrix

$$P = \begin{bmatrix} 1/6 & 5/6 \\ 1/3 & 2/3 \end{bmatrix}.$$

First of all, we have

$$\mathbb{P}[X_{n+1} = 0] = p_{0,0}\mathbb{P}[X_n = 0] + p_{1,0}\mathbb{P}[X_n = 1]$$

and

$$\mathbb{P}[X_{n+1} = 1] = p_{0,1}\mathbb{P}[X_n = 0] + p_{1,1}\mathbb{P}[X_n = 1].$$

Let N_n^0, N_n^1 denote respectively the number of clients for the first and second firms. Note that

$$\begin{array}{rcl} N_{n+1}^0/(N_{n+1}^0 + N_{n+1}^1) & = & \mathbb{P}(X_{n+1} = 0) \\ N_{n+1}^1/(N_{n+1}^0 + N_{n+1}^1) & = & \mathbb{P}(X_{n+1} = 1) \end{array}$$

Month	N^0	N^1
0	500,000	200,000
1	150,000	550,000
2	208,000	492,000
3	199,000	501,000
4	200,000	500,000

where

$$N^0_{n+1} + N^1_{n+1} = N^0_n + N^1_n = \ldots = N^0_0 + N^1_0 = 700,000$$

is the total number of customers.

That leads to the equation in matrix form

$$\begin{bmatrix} N^0_{n+1} \\ N^1_{n+1} \end{bmatrix} = \begin{bmatrix} 1/6 & 1/3 \\ 5/6 & 2/3 \end{bmatrix} \begin{bmatrix} N^0_n \\ N^1_n \end{bmatrix}.$$

Then, after n months, we find that

$$\begin{bmatrix} N^0_{n+1} \\ N^1_{n+1} \end{bmatrix} = (P^T)^{n+1} \begin{bmatrix} N^0_0 \\ N^1_0 \end{bmatrix}.$$

The numbers shown in the previous table are rounded. One observes that the pair of sequences converges to the steady state $N^0 = 200,000$, $N^1 = 500,000$.

3.1.11 An example in Insurance (C. Tapiero):

- Every given year, an insured is in one of 2 states $(0, 1)$.
- State 0: no claim is made, state 1: a claim is made.
- $\mathbb{P}[X_{n+1} = 1 | X_n = 0] = \alpha$, $\mathbb{P}[X_{n+1} = 0 | X_n = 1] = \beta$ are given.

  Then the corresponding matrix of transition probabilities is

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}.$$

We assume that $\alpha, \beta \in (0,1)$ One can show by induction that

$$P^{(n)} = \frac{1}{\alpha+\beta}\begin{bmatrix}\beta & \alpha \\ \beta & \alpha\end{bmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{bmatrix}\alpha & -\alpha \\ -\beta & \beta\end{bmatrix}.$$

As $n \to +\infty$, $P^{(n)}$ converges to its limit

$$\overline{P} = \frac{1}{\alpha+\beta}\begin{bmatrix}\beta & \alpha \\ \beta & \alpha\end{bmatrix}.$$

So here

$$\pi = \begin{bmatrix}\frac{\beta}{\alpha+\beta} \\ \frac{\alpha}{\alpha+\beta}\end{bmatrix}.$$

Note that, when $\alpha = 1$ and $\beta = 1$, the chain is no longer ergodic. The transition probability matrix becomes

$$P = \begin{bmatrix}0 & 1 \\ 1 & 0\end{bmatrix}.$$

We can easily see that

$$P^{2k} = \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}, P^{2k+1} = \begin{bmatrix}0 & 1 \\ 1 & 0\end{bmatrix}.$$

This chain is periodic and its long term behavior depends on the starting point and whether the number of time steps is odd or even.

3.1.12 A example with three states

We consider a finite state Markov chain with state space $\{0,1,2\}$ and the transition probability matrix

$$P = \begin{bmatrix}1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & 1/2 & 0\end{bmatrix}.$$

We compute $P^2, P^3, P^4, \cdots$. We find

$$P^n = \begin{bmatrix}1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & 1/2 & 0\end{bmatrix} = P.$$

We conclude that this Markov chain is non-ergodic. In other words, the evolution of the chain depends on the initial state.

3.2 The gambler's ruin problem: an absorption problem

- A player at each round of the game wins one dollar with probability p and loses one dollar with probability $q = 1 - p$. Suppose that he has an initial capital equal to i dollars. Then he repeats independent rounds of the game until either he goes bankrupt (his capital reaches 0) or his wealth reaches k dollars. We denote by X_n his total wealth. $\{X_n, n = 0, 1, ...\}$ is a Markov chain. It satisfies $X_0 = i$.
- The state space of X_n is $S_X = \{0, 1, 2, ..., k\}$.
- The states 0 and k are *absorbing* (boundary conditions at the edges of the state space 0 and k):

$$p_{0,0} = p_{k,k} = 1$$

- For all the other states, $i = 1, ..., k - 1$, we have

$$p_{i,i+1} = p = 1 - p_{i,i-1}.$$

- This is a random walk with absorbing conditions 0 and k.

  The transition probability matrix for this random walk is of finite dimension. Here is this $(k + 1) \times (k + 1)$ matrix:

$$P = \begin{bmatrix} 1 & 0 & & & & & & \\ q & 0 & p & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & q & 0 & p & & O & \\ & O & & q & 0 & p & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & q & 0 & p \\ & & & & & & 0 & 1 \end{bmatrix}.$$

- In addition, for $p = q = \frac{1}{2}$, this type of Markov chain satisfies the martingale property:

  $$\mathbb{E}[X_{n+1}|X_n = i] = i, \text{ for all } i.$$

  Indeed,

  $$\mathbb{E}[X_{n+1}|X_n = i] = p(i+1) + q(i-1) = (p+q)i + p - q = i + p - q = i.$$

  Note that this property no longer holds when $p \neq q$.

Next, we want to compute the probability of eventual ruin, given the initial wealth i, denoted by r_i. We first define the probability of ruin in n steps $r_i^{(n)}$ as

$$r_i^{(n)} = \mathbb{P}[X_n = 0|X_0 = i] = p_{i,0}^{(n)}.$$

To clarify, note that the probability of ruin in n steps is not the probability of hitting bankruptcy in exactly n steps but rather the probability of going bankrupt in either $1, 2, ...$ or at most n steps.

Since, $X_n = 0$ implies that $X_{n+m} = 0$, for all $m > 0$, the probability of eventual ruin is obtained by taking the limit

$$r_i = \lim_{n \to \infty} r_i^{(n)}.$$

We now turn to the calculation of r_i. First of all, we recall that $r_0 = 1$ and $r_k = 0$. Next, we apply the Chapman-Kolmogorov equations with $j = 0, m = 1$

$$p_{i,0}^{(n+1)} = \sum_{l=0}^{+\infty} p_{i,l}^{(1)} p_{l,0}^{(n)}$$

and by taking the limit $n \to +\infty$ in these equations, we obtain the invariance relation

$$r_i = \sum_{l=0}^{+\infty} p_{i,l}^{(1)} r_l.$$

These lead to the system

$$\begin{aligned} r_1 &= pr_2 + q \\ r_i &= qr_{i-1} + pr_{i+1} \text{ for } i = 2, ..., k-2 \\ r_{k-1} &= qr_{k-2}, (\text{ for } i = k-1) \end{aligned}$$

By rewriting the general equation

$$(p+q)r_i = qr_{i-1} + pr_{i+1},$$

and applying it iteratively, together with the first equation $r_1 = pr_2 + q$, one can deduce

$$r_{i+1} - r_i = \frac{q}{p}(r_i - r_{i-1}) = (\frac{q}{p})^i(r_1 - 1), \text{ for all } i = 1 \cdots k-2$$

Now, adding the above equations for $i = 1, ..., j-1$ we obtain

$$r_j - r_1 = (r_1 - 1)\sum_{i=1}^{j-1}(\frac{q}{p})^i, \text{ for all } j = 1 \cdots k-1.$$

When $p = q = \frac{1}{2}$, this yields

$$r_j = 1 + j(r_1 - 1).$$

Note that we obtained this solution for $j = 1, ..., k-1$. We use the last equation $r_{k-1} = \frac{1}{2}r_{k-2}$ together with the above result for $j = k-1$ and $j = k-2$ to compute the value of r_1:

$$r_{k-1} = \frac{1}{2}r_{k-2} = \frac{1}{2}(1 + (k-2)(r_1 - 1)) = 1 + (k-1)(r_1 - 1).$$

We obtain

$$r_1 - 1 = -\frac{1}{k}.$$

Finally, this yields the value of r_1 and the complete solution

$$r_j = 1 - \frac{j}{k}.$$

In the case $p \neq q$, we obtain

$$r_j - r_1 \quad = \quad (r_1 - 1)\left(\frac{q}{p}\right)\left(\frac{1 - \left(\frac{q}{p}\right)^{j-1}}{1 - \frac{q}{p}}\right), \text{ for } j = 2, \cdots k-1. \tag{3.1}$$

We still need to compute r_1 explicitly. To this end, we rewrite (3.1) for $j = k-1$

$$r_{k-1} = r_1 + (r_1 - 1)\left(\frac{q}{p}\right)\left(\frac{1 - \left(\frac{q}{p}\right)^{k-2}}{1 - \frac{q}{p}}\right),$$

as well as the condition

$$r_{k-1} = qr_{k-2},$$

and equate the right hand sides of the two previous equations, i.e.

$$r_1 + (r_1 - 1)\left(\frac{q}{p}\right)\left(\frac{1 - \left(\frac{q}{p}\right)^{k-2}}{1 - \frac{q}{p}}\right) = qr_{k-2}.$$

By using (3.1) again with $j = k - 2$, we get

$$r_1+(r_1-1)\left(\frac{q}{p}\right)\left(\frac{1 - \left(\frac{q}{p}\right)^{k-2}}{1 - \frac{q}{p}}\right) = q\left(r_1 + (r_1 - 1)\left(\frac{q}{p}\right)\left(\frac{1 - \left(\frac{q}{p}\right)^{k-3}}{1 - \frac{q}{p}}\right)\right).$$

Solving the above equation for r_1 yields, after several steps

$$r_1 = \frac{q}{p}\left(\frac{1 - \left(\frac{q}{p}\right)^{k-1}}{1 - \left(\frac{q}{p}\right)^{k}}\right).$$

Note that I used the identity $p + q = 1$ several times in the above calculation.

The above result, coupled with (3.1), gives us the result, after a few simplifications, i.e.

$$r_j = 1 - \frac{1 - (q/p)^j}{1 - (q/p)^k}.$$

Table 3.1: Example of transition probability matrix for the migration of bonds

Rating	Aaa	Aa	A	Baa	Ba	B	C	Default
Aaa	97.763%	2%	0.2%	0.01%	0.01%	0.015%	0.002%	0.000%
Aa	1%	95.862%	3%	0.1%	0.02%	0.015%	0.002%	0.001%
A	0.95%	1.5%	92.215%	5%	0.3%	0.015%	0.01%	0.01%
Baa	0.53%	0.02%	2%	91.4%	4%	0.95%	0.6%	0.5%
Ba	0.1%	0.2%	0.5%	2%	86.2%	7%	3%	1%
B	0.045%	0.078%	0.15%	1%	4%	82.727%	7%	5%
C	0.001%	0.05%	0.15%	0.65%	2%	5%	81.499%	10%

3.3 Credit ratings

Bonds' credit ratings migrate over time from one category to another. The migration process can be described by using the transition probability matrix of a discrete time Markov chain. Rating agencies estimate the coefficients in the matrix by using historical data. So each transition probability represents the probability of migrating from a given rating to another specified rating within the time period. The time period considered typically ranges from one year to 5 years. In order give you an idea of how such a matrix looks like, I present below a simplified example that is not based on historical data for a period of one year. You will find some tables based on real data on the Moody's website.

3.4 Exercises

Problem 1 (21 point)

Consider the Markov chain X with matrix probability matrix

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

1. (4 points) Plot a diagram representing this chain.

2. (4 points) Is this chain irreducible? Justify.

3. (5 points) Is it aperiodic? Justify your answer.

4. (8 points) Is X ergodic? Justify your answer and compute the limit.

Problem 2 (18 points)

We consider the gambler's ruin problem seen in class. The gambler starts with an initial capital equal to i dollars, and stops playing whenever he gets ruined (i.e. his wealth reaches 0 dollar) or gets rich (her wealth reaches k dollars where $k > i$). We denote by P_i the probability that the gambler will eventually get rich (before being ruined). We also assume that the probablility of winning 1 dollar at every round is equal to $p \neq 0.5$.

1. (10 points) Compute P_i by using the techniques in the class notes. Please, show the details of the calculation!

2. (8 points) Compute the limite of P_i as k converges to $+\infty$. Hint: consider separately the two cases $p < 0.5$ and $p > 0.5$.

Problem 3 (21 point)

Consider the Markov chain X with matrix probability matrix

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$$

1. (4 points) Plot a diagram representing this chain.

2. (4 points) Is this chain irreducible? Justify.

3. (5 points) Is it aperiodic? Justify your answer.

4. (8 points) Is X ergodic? Justify your answer and compute the limit.

Problem 4 (25 points)

Consider the gambler's ruin problem and assume that the player starts with an initial wealth of i us dollars. Next, let M_i denote the mean number of games that must be played until the gambler is either ruined or reaches a wealth of $N \geq i$.

1. (10 points) Explain why M_i satisfies

$$M_0 = M_N = 0, M_i = 1 + pM_{i+1} + qM_{i-1}, i = 1, \cdots, N-1.$$

2. (15 points) When $p = 0.5$, show that M_i is given by

$$M_i = i(N-i).$$

Problem 5 (20 points)

The one-step transition probability matrix P of a Markov chain with state space $\{0, 1, 2\}$ is given by

$$P = \begin{bmatrix} 1/4 & 1/4 & 1/2 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{bmatrix}.$$

1. (2 points) Draw a diagram representing this Markov Chain.
2. (2 points) Is this Markov chain irreducible? Justify your answer.
3. (4 points) Compute P^2.
4. (4 points) Deduce from the previous question $\mathbb{E}[X_2|X_0 = 0]$.
5. (4 points) Solve the system $\pi = \pi P$ coupled with $\sum_{i=0}^{2} \pi_i = 1$ to determine the unique candidate for the stationary distribution. Note that you are not asked to show the convergence of the chain to π.
6. (4 points) Suppose the chain starts from 0 at time 0, i.e. $X_0 = 0$. Compute the probability of reaching the state 2 for the first time after taking 3 steps (i.e. at time $n = 3$).

Problem 6 (18 points)

Consider the one-step transition probability matrix P of a Markov chain with state space $\{0, 1\}$

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

1. (4 points) Show that for all $n \geq 1$, $P^{2n+1} = P$ and

$$P^{2n} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

2. (4 points) Is this Markov chain ergodic? Justify your answer.
3. (5 points) Compute $Q_M = \frac{1}{M} \sum_{n=1}^{M} P^n$.
4. (5 points) Determine $\lim_{M \to +\infty} Q_M$. Conclude.

Problem 7 (40 points)

We consider a discrete time and discrete space Markov chain X_n with state space $\{0, 1, 2, \cdots, R\}$ modeling the wealth of a gambler. The gambler starts with an initial capital equal to X_0 dollars at the time $n = 0$. He or she tosses a coin at each round and either gets a head with probability p or gets a tail with probability $q = 1 - p$. Every time the gambler gets a head, he or she earns one dollar and every time he or she gets a tail, he or she loses one dollar. The gambler stops playing whenever he or she is ruined, i.e. his capital X_n goes down to 0 or when his or her capital reaches R, i.e. $X_n = R$.

We suppose here that the gambler's wealth will eventually reach R (before it goes to 0).

1. (5 points) Explain why the probability that she will win the next round, knowing that her wealth will eventually reach R can be expressed as

$$\mathbb{P}[X_{n+1} = i+1 | X_n = i, \lim_{m \to +\infty} X_m = R] = \frac{\mathbb{P}[X_{n+1} = i + 1, \lim_{m \to +\infty} X_m = R | X_n = i]}{\mathbb{P}[\lim_{m \to +\infty} X_m = R | X_n = i]}.$$

2. (4 points) We set $r_i = \lim_{N\to+\infty} p^N_{i,R}$. Rewrite the right-hand-side of the above relation in terms of r_i and r_{i+1}.

3. (4 points) Show also that
$$r_i = pr_{i+1} + qr_{i-1}.$$

4. (5 points) Compute r_i for $p = 1/2$. The answer is
$$\frac{i+1}{2i}.$$

5. (12 points) Compute now probability for $p \neq 1/2$. The answer is
$$\frac{p(1-(q/p)^{i+1})}{1-(q/p)^i}.$$

6. (10 points) We consider the process Y defined by $Y_n = (q/p)^{X_n}$. Show that Y is a X-martingale, in the sense that, for all $p \leq n$,
$$\mathbb{E}[(Y_n|\{X_0, X_1, \cdots, X_p\}] = Y_p.$$

   Note: this question is independent from the previous questions.

   Hint: you can decompose X_n into $X_n = X_p + \sum_{j=p+1}^n Z_j$, where Z_j denotes the one-step increment of the process X.

Problem 8 (26 points)

Consider the Markov chain with states $\{0, 1, \cdots, n\}$ and transition probabilities
$$p_{0,1} = p_{n,n-1} = 1, p_{i,i+1} = p_i = 1 - p_{i,i-1},$$
where $0 < p_i < 1$ are given for all $i = 1, \cdots, n-1$.

1. (10 points) Give the transition probability matrix P.

2. (6 points) Write the system of linear equations satisfied by the candidate stationary probabilities.

3. (10 points) Solve this system to compute the candidate stationary probabilities π_i, for $i = 1, \cdots, n$.

Chapter 4

Module 4: A Markov chain in continuous time: the Poisson process

4.1 Introduction

On can also design models based on Markov chains in continuous-time. Here, we only cover the example of the Poisson process, which is a building block in many financial models.

4.2 The Poisson process

Last but not least, here is another example of a stochastic process with a continuous time space and a discrete state space, namely the Poisson process. In this section, we give the definition of a Poisson process and provide more details. A good reference for this chapter is the book by Durett. Finally, for these notes, we draw heavily from the book by Ross.

Since, the Poisson process is a counting process, we recall the definition of a counting process first.

A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$

represents the total number of events that occur by time t.

For instance, it can be used for modeling the number of insurance claims by time $t \geq 0$.

A counting process enjoys a number of properties that we list below:

- $N(t) \geq 0$
- $N(t)$ is integer valued
- If $s < t$, then $N(s) \leq N(t)$.
- For $s < t, N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$.

We can now give a definition of a Poisson process:

A counting process N is said to be a Poisson process with rate $\lambda > 0$ if

- $N(0) = 0$.
- It has independent increments
- The increment $N(t+s) - N(s)$ is Poisson distributed with mean λt, i.e.

$$\mathbb{P}[N(t+s) - N(s) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \text{ for all } n = 0, 1, \cdots .$$

An immediate consequence of this definition is that N has stationary increments. Furthermore, we know that

$$\mathbb{E}[N(t)] = \mathbb{E}[N(t) - N(0)] = \lambda t$$

and consequently, λ can be seen as the average rate of jumps per unit of time and, in other words, the average number of jumps in the interval $[0, t]$ is equal to λt.

We also know that $\text{var}[N(t)] = \text{var}[N(t) - N(0)] = \lambda t$.

4.2.1 Inter-arrival and waiting time distribution

We denote by T_1, the time of the first event (jump), by T_2, the waiting time between the first and second event, $\cdots$, T_n, the time between the $(n-1)$th and the nth event. The sequence $T_1, T_2, \cdots, T_N$ is called the sequence of inter arrival times.

It is quite easy to show (and we refer to the book by Ross for instance for a proof) that $T_1, T_2, \cdots, T_n$ are independent, identically distributed exponential random variables with mean $\frac{1}{\lambda}$.

There is a fundamental reason why the exponential distribution arises in this situation. The inter-arrival times have no memory, in the sense that, if the process has already waited t units of time after the last event, the probability of waiting an extra T units of time, does no depend on the fact that it has already waited for the time t. this can be written mathematically as

$$\mathbb{P}[T_i > T + t | T_i > t] = \mathbb{P}[T_i > T].$$

Furthermore, the exponential distribution is actually the only distribution that satisfy this *memoryless* property. So the inter-arrival times must have an exponential distribution. This is also connected to the fact that the Poisson process possesses the Markov property.

More generally, all the Markov chains in continuous time enjoy the same property. The time spent in a given state is always distributed exponentially.

4.2.2 The arrival times

Next, we can define the sequence of arrival times $S_1, S_2, \cdots, S_n$ where S_n denotes the arrival time of the nth jump by setting

$$S_n = \Sigma_{i=1}^{n} T_i, n \geq 1.$$

It turns out that S_n has a gamma distribution with parameters n and λ (we refer for instance to the book by Ross for the proof, the key being that $\{S_n > t\} = \{N_t \leq n-1\}$):

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}.$$

We can finally redefine the Poisson process in a way that's more constructive than the earlier definition, by using the arrival times

$$N(t) = \begin{cases} 0 & \text{for } 0 \le t < S_1 \\ 1 & \text{for } S_1 \le t < S_2 \\ 2 & \text{for } S_2 \le t < S_3 \\ \vdots & \end{cases}$$

4.2.3 A simple example

I give below a simple exercise whose answers are straightforward. Let $N(t)$ be the number of insurance claims filed in the interval $[0, t]$ where t is in weeks. We assume that N is a Poisson process with rate $\lambda = 1$ per week.

1. Compute the probability that at least 1 claim will be made during a given week.

   Answer:

   $$\mathbb{P}[N(t+1) - N(t) \ge 1] = 1 - \mathbb{P}[N(1) - N(0) = 0] = 1 - e^{-1}.$$

2. Calculate the probability that no insurance claim is made during 2 given consecutive weeks. Answer:

   $$\mathbb{P}[N(t+2) - N(t) = 0] = \mathbb{P}[N(2) - N(0) = 0] = e^{-2}.$$

3. Compute the probability that exactly 2 claims will be made during a given week, knowing that there were exactly 2 claims in the previous week. Answer:

   $$\mathbb{P}[N(t+1) - N(t) = 2 | N(t) - N(t-1) = 2] = \mathbb{P}[N(1) - N(0) = 2] = e^{-1}/2.$$

4. Compute the expected waiting time between two consecutive claims.
   Answer:

   All of the inter-arrival times are i.i.d. and exponentially distributed, thus

$$\begin{aligned}\mathbb{E}[T_i] &= \int_0^{+\infty} t\,\lambda e^{-\lambda t}dt \\ &= -te^{-\lambda t}|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda t}dt \\ &= \frac{1}{\lambda}\end{aligned}$$

   So on average, one waits 1 week before receiving an insurance claim.

4.2.4 An aggregate loss model in insurance

We model the number of insurance claims by using a Poisson process with parameter λ. We denote by $N(t)$ the number of claims at time t. The loss X_i associated with the *ith* claim is a random variable.

Next, the total discounted accumulated loss at time t is given by

$$L(t) = \sum_{n=1}^{N(t)} e^{-rS_n} X_n,$$

where the variables X_n are independent and identically distributed, S_n is the arrival time of the *nth* claim, and r is a discount factor.

To simplify, we assume that the claim amounts $X_1, X_2, \cdots$ are independent of the claim arrival times $S_1, S_2, \cdots$.

In addition, conditional on $N(t) = n$, the variables $S_1, S_2, \cdots, S_n$ are distributed as the ordered values of n independent uniform random variables in $(0, t)$ (this is a consequence of the definition of the Poisson process). In other words, $S_1, S_2, \cdots, S_n$ can be interpreted as a random permutation of the uniformly distributed over the interval $(0, t)$ random variables $t_1, t_2, \ldots, t_n$, so that

$$\sum_{n=1}^{N(t)} e^{-rS_n} = \sum_{n=1}^{N(t)} e^{-rt_n}$$

Note that you will find a similar model in the book by Ross presented on pages 334-335.

We can now compute the expected loss by time t, $\mathbb{E}[L(t)]$ by using a total probability rule:

$$\begin{aligned}
\mathbb{E}[L(t)] &= \Sigma_{k=0}^{\infty}\mathbb{E}[L(t)|N(t)=k)]\mathbb{P}[N(t)=k] \\
&= \Sigma_{k=0}^{\infty}\mathbb{E}[L(t)|N(t)=k)]exp(-\lambda t)\frac{(\lambda t)^k}{k!} \\
&= \Sigma_{k=0}^{\infty}\mathbb{E}[\Sigma_{i=1}^{N(t)}e^{-rS_i}X_i|N(t)=k)]exp(-\lambda t)\frac{(\lambda t)^k}{k!} \\
&= \Sigma_{k=0}^{\infty}\mathbb{E}[\Sigma_{i=1}^{k}e^{-rS_i}X_i]exp(-\lambda t)\frac{(\lambda t)^k}{k!} \\
&= \Sigma_{k=0}^{\infty}\Sigma_{i=1}^{k}\mathbb{E}[e^{-rS_i}X_i]exp(-\lambda t)\frac{(\lambda t)^k}{k!} \\
&= \Sigma_{k=0}^{\infty}\Sigma_{i=1}^{k}\mathbb{E}[e^{-rS_i}]\mathbb{E}[X_i]exp(-\lambda t)\frac{(\lambda t)^k}{k!} \\
&= \Sigma_{k=0}^{\infty}\Sigma_{i=1}^{k}\mu\mathbb{E}[e^{-rS_i}]exp(-\lambda t)\frac{(\lambda t)^k}{k!} \\
&= \mu\Sigma_{k=0}^{\infty}\mathbb{E}[\Sigma_{i=1}^{k}e^{-rS_i}]exp(-\lambda t)\frac{(\lambda t)^k}{k!} \\
&= \mu\Sigma_{k=0}^{\infty}\mathbb{E}[\Sigma_{i=1}^{k}e^{-rt_i}]exp(-\lambda t)\frac{(\lambda t)^k}{k!} \\
&= \mu\Sigma_{k=0}^{\infty}\Sigma_{i=1}^{k}\mathbb{E}[e^{-rt_i}]exp(-\lambda t)\frac{(\lambda t)^k}{k!}
\end{aligned}$$

where μ is the common mean of the variables X_i.

Since

$$\mathbb{E}[e^{-rt_i}] = \int_0^t \frac{e^{-rx}}{t}dx = \frac{1}{rt}(1-e^{-rt}),$$

we finally have

$$\begin{aligned}
\mathbb{E}[L(t)] &= \mu\Sigma_{k=0}^{\infty}k\cdot\frac{(1-e^{-rt})}{rt}e^{-\lambda t}\frac{(\lambda t)^k}{k!} \\
&= \mu\frac{(1-e^{-rt})}{rt}\Sigma_{k=0}^{\infty}k\cdot e^{-\lambda t}\frac{(\lambda t)^k}{k!} \\
&= \mu\lambda\frac{(1-e^{-rt})}{r}
\end{aligned}$$

because the mean of a Poisson distribution with parameter λt is $\mathbb{E}[N(t)] = \lambda t$.

4.2.5 An arithmetic random walk driven by a Poisson process

We consider a Poisson process N with rate $\lambda > 0$ and define the Poisson random walk

$$Y_{t+1} = Y_t + \sigma\epsilon_{t+1},$$

where

$$\epsilon_{t+1} = N(t+1) - N(t),$$

and σ is a positive constant. As seen earlier, the variables ϵ_{t+1} are independent and identically distributed. They have a Poisson distribution with mean λ and variance λ.

If there is no jump between the time t and $t+1$, the process is constant on the interval $(t, t+1]$. Every time a jump occurs in the interval $(t, t+1)$, the process Y_t jumps by the amount σ. Overall, given Y_0, we have that

$$Y_{t+1} = Y_0 + \sigma\sum_{i=1}^{t+1}\epsilon_i$$

which can be also rewritten as

$$Y_{t+1} = Y_0 + \sigma(N(t+1) - N(0)).$$

4.2.6 Simulations

There are several methods available. For instance, I suggest that you simulate the jump inter-arrival times explicitly, by drawing independent samples from the exponential distributions with mean $1/\lambda$. Then, at each jump time, you just add 1 to the Poisson process.

4.3 Exercises

Problem 1 (16 points)

Consider a continuous nonnegative random variable T with mean μ and variance σ^2 and a Poisson process with rate $\lambda > 0$.

1. (6 points) Compute $\mathrm{var}(N(T))$.
2. (10 points) Compute $\mathrm{cov}(T, N(T))$.

Problem 2 (26 points) drawn from "Applied Probability Models" by Ross

Consider the sequence $Y_1, Y_2, \cdots$ of independent random variables with common mean μ, common variance σ^2, and the Poisson process with rate $\lambda > 0$. We also assume that all the Y_i random variables are independent of $N(t)$, for all t. We define the asset price process

$$X(t) = x_0 \prod_{i=1}^{N(t)} Y_i,$$

where x_0 is a deterministic initial condition.

1. (5 points) Compute $\mathbb{E}[X(t)]$.
2. (6 points) Compute $\mathrm{var}[X(t)]$.
3. (10 points) Is $X(t)$ a martingale with respect to the information generated by observing it? Justify your answer.

4. (***Difficult 5 points) Is there a value of μ that makes X a martingale? Justify your answer.

Problem 3 (15 points)

We consider a Poisson process with rate $\lambda > 0$ and we recall that $\mathbb{E}[N(t)] = \text{var}[(N(t))] = \lambda t$. We also consider a sequence X_1, $X_2 \cdots X_3$ of i.i.d. random variables with common mean μ and common variance σ^2, that are also independent of $N(t)$.

1. (15 points) Compute

$$\text{cov}(N(t), \sum_{i=1}^{N(t)} X_i).$$

Problem 4 (15 points)

Consider three independent Poisson processes $M_1(t), M_2(t), M_3(t)$ with respective rates $\lambda_1, \lambda_2, \lambda_3 > 0$. We also define

$$N_1(t) = M_1(t) + M_2(t), N_2(t) = M_2(t) + M_3(t).$$

1. (10 points) Write the joint probability mass $\mathbb{P}[N_1(t) = n, N_2(t) = m]$ as a sum. Note: do not attempt to compute the sum.
2. (5 points) Compute $\text{cov}(N_1(t), N_2(t))$.

Problem 5 *(difficult) (15 points)**

We consider the following conditional Poisson process. First of all, let L be a continuous random variable and let f be its density function, defined on $(0, +\infty)$. Next, let N be a counting process. Finally, conditional on $L = \lambda > 0$, N is a Poisson process with rate λ.

1. (5 points) Write the probability mass $\mathbb{P}[N(t + s) - N(s) = n]$ as an integral over the interval $(0, +\infty)$.

2. (5 points) Does N have stationary increments? Justify your answer by using the formula derived in the first question.

3. (5 points) Does N have independent increments? Justify your answer in plain english.

Chapter 5

Module 5: The Binomial asset pricing model

We present first the one-period Binomial model before turning to the more general multi-period binomial model. This class draws heavily on the textbook by S. Shreve, *Stochastic Calculus for Finance*, Volume I.

5.1 The one-period Binomial model

In this model, the beginning of the time period is called time 0 and the end is called time 1. At time 0, we have a stock whose price per share is denoted by S_0. We assume that S_0 is a positive number. An experiment is performed: one tosses a coin once and obtains either a Head with probability p or a Tail with probability $1-p$. If a Head is obtained, the price of a share of the stock at time 1 will be $S_1(H)$ whereas if a Tail is obtained, the price of a share of the stock will be $S_1(T)$. We assume that p and $1-p$ are both positive but they are not necessarily equal. In other words, the coin is not necessarily a fair coin. The up and down factors are denoted by u and d respectively

$$u = \frac{S_1(H)}{S_0}, d = \frac{S_1(T)}{S_0}.$$

$$\begin{tikzcd} & S_1(H) = S_0u \\ S_0 \arrow[ur, dash] \arrow[dr, dash] & \\ & S_1(T) = S_0d \end{tikzcd}$$

We assume that $d < u$. We also incorporate in this model a constant short interest rate $r \geq 0$. Often, the rate used in this model is called *risk-free* rate and I will sometimes use this terminology, although this is only a mathematical construction. In practice, for the applications to option pricing, one typically uses a 3-month or 6-month LIBOR rate (London Inter-Bank Offered rate) in a specified currency.

Next, in order to rule out *arbitrage* in this model, we must have the inequalities

$$0 < d < 1 + r < u. \tag{5.1}$$

In order to see this, we first need to define the concept of *arbitrage*. An *arbitrage strategy* is a trading strategy that allows an investor starting with a zero initial wealth at time 0, to make a profit with a nonzero probability, without taking any risk.

We can exhibit such arbitrage opportunities if $d \geq 1 + r$ or $u \leq 1 + r$. For instance, let us consider the case $u > d \geq 1 + r$.

An investor, with initial wealth $X_0 = 0$ could borrow from the money market in order to buy one share of stock. His debt would then be S_0 at time 0. At time one, his debt would be $S_0(1 + r)$ and the stock price would be either S_0u or else S_0d. Consequently, in the worst case scenario, his wealth at time 1 would be

$$S_0d - S_0(1 + r) \geq 0,$$

and in the case when a head is obtained, his net worth at time 1 would be

$$S_0u - S_0(1+r) > 0.$$

He is therefore guaranteed to finish the period with a nonnegative wealth a time 1 and he even has a positive probability of having a strictly positive wealth at time 1.

Now, similarly, if $(1+r) \geq u > d$, an investor can sell the stock short and invest it in the money market. At time 1, her wealth would be either

$$(1+r)S_0 - S_0d > 0$$

or

$$(1+r)S_0 - S_0u \geq 0.$$

This exhibits another arbitrage opportunity.

We just proved that if we want to rule out arbitrage opportunities in the market model, we must have the inequalities (5.1). The converse is also true: if the inequalities (5.1) hold, there is no arbitrage opportunity in the market.

5.1.1 European call option

The holder of a European call option has the right (but not the obligation) to buy one share of the stock at time 1 for the strike price K. The only interesting case is when

$$S_1(T) < K < S_1(H). \tag{5.2}$$

In this case, if a head is obtained, the holder of the option can exercise it and her profit is, in this scenario, $S_1(H) - K$. If a tail is obtained, the option expires worthless. In other words the value of the option at time 1 is

$$V_1 = (S_1 - K)^+.$$

Now, what is the value V_0 of the option at time 0, before we know the outcome of the coin toss? The general approach of the no arbitrage pricing theory consists in replicating the option by trading in the stock and money markets.

Imagine that the investor begins with wealth X_0 and buys Δ_0 shares of the stock at time 0. His cash position is then

$$X_0 - \Delta_0 S_0.$$

The investor invests his cash position at time 0 in the money market. Consequently, at time 1, the value of the investor's Portfolio is

$$X_1 = (X_0 - \Delta_0 S_0)(1+r) + \Delta_0 S_1 = (1+r)X_0 + \Delta_0(S_1 - (1+r)S_0).$$

Note that his cash position may be negative or positive. One can say alternately that the investor finances his position in the stock market by lending or borrowing from the money market account. Also, the investor's strategy is said to be *self-financing* in the sense that he only trades by investing in the 2 assets available (money market and stock) and does not receive any additional cash from any other source.

Next, we want to choose Δ_0 such that

$$V_1(T) = X_1(T), V_1(H) = X_1(H).$$

This lead to the equations

$$V_1(T) = (1+r)X_0 + \Delta_0(S_1(T) - (1+r)S_0) \tag{5.3}$$
$$V_1(H) = (1+r)X_0 + \Delta_0(S_1(H) - (1+r)S_0) \tag{5.4}$$

or equivalently,

$$\frac{1}{1+r}V_1(T) = X_0 + \Delta_0(\frac{1}{1+r}S_1(T) - S_0) \tag{5.5}$$
$$\frac{1}{1+r}V_1(H) = X_0 + \Delta_0(\frac{1}{1+r}S_1(H) - S_0). \tag{5.6}$$

We multiply the first equation by a real number $1-\tilde{p}$ and the second equation by the number $\tilde{p}$ and add them. We get

$$X_0 + \Delta_0(\frac{1}{1+r}(\tilde{p}S_1(H) + (1-\tilde{p})S_1(T))) - S_0) = \frac{1}{1+r}[\tilde{p}V_1(H) + (1-\tilde{p})V_1(T)].$$

Next, we choose $\tilde{p}$ so that

$$S_0 = \frac{1}{1+r}(\tilde{p}S_1(H) + (1-\tilde{p})S_1(T)).$$

We then have

$$X_0 = \frac{1}{1+r}[\tilde{p}V_1(H) + (1-\tilde{p})V_1(T)].$$

Finally, we can solve the previous equation for $\tilde{p}$

$$S_0 = \frac{1}{1+r}(\tilde{p}uS_0 + (1-\tilde{p})dS_0) = \frac{S_0}{1+r}(\tilde{p}(u-d) + d].$$

We find

$$\tilde{p} = \frac{1+r-d}{u-d}, 1-\tilde{p} = \frac{u-1-r}{u-d}.$$

Finally, we can compute Δ_0 by subtracting (4) from (3) to derive the so-called *delta-hedging formula*

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}.$$

To recapitulate, consider an investor who is short a European call. Her initial wealth is X_0; she then buys at that time, Δ_0 shares of stock, where Δ_0 is given by the above formula and invests her cash position in a money market account. At time 1, her Portfolio will be worth $V_1(H)$ if the coin toss is a head and $V_1(T)$ is the coin toss is a tail. Note that this argument is valid for any derivative security, not just a European call. The investor has *hedged a short position in the derivative security.* Finally, the price of the derivative security at time 0 should allow the investor to hedge the short position in the claim without introducing any arbitrage opportunity. This price is given by the so-called *risk-neutral pricing formula*

$$V_0 = X_0 = \frac{1}{1+r}[\tilde{p}V_1(H) + (1-\tilde{p})V_1(T)].$$

We can interpret $\tilde{p}$ and $1-\tilde{p}$ as probabilities of head and tail respectively. However, they are distinct from the actual probabilities $p, 1-p$ and they are called *risk-neutral probabilities.* Under the risk-neutral probabilities, the mean rate of growth of the stock is the same as the mean rate of growth of the money market account, i.e.

$$u\tilde{p} + d(1-\tilde{p}) = 1 + r.$$

In order words, in this framework, there is basically no uncertainty since the risk is fully eliminated and stock prices grow at the same rate on average as a money market account.

Finally, you may rewrite the *risk-neutral pricing formula* as the expectation, under the *risk-neutral probability measure* of the discounted payoff at the end of the period, i.e.

$$V_0 = \frac{1}{1+r}\mathbb{E}^{\tilde{\mathbb{P}}}[V_1],$$

where $\tilde{\mathbb{P}}$ denotes the probability defined by setting

$$\tilde{\mathbb{P}}(H) = \tilde{p}, \tilde{\mathbb{P}}(T) = 1 - \tilde{p}.$$

5.2 The multi-period asset pricing model

We now turn to the more general multi-period asset pricing model which is widely used in practice for computing the price of a derivative security. The experiment is now as follows: one tosses a coin infinitely many times and the outcome of each toss i is either a head ($\omega_i = H$) with probability p or a tail ($\omega_i = T$) with probability $1-p$. We start with an initial Stock price S_0 at time 0. At time 1, after the first toss, we define the price of the asset S_1 at time 1 in the following manner; if one obtains a tail at the first toss, the corresponding price $S_1(T)$ is defined by

$$S_1(T) = dS_0$$

whereas if the first toss yields a head, the stock price becomes

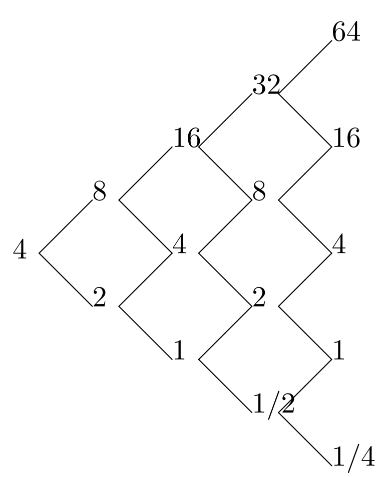
$$S_1(H) = uS_0$$

where $0 < d < u$. Next, we iterate this procedure. After the toss number 2, the price at time 2 becomes

$$\begin{aligned} S_2(HH) &= uS_1(H) = u^2 S_0, S_2(HT) = dS_1(H) = duS_0 \\ S_2(TH) &= uS_1(T) = duS_0, S_2(TT) = dS_1(T) = d^2 S_0 \end{aligned}$$

The third toss has 8 possible outcomes and, given S_2, you can compute the asset price S_3 for each of these outcomes. To illustrate this procedure, we show a plot of a 4-period binomial asset-pricing model for the values $S_0 = 4$, $u = 2$ and $d = 1/2$.

5.2.1 A Binomial tree



As in the one-period case, we introduce the risk-free interest rate $r \geq 0$. We assume again that the following set of inequalities which includes the no arbitrage condition holds

$$0 < d < 1 + r < u.$$

5.2.2 European derivative security

Consider now a European derivative security that expires at a time $N \geq 2$. For instance, the payoff of a European call with strike price K and expiration $N = 2$ is

$$V_2 = (S_2 - K)^+$$

where S_2 and hence, V_2, depend on the first two coin tosses. We want to determine the fair price of the European call a time 0, or more generally the price of a European derivative security at time 0. We proceed in a similar fashion as for the one-period model and we can define recursively the value of the Portfolio, the value of the derivative security and the hedging strategy.

The wealth equation reads

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n).$$

As in the one-period case, the *risk-neutral probabilities* are

$$\tilde{p} = \frac{1+r-d}{u-d}, 1-\tilde{p} = \frac{u-1-r}{u-d}.$$

The value of the derivative security is defined recursively backward

$$V_n(\omega_1\omega_2...\omega_n) = \frac{1}{1+r}[\tilde{p}V_{n+1}(\omega_1\omega_2...\omega_n H) + (1-\tilde{p})V_{n+1}(\omega_1\omega_2...\omega_n T)]$$

and the hedging strategy is given by

$$\Delta_n(\omega_1\omega_2...\omega_n) = \frac{V_{n+1}(\omega_1...\omega_n H) - V_{n+1}(\omega_1...\omega_n T)}{S_{n+1}(\omega_1...\omega_n H) - S_{n+1}(\omega_1...\omega_n T)}.$$

In all of the above equations, $n = 0...N-1$. Then we have at maturity, i.e. for $n = N$,

$$X_N(\omega_1\omega_2...\omega_N) = V_N(\omega_1\omega_2...\omega_N)$$

for all the outcomes $\omega_1\omega_2...\omega_N$.

This model is said to be complete because every derivative security can be replicated in the market model.

5.2.3 Example

Take $S_0 = 4, u = 2, d = 1/2$ and $r = 1/4$. Consider the European call with strike 3 and maturity 3. Its payoff at time 3 is

$$V_3 = (S_3 - 3)^+.$$

The risk-neutral probabilities are

$$\tilde{p} = 1/2, 1 - \tilde{p} = 1/2.$$

The payoff at time $N = 3$ for the various outcomes is

$$\begin{aligned}
V_3(HHH) &= 32 - 3 = 29, V_3(HHT) = 8 - 3 = 5\\
V_3(HTH) &= 8 - 3 = 5, V_3(HTT) = (2 - 3)^+ = 0\\
V_3(TTH) &= (2 - 3)^+ = 0, V_3(TTT) = (1/2 - 3)^+ = 0\\
V_3(THH) &= 8 - 3 = 5, V_3(THT) = (2 - 3)^+ = 0
\end{aligned}$$

We can compute recursively backward V_2, V_1 and V_0:

$$\begin{aligned}
V_2(HH) &= \frac{4}{5}[\frac{1}{2}V_3(HHH) + \frac{1}{2}V_3(HHT)] = 68/5 = 13.6\\
V_2(HT) &= \frac{4}{5}[\frac{1}{2}V_3(HTH) + \frac{1}{2}V_3(HTT)] = 2\\
V_2(TH) &= \frac{4}{5}[\frac{1}{2}V_3(THH) + \frac{1}{2}V_3(THT)] = 2\\
V_2(TT) &= \frac{4}{5}[\frac{1}{2}V_3(TTH) + \frac{1}{2}V_3(TTT)] = 0.
\end{aligned}$$

Next we have

$$
\begin{aligned}
V_1(H) &= \frac{4}{5}[\frac{1}{2}V_2(HH) + \frac{1}{2}V_2(HT)] = 6.24 \\
V_1(T) &= \frac{4}{5}[\frac{1}{2}V_2(TH) + \frac{1}{2}V_2(TT)] = 0.8
\end{aligned}
$$

Finally,

$$V_0 = \frac{4}{5}[\frac{1}{2}V_1(H) + \frac{1}{2}V_1(T)] = 2.816$$

Also the hedge at time 0 is

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)} = \frac{5.44}{6} = 0.907$$

If you write a naive program implementing this algorithm, its complexity is exponential (the number of computation grows exponentially with the number of periods. For example, for $n = 100$ periods, there are $2^{100} \approx 10^{30}$ possible outcomes. You need to make your code more efficient and this is feasible because the number of possible values that the option takes is much smaller than the number of possible outcomes. For instance, in our previous application, there are 8 possible outcomes for V_3 but only 4 different values, $1/2, 2, 8$ and 32.

Similarly, you may apply this algorithm to different types of derivative securities. For instance, you may compute the value of a *lookback option* with payoff

$$V_3 = \max_{0 \leq n \leq 3} S_n - S_3.$$

It turns out that the option's value at time 0 is $V_0 = 1.376$ and that the corresponding hedge at time 0 is $\Delta_0 = 0.1733$.

Chapter 6

Module 6: Introduction to stochastic processes

Stochastic processes are used for modeling the time evolution of financial assets. In this course, we have already encountered some well-known examples of *stochastic processes*, i.e. random variables that depend on time. For instance, sequences of random variables are stochastic processes in discrete time. Another example is Markov chains, which refers to the whole class of stochastic processes with a discrete state space, that satisfy the Markov property.

In this lecture, we present the some of the basic mathematical concepts and definitions of the theory of stochastic processes such as stationarity and independence of increments. These concepts constitute the foundation for conducting empirical studies and building sound mathematical models. We illustrate them with the random walk and the Poisson process, which are the building blocks of many financial models.

We also discuss the features of real financial data that have been discovered through a number of empirical studies. Unfortunately, it is very difficult, to come up with a model that possesses the same features as real financial data. In this course, you are studying the most simple models for the evolution of financial assets and these are somewhat unrealistic. In addition to these notes, I refer for this discussion to the very well written and pedagogical article by Rama Cont, *Empirical properties of asset* *returns: stylized facts and statistical issues*, published in Quantitative Finance, Volume 1 (2001), 223-236.

6.1 Preliminary concepts and examples

First, we list some elementary definition. The story often starts with a *random experiment* E, i.e. an experiment that can be repeated under the same conditions and whose result cannot be predicted with certainty.

Next, we denote by Ω the sample space, i.e. the set of possible outcomes for the experiment E.

Thirdly, we model the time by using a set. Generally, we will either have $T = [0, +\infty)$ (continuous-time process) or $T = \{0, 1, 2, 3...\}$ (discrete-time process).

Then a stochastic process is a function X of the time and of an outcome, $X(t, \omega)$ where $t \in T$, $\omega \in \Omega$. For ω fixed, $\{X(t, \omega), t \in T\}$ is called a sample path. For t fixed, $X(t)$ is a random variable. Roughly, a stochastic process is just a sequence on random variables.

Often, the notation S_X is used for the state space, which is the set of values that the stochastic process X can take. If S_X is finite or countably infinite, X is said to be a discrete-state process ($S_X = \{0, 1, 2, \cdots N\}$, $S_X = \mathbb{N}$ or $S_X = \mathbb{Z}$). If S_X is uncountably infinite (for instance $S_X = \mathbb{R}$), X is said to be a continuous-state process. We have seen in the first lectures some examples of stochastic processes. We list below a few easy examples.

6.1.1 Some elementary examples

- An elementary continuous-time process $X(t, \omega) = tY(\omega)$ where $t \in (0, +\infty)$ and $Y > 0$ is a random variable. Since $S_X = [0, +\infty)$, it is a continuous-state process.

- An elementary discrete-time and discrete-state process that we have seen in the first lecture: the symmetric random walk. We reintroduce it here by using the vocabulary that we just defined above.

  E : a coin is tossed infinitely many times.

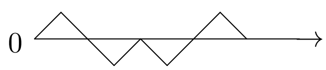
$T = \{0, 1, 2, ...\}$. Let $n \in T$ denotes the toss number. At each toss, a head or a tail is obtained.

We assume that the random walk starts at the origin, i.e. $X_0 = 0$.

When a tail is obtained at the $(n + 1)$th toss: $X_{n+1} = X_n - 1$.

When a head is obtained at the $(n + 1)$th toss: $X_{n+1} = X_n + 1$.

To illustrate the random walk



6.1.2 Distribution of a stochastic process

You may simply fix the time $t \in T$ and look at the distribution of the random variable $X(t, \cdot)$. However, this is not a very rich description of the stochastic process and more generally you may prefer to look at the joint distribution of the process taken at different times.

More precisely, the distribution function of a stochastic process is defined in the following manner. One considers the so-called *distribution function of order k of the stochastic process X*. Given an arbitrary set of k times $t_1, t_2, ..., t_k$, the joint distribution of the random vector $(X(t_1), X(t_2), ..., X(t_k))$ is given by

$$F(x_1, ..., x_k; t_1, ..., t_k) = \mathbb{P}[X(t_1) \le x_1, ..., X(t_k) \le x_k].$$

Probability mass of order k of the discrete-time and discrete-state s.p. X:

In the discrete case, the distribution of order k can also be described by the probability mass function

$$P(x_1, ..., x_k; n_1, ..., n_k) = \mathbb{P}[X_{n_1} = x_1, ..., X_{n_k} = x_k].$$

Continuous-time and continuous-space s.p.

In the continuous case, when the joint cumulative distribution function is differentiable k times, the joint density function of order k is given for a set of k times, $t_1, t_2, ..., t_k$, by

$$f(x_1, ..., x_k; t_1, ..., t_k) = \frac{\partial^k}{\partial x_1 ... \partial x_k} F(x_1, ..., x_k; t_1, ..., t_k).$$

6.1.3 First and second moments

We define the first and second moments by fixing the time variable t. The first moment is simply $\mathbb{E}[X(t)]$, whereas the second moment is given by

$$Var[X(t)] = \mathbb{E}[(X(t) - \mathbb{E}[X(t)])^2].$$

6.1.4 Independence and stationarity of increments

In financial models, we use as building blocks processes with independent and stationary increments. We give these two definitions below.

Independence:

If for all $0 \leq t_1 < t_2 < t_3 < \cdots < t_n$, the random variables $X(t_2) - X(t_1), X(t_3) - X(t_2), \cdots, X(t_n) - X(t_{n-1})$ are independent, X is said to be a process with *independent increments.*

Stationarity:

If for all $t_1 < t_2$, $s \geq 0$, the random variables $X(t_2) - X(t_1)$ and $X(t_2 + s) - X(t_1 + s)$ have the same distribution, X is said to be a process with *stationary increments.*

6.1.5 An example: The Bernouilli Process revisited

We recall the Bernouilli trials: One rolls a die independently an infinite number of times and define a success as rolling a six. A Bernouilli process is a sequence of Bernouilli random variables associated with Bernouilli trials $X_1, X_2, ..., X_k$

$$X_k = \begin{cases} 1 & \text{if } k\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

The Bernouilli variable X_k has a Bernouilli distribution

$$p(x) = p^x(1-p)^{1-x}, \text{ for } x = 0, 1$$

with parameter $p = 1/6$. In addition, its expectation is given by

$$\mathbb{E}[X_k] = \text{ probability of success } \times 1 = \frac{1}{6}.$$

One may also define the stochastic process Y_n representing the number of successes after n trials

$$Y_n = \sum_{k=1}^{n} X_k$$

where X_k are the Bernouilli random variables. The distribution of Y_n is Binomial with parameters n and p. By inverting this formula, we also obtain the relationship

$$X_{n+1} = Y_{n+1} - Y_n.$$

The process $X_{n+1} = Y_{n+1} - Y_n$ is called *one-step increment* of the process Y_n.

Furthermore, since $X_1, X_2, ..., X_n$ are independent random variables, the process Y_n has *independent increments.*

Besides, since $Y_{k+n+1} - Y_{k+n} = X_{k+n+1}$ and $Y_{n+1} - Y_n = X_{n+1}$ have the same distribution, the process Y_n also has *stationary increments.* Note, that we are not proving these last two properties rigorously here since we are considering only one-step increments for a sake of simplicity, instead of more general m-step increments. However, the general argument is very similar as the above simple argument and these properties of Y_n really do hold!

6.1.6 Another important example: the Random Walk

Let us go back to the example of the symmetric Random Walk. One tosses a fair coin infinitely many times. Since the coin is unbiased, the probability of a tail is equal to the probability of a head,i.e. $p = \mathbb{P}(H) = \mathbb{P}(T) = 1 - q = 1/2$.

The successive outcomes of the tosses are denoted by $\omega = \omega_1\omega_2\omega_3...\omega_n...$ where ω_n is the outcome of the toss number n.

We define the one-step increment of the random walk

$$Y_i = \begin{cases} -1 & \text{if } \omega_i = T \\ 1 & \text{if } \omega_i = H \end{cases}$$

and we define the random walk by initializing it

$$X_0 = 0$$

and by adding up all the one-step increments:

$$X_k = \sum_{i=1}^{k} Y_i \text{ for } k = 1, 2, ...$$

Given a set of integers $0 = k_0 < k_1 < ... < k_i < k_{i+1} < ... < k_m$, we can further define the random variables called increments of the random walk

$$X_{k_{i+1}} - X_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} Y_j.$$

The increments $X_{k_1} - X_0$, $X_{k_2} - X_{k_1}$,...,$X_{k_{i+1}} - X_{k_i}$,...,$X_{k_m} - X_{k_{m-1}}$ are independent. So we can state, that the random walk has *independent increments*. Clearly, the increments of the random walk are also *stationary* since the coin tosses are independent and identical.

In addition,

$$\mathbb{E}[X_{k_{i+1}} - X_{k_i}] = \sum_{j=k_i+1}^{k_{i+1}} \mathbb{E}[Y_j] = 0.$$

$$
\begin{aligned}
Var[X_{k_{i+1}} - X_{k_i}] &= \mathbb{E}[(\sum_{j=k_i+1}^{k_{i+1}} Y_j)^2] \\
&= \mathbb{E}[\sum_{j=k_i+1}^{k_{i+1}} Y_j^2 + \sum_{j=k_i+1}^{k_{i+1}} \sum_{k \neq j} Y_j Y_k] \\
&= \sum_{j=k_i+1}^{k_{i+1}} \mathbb{E}[Y_j^2] + \sum_{j=k_i+1}^{k_{i+1}} \sum_{k \neq j} \mathbb{E}[Y_j Y_k] \\
&= \sum_{j=k_i+1}^{k_{i+1}} 1 + \sum_{j=k_i+1}^{k_{i+1}} \sum_{k \neq j} 0 \\
&= k_{i+1} - k_i.
\end{aligned}
$$

The variance of the increment over the time interval $[k_i, k_{i+1}]$ is equal to $k_{i+1} - k_i$.

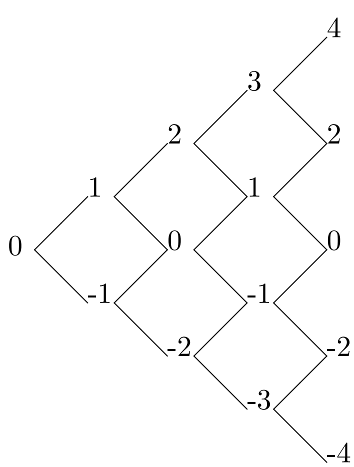
We also deduce the expectation and variance of the random walk from the last 2 results

$$\mathbb{E}[X_k | X_0 = 0] = \mathbb{E}[X_k - X_0 | X_0 = 0] = \mathbb{E}[X_k - X_0] = 0.$$

$$Var[X_k] = Var[X_k - X_0] = k.$$

Representation by a tree

We can represent the random walk process by using a tree.



6.1.7 The Auto Regressive process (abbreviated AR)

I present in this section the well-known discrete time $AR(1)$ process, in econometrics.

We consider a finite time horizon $T > 0$, and a partition of the time interval $[0, T]$ into subintervals of length Δt.

The auto-regressive process $(X_n)_n$, of type AR(1) in discrete time, where X_n represents the value of the process at time $n\Delta t$, can be written as

$$X_{n+1} = aX_n \Delta t + \epsilon_{n+1},$$

where a is a nonnegative real number and $\epsilon_1, \epsilon_2, \cdots, \epsilon_n \cdots$ are independent and identically distributed centered random variables.

In financial applications, X_n would for instance refer to an asset price, or its log price.

For instance, one can assume that ϵ_n is normally distributed but this is not necessarily the case. Note that, in econometrics, one often assumes that $\Delta t = 1$, which leads to the equation

$$X_{n+1} = aX_n + \epsilon_{n+1},$$

6.1.8 Special case: $a = 1$

When $a = 1$, the dynamics of X_n read

$$X_{n+1} = X_n + \epsilon_{n+1}.$$

This can be rewritten as

$$X_{n+1} - X_n = \epsilon_{n+1}.$$

In other words, one can see that $\epsilon_1, \cdots \epsilon_{n+}$, the increments of $(X_n)_n$ are independent and strictly stationary. However, the successive values of the process $(X_n)_n$ itself are not independent and the process $(X_n)_n$ is not stationary!

The above model is driven by the i.i.d variables $\epsilon_1, \cdots, \epsilon_n$. In other words, the sequence $\epsilon_1, \cdots, \epsilon_n, \cdots$ represents the independent and stationary increments of the random walk process (X_n).

The model in this case is basically the random walk model introduced in chapter 1. Given an initial condition X_0, since ϵ_n is assumed to be centered, the expectation of X_n is given by

$$\mathbb{E}[X_n|X_0] = \mathbb{E}[X_{n-1}|X_0] = X_0.$$

Furthermore, we can rewrite X_n as

$$\begin{aligned} X_{n+1} &= X_n + \epsilon_{n+1} && (6.1) \\ &= X_{n-1} + \epsilon_{n+1} + \epsilon_n && (6.2) \\ &= X_{n-2} + \epsilon_{n+1} + \epsilon_n + \epsilon_{n-1} && (6.3) \\ &= X_0 + \sum_{i=1}^{n+1} \epsilon_i. && (6.4) \end{aligned}$$

We can compute the variance of X_n

$$var[X_{n+1}] = var[\sum_{i=1}^{n+1} \epsilon_i] = \sum_{i=1}^{n+1} var[\epsilon_i].$$

Denoting by σ^2 the common variance of ϵ_i, we obtain

$$var[X_{n+1}] = (n+1)\sigma^2.$$

Consequently, in a random walk model, the variance increases linearly with time and as n converges to $+\infty$, it becomes infinite. So the variance of the random walk is explosive.

6.1.9 Second case: $a < 1$

In this second case, we can rewrite X_n as

$$\begin{aligned} X_{n+1} &= aX_n + \epsilon_{n+1} && (6.5)\\ &= a^2 X_{n-1} + \epsilon_{n+1} + a\epsilon_n && (6.6)\\ &= a^{n+1}X_0 + \sum_{i=0}^{n} a^i \epsilon_{n+1-i}. && (6.7) \end{aligned}$$

As n converges to $+\infty$, the term $a^{n+1}X_0$ converges to 0. So, on average, X_n converges to its long-run mean 0! This process is a mean-reverting process.

Note that, by contrast, when $a > 1$, the mean of the process X_n generally explodes to ∞, unless $X_0 = 0$ (the sign depends on the sign of X_0).

Furthermore, it can be shown, that, since a^i converges to 0 as i converges to $+\infty$, this process is stationary in the long run, i.e. as n converges to $+\infty$, thanks to the dampening effect of the a^i terms. In other words, it is considered stable, whereas the case wen $a > 1$ is unstable (or also called explosive). Note that the proof is not trivial and I won't present it in these notes.

In particular, the variance of X_n becomes in the case when $a < 1$:

$$var[X_{n+1}] = var[\sum_{i=0}^{n} a^i \epsilon_{n+1-i}] = \sum_{i=0}^{n} a^{2i} var[\epsilon_i] = \sigma^2 \sum_{i=0}^{n} a^{2i} = \sigma^2 \frac{1-a^{2(n+1)}}{1-a^2}.$$

So, as $n \to +\infty$, $var[X_{n+1}]$ converges to the constant $\frac{\sigma^2}{1-a^2}$.

Alternately, you can define a AR(1) process with a non zero log run mean. For instance, you can use the same equation and assume the i.i.d variables have a common mean μ, instead of 0. An alternate way of introducing a non zero long run mean is by writing the equation

$$X_{n+1} - \mu = a(X_n - \mu) + \epsilon_{n+1},$$

where $\epsilon_1, \cdots \epsilon_n$ are centered i.i.d random variables.

The recursive iteration procedure gives us

$$X_{n+1} - \mu = a^{n+1}(X_0 - \mu) + \sum_{i=0}^{n} a^i \epsilon_{n+1-i}.$$

Here, as n converges to $+\infty$, the long-run mean is given by μ.

This equation is often rewritten as

$$X_{n+1} - X_n = (a-1)(X_n - \mu) + \epsilon_{n+1}.$$

The long run variance is the same as before.

Finally, in the case when the noise terms are normally distributed, the analogue model in continuous-time would be

$$dX(t) = (1-a)(\mu - X(t))dt + \sigma dW(t),$$

where $W(.)$ is a Wiener process (also called a Brownian motion). This model is often used for the spread of two cointegrated assets as we wil see in a later chapter..

To conclude, the above discussion is meant to clarify the concept of invariance (independence and stationarity), in the univariate case. Indpendence and stationarity are important because our statistical methods are based on such assumtpions. We covered the two main types of basic models, that is, the random walk model and the mean-reverting AR(1) model with coefficient $a < 1$. The random walk model is often used for modeling the log prices of stocks whereas the mean-reverting model can be used for interest rates, volatility, or the spread of cointegrated assets.

6.1.10 A distraction: The Saint Petersburg Paradox

(Daniel Bernouilli, early 1700's)

A fair coin is tossed until the player gets a Head. If this happens at the toss number n, the player is paid 2^n dollars by the bank. What is the fair amount to pay to play this game?

Answer:

Since the probability of obtaining a head for the first time at the k throw is $1/2^k$, the expected payoff of the game is

$$\sum_{k=1}^{\infty}(1/2)^k \times 2^k = 1 + 1 + 1 + ... = +\infty.$$

So, the fair price to pay for playing this game is infinite!

This implies that, theoretically, if you were asked to pay a large sum of money, say $2^{20} \approx 1$ million dollars to play this game, then that would be a 'good' deal. Note that the probability of earning 2^{20} dollars or more is equal to 2^{-20}.

What happens is that some events of very small probability have an even bigger payoff, which distorts the expectation.

One way of 'resolving' the paradox is to introduce the concept of a utility function. The rational behind it is that the value of 1 dollar isn't the same for a person that is poor and a person that is wealthy; more precisely, it is expected that the value of our wealth increases with the amount of wealth we own (it is always better to have more) but the rate of increase in this value decreases (the more we have, the less importance we give on having an extra dollar). So the value of wealth should be a *concave* function of wealth.

Let u be an increasing concave utility function. Example of such functions can be $\log(x)$, $\sqrt{x}$. We could compute our expected utility of wealth $\mathbb{E}[u(w)]$ instead of our expected wealth $\mathbb{E}[w]$. The computations for $u(x) = \log(x)$ is left to the reader, and one can check that in this case the expectation is finite.

6.2 Autocovariance and autocorrelation functions

One can study how the stochastic process correlates to itself as it evolves in time. For this purpose, we define the the autocovariance and autocorrelation functions. Note that this terminology is not fully universal and may depend on the particular textbook you are using.

- **Autocovariance function:** at the point (t_1, t_2), it is defined by

$$C_X(t_1, t_2) = \mathbb{E}[X(t_1)X(t_2)] - \mathbb{E}[X(t_1)]\mathbb{E}[X(t_2)].$$

- **Autocorrelation function:** at the point (t_1, t_2), it is given by

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{(Var[X(t_1)]Var[X(t_2)])^{\frac{1}{2}}}.$$

6.2.1 Example of the $AR(1)$ process for $a = 1$

We compute the corresponding autocovariance function as follows. By (6.4), we have that

$$\begin{aligned}
&\mathbb{E}[X_{n+p}X_n] = \\
&\mathbb{E}[(X_0 + \sum_{i=1}^{n+p} \epsilon_i)(X_0 + \sum_{i=1}^{n} \epsilon_i)] = \\
&X_0^2 + \sum_{i=1}^{n+p}\sum_{j=1}^{n} \mathbb{E}[\epsilon_i \epsilon_j] = \\
&X_0^2 + \sum_{i=1}^{n} \mathbb{E}[\epsilon_i^2] = \\
&X_0^2 + n\sigma^2.
\end{aligned}$$

Subtracting the product of the expectations, we find that the autocovariance function is given by

$$C_X(n + p, n) = n\sigma^2.$$

6.2.2 Strict sense and wide sense stationarity of a stochastic process

Previously, we only applied the concept of stationarity to the increments of a stochastic process. However, this concept is generally applicable to any stochastic process. We give below a modified definition of stationarity that takes this remark into account. Note that, although the definition below is worded differently copmared to the earlier definition that applied to increments, it is essentially the same as before. Furthermore, we also introduce in this section a weaker concept of stationarity called *wide sense stationarity* or *weak sense stationarity*, which is very useful in practice. We start below with the definition of stationarity, which has been renamed *strict sense stationarity*, for the purpose of differentiating it from its weaker countepart *wide sense stationarity*. Note that a strict sense stationary process is always wide sense stationary.

- X is strict sense stationary (sss in short) if for all $s, n, t_1, \cdots, t_n$, $x_1, \cdots, x_n$,
$$F(x_1, ..., x_n; t_1, ..., t_n) = F(x_1, ..., x_n; t_1 + s, ..., t_n + s),$$
- X is wide-sense stationary (wss in short) if $\mathbb{E}[X(t)]$ is independent of t and $R_X(t, t+s) = \mathbb{E}[X(t)X(t+s)]$ depends only on s.

In what follows, we apply these concepts to several elementary examples.

First example:

We let $X(t) = Y$ for $t \geq 0$ where Y is a random variable. Clearly, X is sss because its distribution does not depend on the time variable.

Second example:

The Bernouilli process described earlier is not wss because its expectation $\mathbb{E}[Y_n] = np$, which is not independent of the time n.

Third example:

The process $X(t, \cdot) = tY(\cdot)$ where Y is a r.v, is not wss since $\mathbb{E}[X(t)] = t\mathbb{E}[Y]$ is not independent of t.

Fourth example: the symmetric random walk

First of all, $\mathbb{E}[X_k] = 0$ is independent of the time index k. Next, we need to compute

$$
\begin{aligned}
&R_X(k, k+j) = \mathbb{E}[X_k X_{k+j}] = \\
&\mathbb{E}[X_k(X_{k+j} - X_k) + X_k^2] = \\
&\mathbb{E}[X_k]\mathbb{E}[X_{k+j} - X_k] + \mathbb{E}[X_k^2] = \\
&0 + \mathbb{E}[X_k^2] = k.
\end{aligned}
$$

So the symmetric random walk is not wss.

6.3 The Gaussian Processes

Gaussian processes constitute a particular subclass of continuous-time stochastic process. We dedicate to it a separate section because this is one of the most important examples. Later on, you will study the Wiener process, which is the most important Gaussian process. Other examples of processes belonging to this class are the fractional Brownian motions, which won't be covered in this basic core course.

6.3.1 Definition

A stochastic process X is said to be a Gaussian Process if the vector

$$(X(t_1), X(t_2), ..., X(t_n))$$

has a multi-normal distribution for any n and for all $t_1, t_2, ..., t_n$.

6.3.2 Multi-normal distribution

For a sake of completeness, we recall the definition of a multi-normal definition.

A random vector $(X_1, ..., X_n)$ has a multi-normal distribution if each random variable X_k can be expressed as a linear combination of m

independent standard gaussian random variables with $m \leq n$, i.e.:

$$X_k = \mu_{X_k} + \sum_{j=1}^{j=m} c_{kj} Z_j \text{ where } \mu_{X_k} \in \mathbb{R}, Z_j \sim N(0,1).$$

An immediate consequence of this definition is that any affine combination of multi-normally distributed random vectors has a multi-normal distribution.

Furthermore, $\mathbb{E}[X_k] = \mu_{X_k}, Var[X_k] = \sum_{j=1}^{m} c_{kj}^2, Cov[X_i, X_k] = \sum_{l=1}^{m} c_{il} c_{kl}$.

The joint distribution of the vector $(X_1, X_2, ..., X_n)$ is completely determined by the vector of means $\mu = (\mu_{X_1}, ..., \mu_{X_n})$ and the covariance matrix K.

$$K = \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & ... & Cov[X_1, X_n] \\ \\ Cov[X_1, X_2] & Var[X_2] & ... & Cov[X_2, X_n] \\ . & . & . & \\ . & . & . & \\ . & . & . & \\ Cov[X_1, X_n] & ... & ... & Var[X_n] \end{bmatrix}.$$

Note that, since $Cov[X_i, X_j] = Cov[X_j, X_i]$, the matrix K is symmetric.

Furthermore, if K is nonsingular, the density of $(X_1, X_2, \cdots, X_n)$ exists and can be written as

$$f_X(x) = \frac{1}{2\pi^{\frac{n}{2}}} \frac{1}{(\det K)^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x-\mu)K^{-1}(x^T - \mu^T)\}.$$

If K is singular, the vector of random variables $(X_1, \cdots, X_n)$ does not even have a density function!

K is also nonnegative definite in the following sense:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \xi_i \xi_j Cov[X_i, X_j] \geq 0 \text{ for all } \xi = (\xi_1, \cdots, \xi_n) \in \mathbb{R}^n.$$

We also recall that in the case when $(X_1, X_2, \cdots, X_n)$ has a multi-normal distribution, the random variables $X_1, X_2, \cdots, X_n$ are independent if and only if the matrix K is a diagonal matrix.

Note, that, in general, zero linear correlation does not imply independence! The random variables could depend on one another in some nonlinear way even if the linear correlation coefficient is equal to 0.

Finally, we end this section by some facts on the stationarity of Gaussian processes: a Gaussian process is not necessarily weak sense stationary. A well-know Theorem states that a Gaussian process which is weak sense stationary is also strict sense stationary.

Theorem If a Gaussian process is weak sense stationary, then it is also strict sense stationary.

Since the proof of the last result is very straightforward, it is included here.

Proof: $R_X(t_1, t_2) = R_X(t_1 - t_2), m_X(t) = \mu$. The joint distribution is completely defined by μ and $C_X(t_1, t_2) = R_X(t_1 - t_2) - \mu^2$ and hence is invariant with respect to a translation $+s$.

6.4 Ergodicity

We saw a definition of ergodicity of Markov chains in the third lecture. I describe this concept in a more general setting now. A process is said to be ergodic if any of its characteristics can be obtained from a single sample path. A stochastic process can be *mean-ergodic*, *distribution-ergodic*, *autocorrelation function-ergodic*, etc ...

In this lecture, we focus on *mean-ergodicity.* First of all, we define the temporal mean of the process X as follows:

$$< X >_\theta = \frac{1}{\theta} \int_0^\theta X(t, \omega) dt.$$

Note that the temporal mean is defined path-wise and is a random variable. Next, we assume that the expectation of the process X is independent of the time and we let $m = \mathbb{E}[X(t)]$ be the expectation of X.

Definition

A process X with constant mean m is said to be mean-ergodic if

$$\lim_{\theta \to +\infty} \mathrm{var}[< X >_\theta] = 0.$$

Note that

$$\mathbb{E}[< X >_\theta] = \frac{1}{\theta} \int_0^\theta \mathbb{E}[X(t)]dt = m.$$

Consequently the above definition is equivalent to the convergence of the temporal mean to m in mean-squares as $\theta \to +\infty$:

Equivalent definition

A process X with constant mean m is said to be mean-ergodic if

$$\lim_{\theta \to +\infty} \mathbb{E}[(< X >_\theta - m)^2] = 0.$$

6.4.1 Examples

1. Let $X(t) = Y$, where Y is a random variable, for all t. X is not mean-ergodic:

$$\mathbb{E}[X(t)] = \mathbb{E}[Y], < X >_s = \frac{1}{s} \int_0^{+s} X(u)du = \frac{1}{s} \int_0^{+s} Y du = Y.$$

2. The Bernouilli variables are a mean-ergodic process; indeed

$$< X_i >_k = \frac{1}{k} \sum_{i=1}^{k} X_i.$$

Next

$$\mathbb{E}[< X_i >_k] = \frac{1}{6} = \mathbb{E}[X_i],$$

and

$$\lim_{k \to +\infty} var[< X_i >_k] = \lim_{k \to +\infty} \frac{1}{k^2} k(\frac{1}{6} - \frac{1}{6^2}) = 0.$$

3. The increment $(k,\omega) \to X_{k+1}(\omega) - X_k(\omega)$ of the symmetric random walk is a mean-ergodic process; indeed

$$< X_{i+1} - X_i >_k = \frac{1}{k+1}\sum_{i=0}^{k}(X_{i+1} - X_i) = \frac{1}{k+1}(X_{k+1} - X_0) = \frac{X_{k+1}}{k+1}.$$

   Hence,

$$\mathbb{E}[< X_{i+1} - X_i >_k] = 0 = \mathbb{E}[X_{k+1} - X_k],$$

   and

$$\lim_{k\to+\infty} var[< X_{i+1} - X_i >_k] = \lim_{k\to+\infty} \frac{1}{(k+1)^2}(k+1) = 0.$$

The above examples are very simple but, generally it is very hard to determine whether a specific stochastic process is ergodic or not.

6.5 The features of real financial data

This section draws heavily from the article by R. Cont cited in the introduction. In the past ten years, a large amount of data has become available and has been studied by using *non-parametric* statistical methods. While parametric statistical methods assume that the data are realizations of a specific stochastic process whose parameters can be estimated, the non-parametric approach does not and only assumes some qualitative properties of the stochastic process the data are drawn from. The downside of this approach is that it only allows to study the properties of the underlying process and does not describe it precisely. However, one can resort to so-called *semi-parametric* methods to measure some characteristics of the data without describing it completely. For instance, the tail behavior of the data can be measured by using a single parameter.

Typically, one studies the characteristics of the evolution of the log return of a financial asset, where, if $S(t)$ denotes the price of the asset, the log return over the time interval $[t, t + \Delta t]$ is defined as

$$r(t) = logS(t + \Delta t) - logS(t).$$

The properties of the time series $r(t)$ depends on the time scale considered, that is, the exact value of Δt. In particular, high-frequency data reflects the microstructure of financial markets and possess different characteristics than data at lower frequencies. However, researchers have established some common features shared by a wide range of financial assets across a range of time scales, excluding small intraday time scales. Following the article by R. Cont, I list below some of these features and refer to this article for more details.

- The linear autocorrelations, as defined in section 2.4 are almost zero. Note that this is no longer true for small intraday time scales (20 minutes). However, note that the fact that the linear autocorrelation $\rho_X(t, t+\tau)$, where τ denotes the time lag, is zero, do not imply the independence of the log returns $r(t)$ and $r(t+\tau)$!

- The distribution of returns has *fat tails*, which means that very brutal market events have a non negligible probability of happening. The tail behavior is power-law or Pareto like. Unfortunately this excludes the normal distribution and other well-known distributions such as the stable laws with infinite variance.

  For a sake of completeness, I give below the cumulative distribution F of Pareto type:

$$F(x) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^\alpha & x \geq x_m \\ 0 & x < x_m \end{cases}$$

  where $x_m > 0$ and α, known as the tail index, are parameters.

  I also give you the definition of a stable law:

  Take two independent copies X_1 and X_2 of a random variable X. Then, X is said to be stable if for every $a > 0, b > 0$, $aX_1 + bX_2$ has the same distribution as $cX + d$ for some real numbers $c > 0$ and d.

  The *tail index* k of the distribution, which measures how thick the tail is, is defined as the highest absolute moment which is finite. For both the Gaussian and exponential distribution, $k = +\infty$, which means that all moments are finite. For a power-law distribution, the tail index is the exponent of the distribution. The tail index can be measured using statistical methods. In practice, simple studies show that US and French stocks as well as exchange rates seem to have a finite variance. Techniques based on extreme values confirm that the index is generally strictly bigger than two and show that it is lower or equal to five, its value being typically around three.

- One observes an up-down move asymmetry. Indeed, large drawdowns in stock prices are not compensated by equally large gains.

- As the time scale Δt increases, the distribution of the log returns looks more and more like a normal distribution.

  This explains and partly justifies the widespread use of the normal distribution for the log returns.

- Returns are highly variable at every time scale. This phenomenon is often called *intermittency*. The presence of heavy tails support this statement.

- Researchers found some evidence supporting the existence of a phenomenon that they named *volatility clustering*; indeed, they observed that measures of volatility have positive autocorrelations over several days. This shows that high-volatility events are clustered in time.

- Unlike the linear autocorrelations of log returns, the autocorrelations of absolute log returns, which can be seen as a measure of nonlinear dependence, decay slowly as the time lag increases. One can also look at the autocorrelation function of the squared returns, another measure of nonlinear dependence that decays slowly over several days or even weeks. This shows that the log returns are not independent and this contradicts the hypothesis that the increments of the log asset prices follow a random walk.

  Some researchers claim that this proves the existence of *long-range dependence*. A consequence of this statement is that the Markov property would not be satisfied in general. However, this is no consensus on this matter and it is difficult to test whether time series satisfies the Markov property.

- The changes in volatility are generally negatively correlated with the returns.

- The trading volume is positively correlated with volatility.

- An estimate of the volatility at a coarse time scale predict fairly accurately the volatility at finer time scales.

6.6 The hypotheses behind common Statistical methods

You have to be aware of the various hypotheses that one must makes when applying common Statistical methods. One can identifies three main issues that come up in Statistical estimation of Financial data, namely

1. Stationarity
2. Ergodicity
3. Finite sample properties of estimators.

The log returns are generally assumed to be stationary because this makes the statistical properties invariant in time. In particular, it allows us to estimate moments of the returns. The data actually fail to be stationary in calendar time because of seasonality effects, such as intraday variability, weekend effects, January effects, and so on. It is possible to correct this problem by changing the calendar into a *business time*, in which the data appear stationary.

Next, stationarity alone is not sufficient to prove the convergence of the sample mean estimator to the model expectation and ergodicity is also required to ensure that the empirical average do indeed converge toward the mean of the stochastic process. For instance, there are some multifractal processes used in the modeling of high-frequency data which are not ergodic.

Finally, in Finance, sample sizes are typically small; consequently the error of the estimate of the mean is substantial and it is therefore important to pair the estimate with a confidence interval. However deriving a confidence interval relies on a central limit Theorem and the required assumptions are quite strong. The residuals must be independent and identically distributed and the existence of the 4 first moments is usually required. These assumptions do not always hold for every model considered in Finance and you may question the validity of the estimates and of statistical tests in that case.

In this course, for the most part, we focus on the random walk model for the log returns of financial assets and its continuous time limit, the Wiener process. This model does not match the data's observed features. However its merits are it simplicity and tractability. In particular, its properties, such as stationarity, ergodicity and the finiteness of all its moments make it possible to apply statistical methods to estimate the moments of its distribution.

6.7 Exercises

Problem 1 (30 points)

Consider the process $\{X(t), t \geq 0\}$ defined by

$$X(t) = \begin{cases} 1 & \text{if } t \leq Y \\ 0 & \text{if } t > Y \end{cases}$$

where Y is a uniformly distributed random variable on the interval $(0, 1)$.

1. (5 points) Compute, for $t \in [0, 1]$, the first-order probability mass of X, $f(t, x)$.
2. (5 points) What is the first-order probability mass of X for for $t > 1$?
3. (5 points) Give the expectation and the variance of X.
4. (5 points) Compute the autocovariance function of X, $C_X(t_1, t_2)$.
5. (5 points) Is X wide sense stationary? is X strict sense stationary?
6. (5 points) What is the distribution of an increment of X? Does X have stationary increments?

Problem 2 (15 points)

Consider the stochastic process

$$X(t) = e^{-Yt}, \text{ for } t \geq 0,$$

where Y is a random variable with a uniform distribution on the interval $(0, 1)$.

1. (5 points) Calculate the first-order density function of the process $\{X(t), t \geq 0\}$
2. (5 points) Compute $\mathbb{E}[X(t)]$, for $t \geq 0$.
3. (5 points) Compute the auto-covariance function $C_X(t, t+s)$ for $s, t \geq 0$.

Problem 3 (35 points)

Consider the process $\{X(t), t \geq 0\}$ defined by

$$X(t) = N(t) - \lambda t,$$

where N is a Poisson process with rate $\lambda > 0$

1. (5 points) Compute, for $t_1, t_2, n_1, n_2 > 0$, the second-order probability mass of N, $G(t_1, t_2; n_1, n_2)$.
2. (5 points) Give the expectation and the variance of $X(t)$.
3. (5 points) Compute $\mathbb{E}[X(t_1)X(t_2)]$, for $0 < t_1, \leq t_2$.
4. (5 points) Is X wide sense stationary? is X strict sense stationary?
5. (5 points) is X a martingale? Justify your answer.
6. (10 points) Is $\frac{X(t)}{t}$ mean ergodic?

Problem 4 (30 points) Consider the process $\{X(t), t \geq 0\}$ defined by

$$X(t) = te^Y,$$

where Y is a random variable with a uniform distribution on $(0, 1)$.

1. (6 points) Compute, for $t_1, t_2 > 0$ and $x_1 > 0, x_2 > 0$, the second-order cumulative distribution function of X, $G(t_1, t_2; x_1, x_2)$.
2. (5 points) Compute the expectation and the variance of X.

3. (5 points) Compute $\mathbb{E}[X(t_1)X(t_2)]$, for $0 < t_1 < t_2$.

4. (4 points) Is X wide sense stationary? is X strict sense stationary?

5. (5 points) What is the distribution of an increment of X, $X(t_2) - X(t_1)$, for $0 < t_1 < t_2$ given?

6. (5 points) Does X have stationary increments? Does X have independent increments?

Problem 5 (30 points) Consider the process $\{X(t), t \geq 0\}$ defined by

$$X(t) = 1 + tY,$$

where Y is a continuous random variable with cumulative distribution function F, and such that $\mathbb{E}[Y] > 0$.

1. (6 points) Compute, for $t_1, t_2 > 0$, the second-order cumulative distribution function of X, $G(t_1, t_2; x_1, x_2)$.

2. (5 points) Give the expectation and the variance of X in terms of the expectation and variance of Y respectively.

3. (5 points) Compute $\mathbb{E}[X(t_1)X(t_2)]$, for $0 < t_1, \leq t_2$ in terms of the expectation of Y and the expectation of Y^2.

4. (4 points) Is X wide sense stationary? is X strict sense stationary?

5. (5 points) What is the distribution of an increment of X, $X(t_2) - X(t_1)$, for $0 < t_1 < t_2$ given?

6. (5 points) Does X have stationary increments? Does X have independent increments?

Chapter 7

Module 7: The continuous-time limit of the random walk

7.1 Introduction

In this lecture, we draw heavily from the book by S. Shreve, *Stochastic Calculus for Finance*, II. In particular you will find the proofs in this text. We present both the arithmetic and the geometric random walks with and without drift. In the course of this presentation, we introduce the concept of quadratic variation.

7.2 Passing to the continuous-time limit

I start with the symmetric random walk, that we saw in earlier classes, rescale it and obtain the Brownian motion as the limit of the scaled random walks. The Brownian motion, whose thorough study is postponed to a later class, is the cornerstone of the arithmetic random walk in continuous-time.

First of all, I recall the definition of the basic memoryless symmetric random walk that I presented earlier: we toss a fair coin infinitely many times. On each toss i, the probability of getting a head is $p = \frac{1}{2}$ and the probability of getting a tail is $q = 1 - p = \frac{1}{2}$. We denote the one-step increment by Y_i and the random walk itself by X_n. We assume that $X_0 = 0$

7.2.1 Quadratic variation of the random walk

For reasons that will be made clearer later, the quadratic variation of the random walk is quantity of great interest for the Black-Scholes theory. It is defined as follows.

The quadratic variation up to time k is denoted by $[X, X]_k$ and is defined by

$$[X, X]_k = \sum_{j=1}^{k} (X_j - X_{j-1})^2.$$

We compute the quadratic variation of the random walk along each path

$$[X, X]_k = \sum_{j=1}^{k} (X_j - X_{j-1})^2 = \sum_{j=1}^{k} 1 = k.$$

Note that the quadratic variation of the random walk is independent of the path considered and is equal to $Var[X_k]$.

7.2.2 Scaled symmetric random walk

We rescale the random walk in order to approximate a Brownian motion:

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} X_{nt}.$$

Of course, the number nt is not always an integer and we do not know what X_{nt} means in the case when nt is not an integer. Here is how do we handle this small technical issue in practice:

- On one hand, if nt is a integer, X_{nt} is already defined without ambiguity.

- On the other hand, if nt is not an integer, we define $W^{(n)}(t)$ by *linear interpolation* on the interval $[s, u]$ where s is the nearest real numbers on the left of t such that ns is an integer and u is the nearest real number to the right of t such that nu is an integer.

Next, our goal here is to let $n \to +\infty$ in order to obtain a standard Brownian motion.

From the practical point of view, we are accelerating time and we are reducing the size of the steps. We now have n tosses per unit of time and a total of nt tosses in the time interval $[0, t]$. At the limit $n \to +\infty$, we toss the die infinitely fast and the size of each step is infinitesimally small. Furthermore, I use a specific scaling: the number of coin tosses per unit time is the square of the time step. The particular scaling is dictated by the Central Limit Theorem. Our goal is to apply the Central Limit Theorem to the rescaled random walk to obtain a normal distribution a the limit.

Next, I state some facts about the scaled random walks:

- The increments of the scaled random walk are independent, i.e. For a collection of times $0 = t_0 < t_1 < t_2 < ... < t_{m-1} < t_m$ such that each nt_j is an integer, the increments

  $$W^{(n)}(t_1) - W^{(n)}(t_0), W^{(n)}(t_2) - W^{(n)}(t_1), ..., W^{(n)}(t_m) - W^{(n)}(t_{m-1})$$

  are independent random variables.

- The increments of the rescaled random walk are stationary.

- Expectation of an increment of the scaled random walk

  $$\mathbb{E}[W^{(n)}(t) - W^{(n)}(s)] = 0$$

- Variance of an increment of the scaled random walk

  $$Var[W^{(n)}(t) - W^{(n)}(s)] = \mathbb{E}[(\frac{1}{\sqrt{n}}X_{nt} - \frac{1}{\sqrt{n}}X_{ns})^2]$$

  $$= \frac{1}{n}\mathbb{E}[(X_{nt} - X_{ns})^2] = \frac{1}{n}(nt - ns) = t - s.$$

- Martingale property: let $0 \leq s \leq t$ such that nt and ns are integers. We have

  $$\mathbb{E}[W^{(n)}(t) | \{W^{(n)}(\theta), 0 \leq \theta \leq s\}] = W^{(n)}(s).$$

- Quadratic variation: let $t \geq 0$ be such that nt is an integer. We compute

$$[W^{(n)}, W^{(n)}](t) = \sum_{j=1}^{nt} \{W^{(n)}(\frac{j}{n}) - W^{(n)}(\frac{j-1}{n})\}^2$$

$$= \sum_{j=1}^{nt} \{\frac{1}{\sqrt{n}} Y_j\}^2 = \sum_{j=1}^{nt} \frac{1}{n} = t.$$

Finally, I can state result concerning the limit distribution of the scaled random walk.

Central Limit Theorem As $n \to +\infty$ the distribution of the scaled random walk at time t converges to the normal distribution with mean 0 and variance t.

We recall a normal distribution with mean 0 and variance t:

$$f(x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t}).$$

As t varies, we obtain a Brownian motion $W(t)$ as the limit of the scaled random walk $W^{(n)}(t)$ as $n \to +\infty$.

To see this, we can apply the central Limit Theorem to the random walk

$$X_{nt} = \sum_{i=1}^{nt} Y_i,$$

where the variables Y_i have common mean 0 and common variance 1.

Then, by the Central Limit Theorem, we have that $\frac{X_{nt}}{\sqrt{nt}}$ converges in distribution to a standard normal distribution, or, in other words, that $\frac{X_{nt}}{\sqrt{n}}$ converges to a normal distribution with mean 0 and variance t.

Sometimes, the scaling will be presented differently in the literature. Consider a time frame of one year, i.e $T = 1$. You may for instance construct a random walk whose step size is $\sqrt{\Delta}$, where Δ is small. So, here, the relationship between Δ and n is

$$\Delta = \frac{1}{n}.$$

In this situation, the number of coin tosses is $\frac{1}{\Delta}$ on the time interval of length one, or in other words, there is one coin toss every Δ unit of time.

7.3 The random walk in continuous-time

In continuous-time, one can define the random walk without drift as

$$X(t) = X(0) + \sigma W(t),$$

where $W(t)$ is the Brownian motion obtained at the limit of the scaled random walk. Here I incorporate a volatility parameter $\sigma > 0$.

A discretized version of the above random walk is

$$X(\Delta) = X(0) + \sqrt{\Delta}\sigma N(0,1)$$

where $N(0,1)$ denotes a normal random variable with mean 0 and variance 1 and Δ is the time step. This model must be coupled with an initial condition

$$X(0) = X_0.$$

So, in this model with a volatility $\sigma > 0$, we have,

$$\mathbb{E}[X(t)] = X(0), var[X(t)] = \sigma^2 t.$$

More generally, we can define the random walk with drift

$$X(t) = X(0) + \mu t + \sigma W(t),$$

where μ is the drift parameter. If, $\mu > 0$, the random walk drifts upward overtime and X is a submartingale, whereas if $\mu < 0$, the random walk drifts downward and X is supermartingale.

Clearly, the distribution of $X(t)$ at time t is normal with mean $X(0) + \mu t$ and variance $\sigma^2 t$. In particular, if $X(0) = 0$, then $X(t)$ has mean μt at time t.

Finally, the merit of this model is to be simple and it can be used for describing the evolution of an asset price over time. However, it is not really suitable for modeling a stock price because the stochastic process $X(t)$ may take negative values, even if μ and X_0 are positive. We will see next week the geometric Brownian motion model which is more adequate for modeling the price of a stock.

Furthermore, we recall that this model does not incorporate any short term memory effect since every new increment is independent from the preceding increment and therefore is not adequate for modeling a phenomenon called *momentum* , which is often observed in practice. This occurs when, say, a stock price, suddenly grows sharply, following the announcement of some news affecting the firm in a positive way and then, keeps growing a while longer because that initial growth spurt made this stock suddenly so attractive that its demand rose, fueling its growth further.

7.4 The Binomial tree model

First, we present a one-period Binomial model which constitutes another example of a stochastic process, called a geometric random walk in discrete time. For a sake of simplicity, we introduce the one-period model first. In this model, the beginning of the time period is called time 0 and the end is called time 1. At time 0, we have, say, a stock whose price per share is denoted by S_0. We assume that S_0 is a positive number. An experiment is performed: one tosses a coin once and obtains either a head with probability p or a tail with probability $1-p$. If a head is obtained, the price of a share of the stock at time 1 will be $S_1(H)$ whereas if a tail is obtained, the price of a share of the stock will be $S_1(T)$. We assume that p and $1-p$ are both positive but they are not necessarily equal. In other words, the coin is not necessarily a fair coin. The up and down factors are denoted by u and d respectively

$$u = \frac{S_1(H)}{S_0}, d = \frac{S_1(T)}{S_0}.$$

$$\begin{tikzcd}[arrows=dash]
 & S_1(H) = S_0 u \\
S_0 \arrow[ur] \arrow[dr] & \\
 & S_1(T) = S_0 d
\end{tikzcd}$$

We assume that $d < u$.

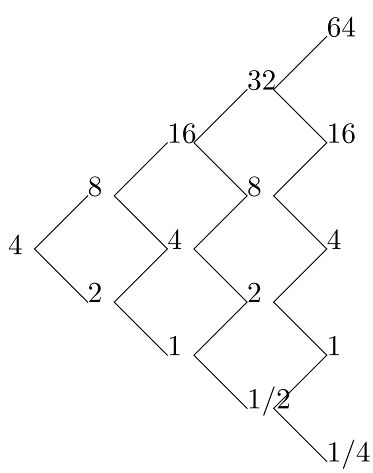
The experiment can be repeated infinitely many times and we can iterate this procedure for determining the next asset price from the previous one. At each step, the next stock price S_{n+1} goes up by a factor u if a head is obtained and goes down by a factor d if a tail is obtained, i.e.

$$S_{n+1}(H) = uS_n,\ S_{n+1}(T) = dS_n.$$

In particular, after the toss number 2, the price at time 2 becomes

$$\begin{aligned} S_2(HH) &= uS_1(H) = u^2S_0, S_2(HT) = dS_1(H) = duS_0 \\ S_2(TH) &= uS_1(T) = duS_0, S_2(TT) = dS_1(T) = d^2S_0 \end{aligned}$$

The third toss has 8 possible outcomes and, given S_2, you can compute the asset price S_3 for each of these outcomes. To illustrate this procedure, we show a plot of a 4-period binomial asset-pricing model for the values $S_0 = 4$, $u = 2$ and $d = 1/2$.



The asset price model we just described, is the Binomial tree model used in Finance. We will study thoroughly in a few weeks the no-arbitrage option pricing theory for an asset modeled using a Binomial tree.

7.5 Rescaled Binomial asset pricing model

Here we rescale the Binomial asset pricing model and pass to the continuous-time limit to derive the geometric Brownian motion model. Here, to simplify, we assume that the probabilities are $p = q = 1/2$, or, in other words, that the coin is fair. At the limit, we derive the geometric random walk without drift.

- We pick n and t such that nt is an integer. We assume that we have n tosses per unit time and hence on the time interval $[0, t]$, we have nt tosses.

- Given that we have n tosses per unit time, we want to determine the asset price $S_n(t)$ at time t.

- the up and down factors are defined by

$$u_n = 1 + \frac{\sigma}{\sqrt{n}}, d_n = 1 - \frac{\sigma}{\sqrt{n}},$$

  where $\sigma > 0$ is the *volatility*.

- If H_{nt} and T_{nt} denote respectively the number of heads and tails in the first nt coin tosses, we have

$$nt = H_{nt} + T_{nt}.$$

- We define the random walk M_{nt} as

$$M_{nt} = H_{nt} - T_{nt}.$$

- We have

$$H_{nt} = \frac{1}{2}(nt + M_{nt}), T_{nt} = \frac{1}{2}(nt - M_{nt}).$$

- We can now write the scaled asset price:

$$S_n(t) = S(0)u_n^{H_{nt}} d_n^{T_{nt}} = S(0)(1 + \frac{\sigma}{\sqrt{n}})^{\frac{1}{2}(nt+M_{nt})}(1 - \frac{\sigma}{\sqrt{n}})^{\frac{1}{2}(nt-M_{nt})}.$$

7.5.1 Central Limit Theorem

As $n \to +\infty$, $S_n(t)$ converges in distribution to the geometric Brownian motion

$$S(t) = S(0)\exp\{\sigma W(t) - \frac{1}{2}\sigma^2 t\}.$$

The distribution of $S(t)$ for fixed t is called *log-normal* because $\log(S(t)/S(0))$ is distributed normally with mean $-\frac{1}{2}\sigma^2 t$ and variance $\sigma^2 t$. To see this, we write

$$\log(\frac{S(t)}{S(0)}) = \sigma W(t) - \frac{1}{2}\sigma^2 t.$$

Since $W(t)$ has a normal distribution with mean 0 and variance t, then, $\log(\frac{S(t)}{S(0)})$ has a normal distribution with mean $-\frac{1}{2}\sigma^2 t$ and variance $\sigma^2 t$. Equivalently, $\log(S(t))$ has a normal distribution with mean $-\frac{1}{2}\sigma^2 t + \log(S(0))$ and variance $\sigma^2 t$.

7.6 The geometric Brownian asset price

$$S(t) = S(0)e^{-\frac{1}{2}\sigma^2 t + \sigma W(t)}, t \in [0, T].$$

The geometric Brownian motion obtained by passing to the continuous-time limit in the Binomial asset pricing model, also called the geometric random walk without drift, is commonly used for modeling stock prices under the risk-neutral measure. Its merits are that the process $S(t)$ stays nonnegative for every time $t \geq 0$ and that it incorporates into the price a multiplicative effect that make it a better fit for historical financial data better over the long term, compared the arithmetic Brownian motion. It has its weaknesses too. In particular, $S(t)$ evolves continuously in time and this model is therefore unable to capture any significant jump in the stock price.

You can add a drift term to the above model which becomes

$$S(t) = S(0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}, t \in [0, T],$$

where $\mu > 0$ is the actual instantaneous mean rate of return of the stock or in other words, the return earned per unit of time. The value of μ, which is notoriously hard to estimate, depends on the risk of the stocks (higher returns should correspond to higher risks) and the level of interest rates in the economy (the higher the interest rates, the higher the expected return).

The coefficient $\sigma > 0$ is called the volatility. It is the proportional change in the standard deviation per unit of time. More accurately, a proportional change in the standard deviation of the stock price over a time interval of length Δt is $\sigma\sqrt{\Delta t}$. Also, $\sigma^2 \Delta t$ is the variance of change in the stock price over the time Δt. Usually, the time unit is *years.* As an example, suppose that $\sigma = 0.3 = 30$ percent per year and the current stock price is 50 dollars. Then the standard deviation of the percentage change in the Stock price in 1 week is $30 \times \frac{1}{\sqrt{52}} = 4.16$ percent. Next, a one-standard deviation move in the stock price in 1 week is $50 \times 0.0416 = 2.08$ dollars.

Taking the log of the left and right hand sides in the above definition of the Geometric Brownian motion with drift, we obtain the log price

$$\log S(t) = \log S(0) + (\mu - \frac{1}{2}\sigma^2)t + \sigma W(t).$$

Furthermore, the distribution of the log price $\log S(t)$ is normal with mean $\log S(0) + (\mu - \frac{1}{2}\sigma^2)t$ and variance $\sigma^2 t$.

Similarly, we can compute the log return of the asset price over the time interval Δt

$$r(t, \Delta t) = \log \frac{S(t + \Delta t)}{S(t)} = (\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma(W(t + \Delta t) - W(t)).$$

It has a normal distribution with mean $(\mu - \frac{1}{2}\sigma^2)\Delta t$ and variance $\sigma^2 \Delta t$.

On the other hand the rate of return is different and defined as

$$
\begin{aligned}
r(t, \Delta t) &= \frac{S(t+\Delta t) - S(t)}{S(t)} \\
&= \frac{S(t)e^{(\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma(W(t+\Delta t) - W(t))} - S(t)}{S(t)} \\
&= e^{(\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma(W(t+\Delta t) - W(t))} - 1.
\end{aligned}
$$

We conclude by adopting a slightly different point of view: we seek the continuously compounded rate of return x earned on a stock between time 0 and time T, that is

$$S(T) = S(0)e^{xT}.$$

Then, using the GBM model for S, one has

$$x = \frac{1}{T}\log\frac{S(T)}{S(0)} = (\mu - \frac{1}{2}\sigma^2) + \frac{1}{T}\sigma W(T).$$

The distribution of x is normal with mean $\mu - \frac{1}{2}\sigma^2$ and standard deviation $\frac{\sigma}{\sqrt{T}}$. So, the expected compounded rate of return is $\mu - \frac{1}{2}\sigma^2$. You may wonder Where the $-\frac{1}{2}\sigma^2$ term come from. This will be made much clearer later, after we study the Brownian motion and the chain rule of Stochastic Calculus.

7.7 Exercises

Problem 1 (15 points)

We consider a time horizon of 1 year and we toss a coin n times per year. We consider time steps of size $\Delta = 1/n$. The probability of landing on heads is $p = \frac{1}{2}(1 + \mu\sqrt{\Delta})$ where $\mu > 0$. We construct the discrete random walk X_i, after i coin tosses, by setting

$$X_0 = 0,\ X_{i+1} - X_i = \begin{cases} \sqrt{\Delta} & \text{if a head is obtained} \\ -\sqrt{\Delta} & \text{if a tail is obtained} \end{cases}$$

1. (2 points) Compute $\mathbb{E}[X_{i+1} - X_i]$.
2. (3 points) Compute $var[X_{i+1} - X_i]$.

3. (10 points) What is the limit distribution of X_n as $n \to +\infty$? Justify your answer.

Chapter 8

Module 8: The Brownian Motion, first hitting time and the gambler's ruin problem in continuous time

8.1 Introduction

In this lecture, I provide more facts about the Brownian Motion. These notes draw heavily from the book by S. Shreve, *Stochastic Calculus for Finance, II.*

First of all, let me recall that I introduced the Brownian motion as the continuous-time limit of the rescaled symmetric random walk. The Brownian motion is a stochastic process in continuous time and with a continuous state space. The set of outcomes of the Brownian motion, Ω, can be seen as the result of an infinite number of coin tosses but the coin is being tossed infinitely fast in the continuous-time limit. In this context, an experiment is the set of all the coin tosses and the outcome of the experiment is ω. To each outcome, corresponds a single sample path. In other words, for ω fixed, $W(t, \omega)$ is a *sample path* of the Brownian motion. We also recall that, as a consequence of the application of the Central Limit Theorem to the rescaled symmetric random walk, the Brownian motion $W(t) = W(t) - W(0)$ is normally distributed with mean 0 and variance t. Next, I am going to present a definition of the Brownian motion and then discuss some of its properties.

8.1.1 Definition of a Brownian motion

We consider a process $\{W(t), t \geq 0\}$ satisfying $W(0) = 0$. $\{W(t), t \geq 0\}$ is said to be a Brownian motion if W is continuous in the variable t, and for all $0 = t_0 < t_1 < t_2 < ... < t_n$, the increments $W(t_1) - W(t_0), W(t_2) - W(t_1), \cdots, W(t_n) - W(t_{n-1})$ are independent, stationary, and $W(t_{j+1}) - W(t_j)$ is normally distributed with mean 0 and variance $t_{j+1} - t_j$ for all $j = 0, \cdots, n-1$.

8.1.2 The Brownian motion as a Gaussian process

We saw earlier, in the introductory lecture on stochastic processes, the definition of a Gaussian process. It turns out, that the Brownian motion is indeed a Gaussian process and, in this subsection, we look at its joint distribution.

For all $0 = t_0 < t_1 < t_2 < ... < t_n$, the random variables $W(t_1), W(t_2), ..., W(t_n)$ are jointly normally distributed. Their joint distribution is characterized by the vector of its means and its covariance matrix.

- The mean of each $W(t_j)$ is 0

$$\mathbb{E}[W(t_j)] = 0.$$

- We want to determine its covariance matrix

$$\begin{bmatrix} \mathbb{E}[W^2(t_1)] & \mathbb{E}[W(t_1)W(t_2)] & ... & \mathbb{E}[W(t_1)W(t_n)] \\ \mathbb{E}[W(t_2)W(t_1)] & \mathbb{E}[W^2(t_2)] & ... & \mathbb{E}[W(t_2)W(t_n)] \\ . & . & . & \\ . & . & . & \\ . & . & . & \\ \mathbb{E}[W(t_n)W(t_1)] & ... & ... & \mathbb{E}[W^2(t_n)] \end{bmatrix}.$$

We have

$$\mathbb{E}[W^2(t)] = t$$

and for $s \le t$,

$$\begin{aligned} & \mathbb{E}[W(s)W(t)] = \mathbb{E}[W(s)(W(t) - W(s)) + W^2(s)] \\ = \quad & \mathbb{E}[W(s)]\mathbb{E}[W(t) - W(s)] + \mathbb{E}[W^2(s)] \\ = \quad & 0 + s = \min(s, t) \end{aligned}$$

This yields the covariance matrix

$$\begin{bmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ t_1 & t_2 & \dots & t_n \end{bmatrix}.$$

8.2 Martingale property, Markov property

The Markov and martingale properties I state in these notes are not very general because I want to avoid defining the concept of filtration in this course. The concept of filtration, which models mathematically the flow of information available at each time, is presented in the course on Stochastic Calculus.

In this course, we say that the Brownian motion satisfies the Markov property if for all $0 \le s \le t$,

$$\mathbb{P}[W(t) \le x | \{W(u) | 0 \le u \le s\}] = \mathbb{P}[W(t) \le x | W(s)]$$

The above property can be interpreted in the following way: when you look at the future distribution of the Brownian motion, conditionally on its past realizations, the only relevant data is the most recent observation instead of the whole path.

Furthermore, we say that the Brownian motion satisfies the martingale property if for all $0 \leq s \leq t$,

$$\mathbb{E}[W(t)|\{W(u), 0 \leq u \leq s\}] = W(s).$$

Also, since the Markov property holds for the Brownian motion, you may rewrite the above relation as

$$\mathbb{E}[W(t)|W(s)] = W(s), \forall 0 \leq s \leq t.$$

I show below that the martingale property holds.

$$\begin{aligned} & \mathbb{E}[W(t)|\{W(u), 0 \leq u \leq s\}] = \mathbb{E}[W(t) - W(s) + W(s)|W(s)] \\ = & \mathbb{E}[W(t) - W(s)|W(s)] + \mathbb{E}[W(s)|W(s)] \\ = & 0 + W(s) = W(s) \end{aligned}$$

8.2.1 Transition density for the Brownian motion

Since the Brownian motion is a Markovian process, it can be characterized alternately by its initial condition $W(0) = 0$, together with its transition density given for $x, y, s \leq t$ by

$$p(y, x, t, s) = \frac{1}{\sqrt{2\pi(t-s)}} \exp(-\frac{(y-x)^2}{2(t-s)}).$$

This is the probability density in the variable y for the random variable $W(t)$, conditioned by the event $W(s) = x$. The transition density for a Markov process in continuous time with a continuous state space is the analogue of the transition probability matrix for Markov chains in discrete time with a discrete state space.

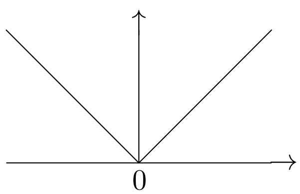
8.3 On the regularity of the Brownian motion

- Each sample path of the Brownian motion $W(t, \omega)$ is a continuous function of time. It does not have jumps!

- However the sample paths of the Brownian motion cannot be differentiated with respect to the time variable at any point! Its derivative is not continuous. A sample path of the Brownian motion does not look smooth and curvy. It has *kinks*.

- Example of a function which does not have a continuous derivative at 0:

$$f(t) = |t|, \text{ for } t \in \mathbb{R}.$$



However, the paths of the Brownian motion look even less smooth than this example. In fact, a Brownian motion is not even Lipschitz continuous unlike the absolute value. If it were Lipschitz continuous, it would be differentiable almost everywhere, which is not the case. We recall, for a sake of completeness the definition of a Lipschitz continuous function f.

Definition:

A continuous function f defined on the real line is said to be Lipschitz continuous if there is a constant C such that, for all $x, y \in \mathbb{R}$,

$$|f(x) - f(y)| < C|x - y|.$$

8.4 Quadratic variation for the Brownian motion

It turns out that the Brownian motion has the same quadratic variation as the random walk and the scaled random walk. Here, we just state this fact without any proof.

For all $T \geq 0$,

$$[W,W](T) = T \textit{ almost surely} .$$

The natural question is how the quadratic variation is defined in general. So far, we only defined the quadratic variation in 2 particular cases, the symmetric random walk in discrete time and the rescaled random walk in continuous time.

To this end, one need to consider a general partition of the time interval $[0,T]$, say, $0 = t_0 < t_1 < \cdots < t_j < t_{j+1} < \cdots < t_n = T$, where n is an positive integer. I also denote by h the size of the largest subinterval, i.e.

$$h = \max_{i=1\cdots n} |t_j - t_{j-1}|.$$

We can then define the quadratic variation of any function f of the variable t by setting

$$[f,f](T) = \lim_{h\to 0,\, n\to+\infty} \sum_{i=1}^{n} (f(t_j) - f(t_{j-1}))^2.$$

Now, for a stochastic process, the quadratic variation is computed pathwise, i.e. for any given fixed outcome ω and hence is itself a stochastic process in general.

In addition, for a sake of completeness, I recall the interpretation of *almost surely*: this means that the statement is true for almost all the sample paths except for a set of sample paths. This set of sample paths for which the equality is wrong has a probability 0. It is worth mentioning that any function which has continuous derivatives has a quadratic variation equal to 0!. The non-zero quadratic variation of the Brownian motion comes from the fact that the Brownian motion is not differentiable.

Finally, we note that the quadratic variation of the Brownian motion is deterministic, i.e. does not depend on the particular outcome ω and is equal to its variance.

8.5 Hitting time

In this section, we define the first time the Brownian motion hits the point a. For instance, we assume that $a > 0$. We then denote the first hitting time by τ_a. It is defined by

$$\tau_a = \inf\{t \geq 0, W(t) \geq a\}.$$

Next, it is natural to ask how we can compute the distribution of τ_a, $\mathbb{P}[\tau_a \leq t]$. This computation is based on the *reflection principle*. First of all, we can write

$$\mathbb{P}[W(t) \geq a] = \mathbb{P}[W(t) \geq a|\tau_a \leq t]\mathbb{P}[\tau_a \leq t] + \mathbb{P}[W(t) \geq a|\tau_a > t]\mathbb{P}[\tau_a > t].$$

Then, the send term in the right hand side is necessarily equal to 0 because if the Brownian motion has not hit a yet by time t ($\tau_a > t$), it cannot be above a.

Next, when $\tau_a \leq t$, the process hits the point a at some time smaller than or equal to t and by the so-called *reflection principle* , we have

$$\mathbb{P}[W(t) \geq a|\tau_a \leq t] = \frac{1}{2}.$$

In plain english, this means that, knowing that the Brownian motion hit a before the time t, the Brownian motion is as likely to be above a than below a at time t.

Consequently, we can deduce the distribution of the first hitting time for any $a > 0$.

$$\begin{aligned}
\mathbb{P}[\tau_a \leq t] &= 2\mathbb{P}[W(t) \geq a] \\
&= \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/2t} dx \\
&= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^\infty e^{-y^2/2} dy \\
&= 2(1 - \Phi(\tfrac{a}{\sqrt{t}})),
\end{aligned}$$

where Φ is the standard normal cumulative distribution function.

Finally, we can argue that, by symmetry, the distribution of τ_a for $a \leq 0$ is the same as the distribution of τ_{-a} and this implies that, in general,

$$\mathbb{P}[\tau_a \leq t] = 2(1 - \Phi(\frac{|a|}{\sqrt{t}})).$$

To see this, we write

$$\tau_a = \inf\{t \geq 0, W(t) \leq a\} = \inf\{t \geq 0, -W(t) \geq -a\},$$

and we argue that $-W(t)$ is also a standard Brownian motion.

We can also compute the density function of the first hitting time by differentiating the above formula:

$$f_{\tau_a}(t) = |a|\Phi'(\frac{|a|}{\sqrt{t}})t^{-3/2}.$$

This constitutes a very useful tool in Finance and can be used for instance to build a gambler's ruin model in continuous time. Suppose that the wealth of a gambler is modeled using the arithmetic Brownian motion without drift

$$X(t) = X(0) + \sigma W(t),$$

where $X(0)$ is the gambler's initial capital and $\sigma > 0$ is a volatility coefficient.

We can define the time of ruin as the hitting time

$$\begin{aligned} \tau_0 &= \inf\{t \geq 0, X(t) \leq 0\} \\ &= \inf\{t \geq 0, W(t) \leq -\frac{X(0)}{\sigma}\} \end{aligned}$$

and deduce the distribution of τ_0. It is worth mentioning that this type of computation is very commonly used in the industry in practice in the area of credit risk.

$$
\begin{aligned}
\mathbb{P}[\tau_0 \leq t] &= \frac{2}{\sqrt{2\pi}} \int_{|X(0)|/(\sigma\sqrt{t})}^{\infty} e^{-y^2/2} dy. \\
&= 2\big(1 - \Phi(\frac{|X(0)|}{\sigma\sqrt{t}})\big).
\end{aligned}
$$

Finally, it is worth mentioning that the *reflection principle* and the first hitting time are very important for the application in Finance, in particular in credit risk and option pricing. In option pricing, the reflection principle is used for pricing a Barrier option, for instance an up and out call (the call becomes worthless when the stock price rises above a prescribed barrier B which is greater than the strike K). The first step consist in writing the payoff of such an option as a function of the maturity date. This uses the distribution of the maximum to date of the Brownian motion

$$
M(t) = \max_{0 \leq s \leq t} W(s),
$$

which is determined by using the reflection principle.

In credit risk, Merton and subsequently Black-Cox developed a model for pricing equity and Bonds that have a risk of default. In this model, defaulting occurs when the value of the asset of the company, $A(t)$, usually modeled as a geometric Brownian motion, falls below a prescribed barrier. The case of a geometric Brownian motion is more complicated than the case of an arithmetic Brownian motion and I won't cover it in this lecture.

8.6 Application to the geometric Brownian motion

I introduced a couple weeks ago the geometric Brownian motion as the limit of the rescaled multi-period Binomial tree model and I explained how it was used for modeling an asset price.

This time, I complement this study by computing the expectation a geometric Brownian motion because tis is a very standard computation that you ought to know.

I consider the arithmetic Brownian motion with drift

$$Y(t) = (\mu - \frac{1}{2}\sigma^2)t + \sigma W(t),$$

where $W(t)$ is a standard Brownian motion and the geometric Brownian motion

$$X(t) = e^{Y(t)}.$$

We compute

$$\begin{aligned}
&\mathbb{E}[X(t)] = \mathbb{E}[e^{(\mu-\frac{1}{2}\sigma^2)t+\sigma W(t)}] \\
&= e^{(\mu-\frac{1}{2}\sigma^2)t}\mathbb{E}[e^{\sigma W(t)}] \\
&= e^{(\mu-\frac{1}{2}\sigma^2)t}\mathbb{E}[e^{\sigma\sqrt{t}Y}]
\end{aligned}$$

where Y is a standard normal random variable. Next, we calculate

$$\begin{aligned}
&e^{(\mu-\frac{1}{2}\sigma^2)t}\mathbb{E}[e^{\sigma\sqrt{t}Y}] \\
&= e^{(\mu-\frac{1}{2}\sigma^2)t}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{\sigma\sqrt{t}x}e^{-x^2/2}dx \\
&= e^{(\mu-\frac{1}{2}\sigma^2)t}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\sigma\sqrt{t})^2}e^{\frac{1}{2}\sigma^2 t}dx \\
&= e^{\mu t}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\sigma\sqrt{t})^2}dx \\
&= e^{\mu t}
\end{aligned}$$

Consequently, when we look now at

$$S(t) = se^{Y(t)},$$

then we clearly have $S(0) = s$ and

$$\mathbb{E}[S(t)|S(0) = s] = se^{\mu t}.$$

We can also generalize and write for every θ, t, with $\theta \leq t$

$$\mathbb{E}[S(t)|\{S(u)|0 \leq u \leq \theta\}] = S(\theta)e^{\mu(t-\theta)}.$$

8.7 Exercises

Question 1 (5 points)

Consider the process $X(t) = \mu t + \sigma W(t)$, where W is a standard Brownian motion.

1. (5 points) What is the conditional distribution of $X(t)$ given that $X(s) = c$ when $s < t$ (c is a given constant)?

Question 2 (16 points)

Consider an investor who is holding one share of a stock whose price per share is given by

$$S(u) = S(0) + \sigma W(u), u \geq 0,$$

where $\sigma > 0$ is constant and W is a standard Brownian motion. This investor purchased the stock at a price $S(0) > 0$ at time 0 and decides to sell the stock whenever it falls below the price $S(0)(1-\mu)$ for the first time, where $0 < \mu < 1$ is a constant.

1. (4 points) What is the distribution of $S(u)$? Give its mean and variance.
2. (4 points) Is the process S a martingale? Justify your answer.
3. (4 points) What is the cumulative distribution function of the hitting time $\tau_{S(0)(1-\mu)}$ of the process S?
4. (4 points) Give also the density function of the distribution of the hitting time $\tau_{S(0)(1-\mu)}$ of the process S.

Question 3 (20 points)

We consider a standard Brownian motion W.

1. (5 points) Is the process $t \in [0, +\infty) \to W(ct^2)$, where c is a positive constant, a standard Brownian motion? Justify your answer.

2. (5 points) Is $t \in [0, +\infty) \to \sqrt{t}W(1)$ a standard Brownian motion? Justify your answer.

3. (5 points) Compute the mean and autocovariance function of the process
$$X(t) = \int_0^t W(s)ds.$$

4. Consider the process $S(t) = \mu t + \sigma W(t)$, for $t \in [0, T]$, where $\mu, \sigma > 0$ are both positive constant. Compute $\mathbb{P}[S(T) < 0]$.

Question 4 (20 points)

We consider a standard Brownian motion W. We also consider the maximum to date of the Brownian motion,
$$M(t) = \max_{0 \leq s \leq t} W(s).$$

1. (4 points) Show that
$$\mathbb{P}[\max_{0 \leq s \leq t} W(s) \geq m] = \mathbb{P}[\tau_m \leq t],$$
where τ_m denotes the first time the Brownian motion W hits the level m.

2. (4 points) Deduce the cumulative function of the randon variable $M(t)$.

3. (4 points) Deduce the density function of $M(t)$.

4. (8 points) Consider now the partition of the time interval $[0, t]$,
$$0, \frac{t}{n}, \frac{2t}{n}, \cdots, \frac{jt}{n}, \cdots \frac{nt}{n} = t,$$
for some integer $n > 1$, and define the discrete-time moving average process
$$X_n = \frac{\sum_{j=n-k}^{n} W(\frac{jt}{n})}{k+1}, \text{ for some } n \geq k.$$
Compute the expectation and the variance of X_n.

Question 5 (20 points)

1. (5 points) Let Y be a standard normal random variable. For all $t \geq 0$, let $X(t) = \sqrt{t}Y$. Is X a standard Brownian motion? Justify.

2. (5 points) Let W be a standard Brownian motion. Show that $W^2(t) - t$ is a martingale.

3. (5 points) Consider a standard Brownian motion W. Show that the following process is a standard Brownian motion:

$$B(t) = cW(t/c^2) \text{ for all } t \geq 0.$$

4. (5 points) Consider a standard Brownian motion W, and the process

$$B(0) = 0, B(t) = tW(\frac{1}{t}), \text{ for all } t > 0.$$

Show that B is a standard Brownian motion by using the definition.

Chapter 9

Module 9: A basic introduction to stochastic calculus and its application to the modeling of dynamic asset prices

9.1 Introduction

The aim of this lecture is to introduce in an elementary manner the most basic concepts of stochastic calculus and their application to the modeling of dynamic asset prices. I draw most of the material from Shreve, *Stochastic Calculus for finance*, II.

9.2 The Ito integral

The goal in this section is to define the so-called Ito integral, which is a stochastic integral of the form

$$\int_0^t \sigma(u)dW(u),$$

where $W(t)$ is a standard Brownian motion and $\sigma(u)$ is a stochastic process.

In addition, we observe as time passes the realizations of the Brownian motion and gather this information. We assume that the information available at time t is sufficient to know σ with certainty up to time t, for each $t \in [0, T]$. We only define the stochastic integral for integrands satisfying this condition. This means that the standard Brownian motion is the only source of uncertainty and that there is no additional source of uncertainty that is introduced through $\sigma(u)$.

However, we know that trajectories of Brownian motion **are not** time differentiable, and so in particular we cannot make sense of the integral with a Brownian motion as a Riemann integral.

Hence we need to give a meaning to integrals of the type $\int \sigma(t)dW(t)$, which is explained in the construction below.

9.2.1 Construction of the integral

We describe briefly the procedure for defining a stochastic integral: first, we define the integral for piecewise constant integrands σ and then pass to the limit to extend it to general integrands.

Consider a partition of $[0, T]$, $0 = t_0 < t_1 < t_2 < .... < t_n = T$ and assume that the integrand σ is constant on each subinterval $[t_j, t_{j+1})$, i.e.

$$\sigma(t) = \sigma(t_j) \text{ for all } t \in [t_j, t_{j+1}).$$

We define

$$\begin{aligned}
I(t) &= \sigma(t_0)[W(t) - W(t_0)] = \sigma(0)W(t), \text{ for all } t \in [0, t_1] \\
I(t) &= \sigma(0)W(t_1) + \sigma(t_1)[W(t) - W(t_1)], \text{ for all } t \in [t_1, t_2] \\
I(t) &= \sigma(0)W(t_1) + \sigma(t_1)[W(t_2) - W(t_1)] + \sigma(t_2)[W(t) - W(t_2)], \\
&\quad \text{for all } t \in [t_2, t_3]
\end{aligned}$$

For $t \in [t_k, t_{k+1}]$,

$$I(t) = \sum_{j=0}^{k-1} \sigma(t_j)[W(t_{j+1}) - W(t_j)] + \sigma(t_k)[W(t) - W(t_k)].$$

A financial application of the Ito's integral

- $W(t)$ represents the price per share of a an asset at time t. Warning: the Brownian motion may take negative values and is not really appropriate for modeling the price of a stock.
- The partition $0 = t_0 < t_1 < ... < t_n$ represent the trading dates or times.
- $\sigma(t_0)$,$\sigma(t_1), ..., \sigma(t_n)$ are the positions in the asset (number of shares) at each trading date and held until the next trading date.
- $I(t)$ is the accumulated gain/loss from trading up to time t

Let $\sigma(t)$ be a process satisfying the condition stated earlier, and satisfying the technical square integrability condition

$$\mathbb{E} \int_0^T \sigma(u)^2 dt < \infty.$$

We define a sequence $\sigma_n(t)$ of piecewise constant processes which converges to $\sigma(t)$ as $n \to \infty$, i.e.

$$\lim_{n \to +\infty} \mathbb{E} \int_0^T |\sigma_n(t) - \sigma(t)|^2 dt = 0.$$

We define the Ito's integral by setting

$$\int_0^t \sigma(u) dW(u) = \lim_{n \to \infty} \int_0^t \sigma_n(u) dW(u),$$

where the limit is meant in the mean squares sense.

9.2.2 Properties of the Ito's integral

We simply list below without any proof some useful properties of the Ito integral.

- $I(t)$ is continuous in t.

- The information available at time t is sufficient to know I with certainty up to time t, for each $t \in [0, T]$
- The Ito's integral is linear, i.e.
$$\int_0^t [c_1\sigma_1(t) + c_2\sigma_2(t)]dW(t) = c_1 \int_0^t \sigma_1(t)dW(t) + c_2 \int_0^t \sigma_2(t)dW(t).$$
- $I(t)$ is a martingale.
- Ito's isometry
$$\mathbb{E}I^2(t) = \mathbb{E}\int_0^t \sigma^2(u)du.$$
- Quadratic variation
$$[I, I](t) = \int_0^t \sigma^2(u)du.$$

9.2.3 An additional result

We complete this section by stating a result of interest for the application to modeling interest rate that we cover in these notes. We look at the particular case when the integrand is simply deterministic.

Theorem Let
$$I(t) = \int_0^t \sigma(s)dW(s)$$
where $\sigma(s)$ is a deterministic function of time.

Then $I(t)$ is normally distributed with mean 0 and variance $\int_0^t \sigma^2(s)ds$.

Furthermore it is a Gaussian process.

9.3 Some well-known facts of stochastic calculus

We saw in an earlier lecture that the quadratic variation of the Brownian motion was equal to the time. This can be rewritten in differential form
$$dW(t)dW(t) = dt.$$

Since the quadratic variation of a differentiable function is 0, we also have

$$dtdt = 0.$$

Finally, it is another well-known fact of stochastic calculus that the cross-variation of a Brownian motion with time is 0, i.e.

$$dW(t)dt = 0.$$

We will need these facts in the next sections of this lecture.

9.4 The chain rule of stochastic calculus

Sometimes, we need to evaluate quantities of the form $f(W(t))$ where $W(t)$ is a Brownian motion and f is a differentiable function. In ordinary calculus, we would have by the chain rule

$$\frac{d}{dt}f(W(t)) = f'(W(t))W'(t)$$

or equivalently, in differential notation,

$$df(W(t)) = f'(W(t))dW(t).$$

However this is not the correct formula for the Brownian motion because its quadratic variation is not equal to 0 and hence, there is an extra-term in the chain rule, i.e. the chain rule for a Brownian motion actually is

$$df(W(t)) = f'(W(t))dW(t) + \frac{1}{2}f''(W(t))dt.$$

The above formula is called Ito's formula. It should be interpreted as the differential form of the Ito's formula in integral form

$$f(W(t)) = f(0) + \int_0^t f'(W(s))dW(s) + \int_0^t \frac{1}{2}f''(W(s))ds,$$

where $\int_0^t f'(W(s))dW(s)$ is an Ito stochastic integral.

It can be generalized to functions who also depend on the time variable

$$df(t, W(t)) = f_t(t, W(t))dt + f_x(t, W(t))dW(t) + \frac{1}{2} f_{xx}(t, W(t))dt,$$

or to more general processes called Ito processes. They are of the form

$$X(t) = X(0) + \int_0^t \theta(s)ds + \int_0^t \sigma(s)dW(s),$$

where σ and θ are stochastic processes, which are known with certainty up to time t, if the standard Brownian motion is known up to time t, for each time $t \in [0, T]$.

Ito's formula for the process X is then given by

$$df(t, X(t)) = f_t(t, X(t))dt + f_x(t, X(t))dX(t) + \frac{1}{2} f_{xx}(t, X(t))dX(t)dX(t).$$

Note that the term $dX(t)dX(t)$ is the differential of the quadratic variation of the Ito process and that we have

$$dX(t)dX(t) = \sigma^2(t)dW(t)dW(t) = \sigma^2(t)dt.$$

Finally, Ito's formula may also be written in the integral form on the interval $[0, t]$:

$$\begin{aligned} f(t, X(t)) \quad &= \quad f(0, X(0)) + \int_0^t f_t(s, X(s))ds + \int_0^t f_x(s, X(s))dX(s) + \\ &\int_0^t \frac{1}{2} f_{xx}(s, X(s))\sigma^2(s)ds, \end{aligned}$$

where

$$\int_0^t f_x(s, X(s))dX(s) = \int_0^t f_x(s, X(s))\theta(s)ds + \int_0^t f_x(s, X(s))\sigma(s)dW(s).$$

9.4.1 Ito's formula in 2 dimensions

Ito's formula can be generalized to two spatial dimensions and beyond. Take two Ito processes X, Y driven by the independent standard Brownian motions W_1, W_2, defined as

$$X(t) = X(0) + \int_0^t \theta_1(s)ds + \int_0^t \sigma_{11}(s)dW_1(s) + \int_0^t \sigma_{12}(s)dW_2(s),$$
$$Y(t) = Y(0) + \int_0^t \theta_2(s)ds + \int_0^t \sigma_{21}(s)dW_1(s) + \int_0^t \sigma_{22}(s)dW_2(s).$$

Ito's formula reads

$$df(t, X(t), Y(t)) = f_t dt + f_x dX(t) + f_y dY(t) + \frac{1}{2} f_{xx} dX(t)dX(t) + \frac{1}{2} f_{yy} dY(t)dY(t) + f_{xy} dX(t)dY(t),$$

where

$$\begin{aligned} dX(t)dX(t) &= (\sigma_{11}^2 + \sigma_{12}^2)dt, \\ dY(t)dY(t) &= (\sigma_{21}^2 + \sigma_{22}^2)dt, \\ dX(t)dY(t) &= (\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22})dt. \end{aligned}$$

A consequence of Ito's formula in two dimensions is the well-known product rule

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t) + dX(t)dY(t).$$

9.5 The geometric Brownian asset price

$$S(t) = S(0)e^{-\frac{1}{2}\sigma^2 t + \sigma W(t)}, t \in [0, T].$$

We recall that the geometric Brownian motion obtained by passing to the continuous-time limit in the Binomial asset pricing model is commonly used for modeling stock prices under the risk-neutral measure. We saw that its merit is that the process $S(t)$ stays nonnegative for every time $t \geq 0$. It has its weaknesses too. In particular, $S(t)$ evolves continuously in time and this model is therefore unable to capture jumps in the stock price. In addition, it is a Markov process and thus, does not have any memory and therefore is not adequate for capturing a *momentum* phenomenon.

Furthermore, the Stochastic Differential Equation that $S(t)$ satisfies can be derived rigorously by using the chain rule of stochastic calculus. I can show you that the correct Stochastic Differential Equation for the evolution of $S(t)$ is given by

$$dS(t) = \sigma S(t) dW(t).$$

Indeed, consider the function $f(t,x) = S(0)e^x$ and the process $X(t) = -\frac{1}{2}\sigma^2 t + \sigma W(t)$. We first compute the derivatives of f

$$f_t(t,x) = 0, f_x(t,x) = S(0)e^x, f_{xx}(t,x) = S(0)e^x.$$

Hence, by the Ito-Doeblin formula, we have

$$\begin{aligned} dS(t) &= df(t, X(t)) = 0dt + S(0)e^{X(t)}dX(t) + \frac{1}{2}S(0)e^{X(t)}dX(t)dX(t) \\ &= S(t)(-\frac{1}{2}\sigma^2 dt + \sigma dW(t)) + \frac{1}{2}S(t)\sigma^2 dt \\ &= S(t)\sigma dW(t) \end{aligned}$$

The above equation is actually the Stochastic Differential Equation for a stock price in the particular case when $\mu = 0$. In the more general case with drift, the stock price satisfies the following SDE with a drift term

$$dS(t) = \mu S(t) dt + \sigma S(t) dW(t).$$

The closed form solution for this more general SDE is

$$S(t) = S(0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}, t \in [0,T].$$

As in the discrete Binomial asset pricing model, the instantaneous mean rate of return of the stock under the risk-neutral measure is again here the risk free rate.

Under the risk-neutral probability measure, the stock satisfies the SDE

$$dS(t) = rS(t)dt + \sigma S(t)d\tilde{W}(t),$$

where r is the risk free rate and $\tilde{W}$ is a standard Brownian motion under the risk-neutral probability measure. I do not explain in this introductory course the detailed mechanisms of the change of measure in continuous-time.

9.5.1 Application of the quadratic variation of the Brownian motion: estimating volatility from market data

We saw earlier, that it was basically impossible to estimate the parameter μ from historical data and this is an issue that portfolio managers must live with. However, estimating the historical volatility is much easier in the sense that it does not requires vey lengthy data sets. Generally, we can estimate the historical volatility from historical market data. A good estimate for the historical volatility is provided by the following procedure, which relies on the well-known facts of stochastic calculus

$$dW(t)dW(t) = t, dtdt = 0, dtdW(t) = 0$$

that we saw earlier.

We pick a partition of the interval $[0, T]$ in n subintervals of equal length, $0 = t_0 \leq t_1 \leq t_2 ... \leq t_j \leq t_{j+1} ... \leq t_n = T$ with $t_j = \frac{jT}{n}$. We define $\Delta t = T/n$. Suppose that we have available $n+1$ observations for the stock price $S_0, S_1, ...S_j, ..., S_n$ where S_j is the stock price in the interval $[t_j, t_{j+1})$.

We can write the *log returns* of the stock S as

$$\log \frac{S_{j+1}}{S_j} = \sigma(W(t_{j+1}) - W(t_j)) + (\mu - \frac{1}{2}\sigma^2)\Delta t.$$

Next, we square the left and right-hand side

$$(\log \frac{S_{j+1}}{S_j})^2 = \left(\sigma(W(t_{j+1}) - W(t_j)) + (\mu - \frac{1}{2}\sigma^2)\Delta t\right)^2.$$

We sum over all the indices

$$\sum_{j=0}^{n-1}(\log \frac{S_{j+1}}{S_j})^2 = \sum_{j=0}^{n-1} \left(\sigma(W(t_{j+1}) - W(t_j)) + (\mu - \frac{1}{2}\sigma^2)\Delta t\right)^2,$$

and expand the right-hand side

$$\begin{aligned}
\sum_{j=0}^{n-1}(\log \frac{S_{j+1}}{S_j})^2 = \sum_{j=0}^{n-1} \Big[\sigma^2(W(t_{j+1}) - W(t_j))^2 + \\
2(\mu - \frac{1}{2}\sigma^2)\Delta t(W(t_{j+1}) - W(t_j)) + \\
(\mu - \frac{1}{2}\sigma^2)^2 \Delta t^2\Big] \\
= \quad \sigma^2 \sum_{j=0}^{n-1}(W(t_{j+1}) - W(t_j))^2 + \\
2(\mu - \frac{1}{2}\sigma^2) \sum_{j=0}^{n-1} \Delta t(W(t_{j+1}) - W(t_j)) + \\
(\mu - \frac{1}{2}\sigma^2)^2 \sum_{j=0}^{n-1} \Delta t^2.
\end{aligned}$$

Then we find, using the quadratic variation of the Brownian motion together with the other properties

$$dtdt = 0, dW(t)dt = 0,$$

that

$$\sum_{j=0}^{n-1}(\log \frac{S_{j+1}}{S_j})^2 \approx \sigma^2 T.$$

This provides us with a estimate of the volatility that does not require a prior estimation of the parameter μ. It is worth mentioning that this is not the only estimator available.

9.6 The Vasicek interest rate model

The model below is another example of a Stochastic Differential Equation that is used in Finance. It constitutes a possible model for the interest rate (short rate) or for the spread of two co-integrated assets.

$$dR(t) = (\alpha - \beta R(t))dt + \sigma dW(t)$$

where α, β, σ are positive constants. This is another example of a stochastic differential equation.

We are going to verify that has the closed-form solution

$$\begin{aligned} R(t) &= \exp(-\beta t)R(0) + \frac{\alpha}{\beta}(1 - \exp(-\beta t)) \\ &+ \sigma \exp(-\beta t) \int_0^t \exp(\beta s) dW(s) \end{aligned}$$

using the Ito's formula. We apply the Ito formula to $f(t, X(t))$ where

$$X(t) = \int_0^t \exp(\beta s) dW(s)$$

and

$$f(t, x) = \exp(-\beta t)R(0) + \frac{\alpha}{\beta}(1 - \exp(-\beta t)) + \sigma \exp(-\beta t)x.$$

Note that

$$\begin{aligned} f_x(t, x) &= \sigma \exp(-\beta t) \\ f_{xx}(t, x) &= 0 \\ f_t(t, x) &= -\beta \exp(-\beta t)R(0) + \frac{\alpha}{\beta}\beta \exp(-\beta t) - \beta\sigma \exp(-\beta t)x \\ &= \alpha - \beta f(t, x). \end{aligned}$$

By Ito formula

$$\begin{aligned} dR(t) &= df(t, X(t)) \\ &= f_t(t, X(t))dt + f_x(t, X(t))dX(t) + \frac{1}{2} f_{xx}(t, X(t))dX(t)dX(t) \\ &= (\alpha - \beta f(t, X(t)))dt + \sigma \exp(-\beta t)dX(t) \\ &= (\alpha - \beta R(t))dt + \sigma dW(t). \end{aligned}$$

We recall the solution of the Vasicek model

$$\begin{aligned} R(t) &= \exp(-\beta t)R(0) + \frac{\alpha}{\beta}(1 - \exp(-\beta t)) \\ &+ \sigma \exp(-\beta t) \int_0^t \exp(\beta s) dW(s). \end{aligned}$$

The integral above is an Ito integral with a deterministic integrand. Hence it is normally distributed with mean 0 and variance

$$\int_0^t \exp(2\beta s)ds = \frac{1}{2\beta}(\exp(2\beta t) - 1).$$

So R is normally distributed with mean $\exp(-\beta t)R(0) + \frac{\alpha}{\beta}(1 - \exp(-\beta t))$ and variance $\frac{\sigma^2}{2\beta}(1 - \exp(-2\beta t))$.

The Vasicek model is said to be *mean-reverting*. This means that, for t large, the mean of the process $R(t)$ tends to the ratio $\frac{\alpha}{\beta}$. Furthermore, when $R(t) = \frac{\alpha}{\beta}$, the drift term becomes 0. In particular, if one starts with $R(0) = \frac{\alpha}{\beta}$ then the mean of $R(t)$ is $\frac{\alpha}{\beta}$ for all times. When $R(t) > \frac{\alpha}{\beta}$, the drift term is negative, which pushes $R(t)$ back toward $\frac{\alpha}{\beta}$. When $R(t) < \frac{\alpha}{\beta}$, the drift term is positive, which also pushes $R(t)$ up toward $\frac{\alpha}{\beta}$.

Finally, it is worth pointing out a drawback of the Vasicek model. Even when $R(0)$ is strictly positive, $R(t)$ may wander in negative territory at times. One remedy consists in replacing the Vasicek model by the nonlinear Cox-Ingersoll-Ross interest rate model

$$dR(t) = (\alpha - \beta R(t))dt + \sigma\sqrt{R(t)}dW(t),$$

where α, β and σ are positive constant. In this latter model, roughly speaking, whenever $R(t)$ reaches 0, the volatility term is equal to zero as well and the evolution is only dictated by the drift which pushes $R(t)$ back into the positive territory. However, unlike the Vasicek model, the above stochastic differential equation does not have a analytical solution and must be approximated numerically.

The mean-reverting models are very useful for a variety of applications. I mentioned earlier that they are used in particular for modeling the short rate. They are also used for modeling the price of commodities, such as gold, coffee, oil, gas and electricity. Last but not least, they can model the spread of two co-integrated stocks to which pairs trading strategies can be applied with the goal of generating abnormally high returns, that cannot be simply explained by risk being taken and by the transaction costs incurred.

9.7 Time-discretization

Consider first the arithmetic Brownian motion with drift μ:

$$X(t) = X(0) + \mu t + \sigma dW(t).$$

We can alternately write this model in differential form

$$dX(t) = \mu dt + \sigma dW(t).$$

You can discretize the above Stochastic Differential Equation by approximating it using the Euler scheme

$$X(t + \Delta t) = X(t) + \mu \Delta t + \sqrt{\Delta t}\sigma N(0, 1),$$

where $N(0, 1)$ is a sample drawn from the standard normal distribution, and Δt is the time step.

The truncation error of this Euler scheme is of order $\sqrt{\Delta}$.

We turn next to the Geometric Brownian motion model with drift μ

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t).$$

The above SDE can be discretized and its approximation by an Euler scheme reads

$$\frac{S(t + \Delta t) - S(t)}{S(t)} \approx \mu \Delta t + \sigma \sqrt{\Delta t} N(0, 1)$$

where $N(0, 1)$ denotes a standard normal random variable, and Δt is the time step.

We approximated the rate of return over the interval $[t, t + \Delta t]$ by a normal distribution with mean $\mu \Delta t$ and variance $\sigma^2 \Delta t$.

9.8 Exercises

Question 1 (15 points)

1. (5 points) Compute the differential of $W^3(t)$, where W is a standard Brownian motion.
2. (4 points) Integrate the above formula on $[0, T]$.
3. (6 points) Take the expectation of the left and right-hand sides and deduce $\mathbb{E}[W^3(T)]$.

Question 2 (15 points)

Your nominal income in us dollars $X(t)$, at time t, satisfies the dynamics

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t),$$

where W is a standard Brownian motion. Inflation is represented by the variable Y satisfying the dynamics

$$dY(t) = \gamma Y(t)dt + \delta Y(t)dB(t),$$

where B is a standard Brownian motion which is independent of W. We define the variable $Z = \frac{X}{Y}$, which represents your real income.

1. (5 points) Use Ito's formula to compute the differential of $d\left(\frac{1}{Y(t)}\right)$.
2. (10 points) Compute the differential of Z by using the product rule and derive the SDE for Z by substituting Z for $\frac{X}{Y}$ in the right-hand side of dZ.

Question 3 (15 points)

Consider the process

$$X(t) = \int_0^t \sigma(s)dW(s),$$

where σ is a deterministic function of t. Furthermore, we define

$$Y(t) = e^{iuX(t)},$$

where i is the complex number, such that $i^2 = -1$.

1. (2 points) Apply Ito's formula to compute $dY(t)$.
2. (2 points) Deduce the SDE satisfied by Y.
3. (3 points) Rewrite the above SDE in integral form, on the interval $[0, t]$.
4. (2 points) Take the expectation of the left- and right-hand sides.
5. (2 points) We define $m(t) = \mathbb{E}[Y(t)]$. Derive the ODE satisfied by m
6. (4 points) Solve the ODE and conclude that

$$m(t) = \exp\{-\frac{u^2}{2}\int_0^t \sigma^2(s)ds\}.$$

Question 4 (20 points) We consider the integrated Brownian motion defined by the Rieman integral

$$Z(t) = \int_0^t W(u)du,$$

where W is a standard Brownian motion.

1. (4 points) Compute the expectation of Z.
2. (9 points) Compute the autocovariance of Z.
3. (5 points) We can define Z as

$$Z(t) = \lim_{n\to+\infty} \frac{t}{n}\sum_{k=1}^{n} W(\frac{tk}{n}).$$

   We denote by S_n

$$S_n = \frac{t}{n}\sum_{k=1}^{n} W(\frac{tk}{n}).$$

   Show that S_n is a sum of independent and normally distributed variables.

4. (2 points) Is S_n normally distributed (justify)? You are not asked to compute the moments of S_n.

Question 5 (10 points) (Drawn from T. Bjork, "Arbitrage theory in continuous time"

Suppose that the stochastic process X satisfies the Stochastic Differential Equation

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t),$$

where W is a standard Brownian motion. Define $Z(t) = \frac{1}{X(t)}$.

1. (10 points) Compute $dZ(t)$ by using Ito's formula and derive the SDE satisfies by Z.

Question 6 (20 points)

Consider the Stochastic Differential Equation

$$dY(t) = (2\mu Y(t) + \sigma^2)dt + 2\sigma\sqrt{Y(t)}dW(t),$$

where W is a standard Brownian motion, μ is a constant, and $\sigma > 0$ is a constant.

1. (5 points) Let $X(t) = \sqrt{Y(t)}$. Compute $dX(t)$ by using Ito's formula and deduce the Stochastic Differential Equation satisfied by $X(t)$.
2. (10 points) Show by using Ito's formula that the following expression is a solution of the SDE derived in the previous question:

$$X(t) = e^{\mu t}X(0) + \sigma \int_0^t e^{\mu(t-s)}dW(s).$$

3. (5 points) Deduce the solution $Y(t)$ of the first SDE introduced.

Chapter 10

Module 10: Pricing a derivative security in continuous-time

10.1 The geometric Brownian motion asset price model

First of all, we recall briefly, the geometric Brownian motion model for an asset price

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), S(0) = S_0$$

where $W(t)$ is a standard Brownian motion under a probability measure $\mathbb{P}$ and μ, σ are constant with $\sigma > 0$. $\mathbb{P}$ is the *true* probability measure or *historical* probability measure. Because of the presence of the drift coefficient μ, the stock price tends to go up or down and is not a martingale under this probability measure.

We also assume that there exists a so-called *risk free* rate, which could be taken to be the interest rate paid by a money market account for instance; we denote it by r and assume that it is positive.

Next, it is actually possible to change the probability measure, from $\mathbb{P}$ to an equivalent probability measure $\tilde{\mathbb{P}}$, under which the discounted stock price, $e^{-rt}S(t)$ is a martingale, i.e.

$$d(e^{-rt}S(t)) = \sigma(e^{-rt}S(t))d\tilde{W}(t),$$

where $\tilde{W}(t)$ is a Brownian motion under the risk-neutral probability measure $\tilde{\mathbb{P}}$.

As you can see, there is no dt term in the above SDE and this makes the discounted asset price a martingale. Another way to look at it is by rewriting it in integral form

$$e^{-rt}S(t) - S(0) = \sigma \int_0^t e^{-ru}S(u)d\tilde{W}(u),$$

and observing that the Ito integral in the right-hand side is a martingale.

This is basically a mathematical trick that will allow us to derive an analytical formula for the price of a derivative security from the elegant risk-neutral pricing formula. The existence of such a risk-neutral probability measure is due to the absence of arbitrage in the Black-Scholes-Merton model. The idea is exactly the same as for the Binomial tree model seen earlier. In this course, I do not present the mathematical tools that allow us to rewrite the discounted asset price as a martingale in continuous time and I only show below how we move the drift coefficient into the diffusion term

Roughly, we "hide" the drift coefficient μ into the Brownian term by introducing the market price of risk $\frac{\mu-r}{\sigma}$ and rewriting the SDE as

$$dS(t) \quad = \quad rS(t)dt + \sigma S(t)[\frac{\mu - r}{\sigma}dt + dW(t)].$$

It is possible to define a risk-neutral probability measure $\tilde{\mathbb{P}}$ under which the process $\tilde{W}(t) = \frac{\mu-r}{\sigma}t + W(t)$ is a standard Brownian motion. Consequently, we can write the new SDE satisfied by the stock price as

$$dS(t) = rS(t)dt + \sigma S(t)d\tilde{W}(t), S(0) = S_0.$$

Consequently, the stock's instantaneous rate of return per unit time is equal to the risk free rate under the new probability measure $\tilde{\mathbb{P}}$.

10.2 The Black-Scholes analytical formula

In this section, the goal is to compute the price of a European call by using the Black-Scholes-Merton no arbitrage theory. We only present an outline of the theory and we refer to the textbook by Paul Wilmott, Sam Howison, Jeff Dewynne or more advanced texts such as the book by Shreve, *Stochastic Calculus for Finance*,II, for further details.

The risk-neutral pricing formula tells us in the continuous-time setting what the fair price, $V(t)$, of a European derivative security is, and allows us to derive an analytical formula for this price.

First of all, we recall that the discounted stock price is a martingale under the risk-neutral probability measure.

Next, consider an agent who is short a call and needs to hedge this short position. We denote by $X(t)$ the value of his hedging portfolio. He starts with an initial capital $X(0)$ and by time T, the portfolio value $X(T)$ must satisfy

$$X(T) = V(T).$$

She invests in a money market account with pays a constant interest rate $r > 0$ and the underlying asset. He holds at time t, $\Delta(t)$ shares of the stock. We need to determine $X(0)$ and $\Delta(t)$ such that

$$X(T) = V(T).$$

For now, we leave out the computation of $\Delta(t)$ and focus on the price $V(t)$. We will address this question later and for now, we simply assume that $\Delta(t)$ exists.

Next, It turns out, that the discounted portfolio is also a martingale under the risk-neutral probability measure. This can be shown by using the techniques of stochastic calculus. This implies that

$$e^{-rt}X(t) = \\ \tilde{\mathbb{E}}[e^{-rT}X(T)|\{\tilde{W}(s), 0 \leq s \leq t\}] = \tilde{\mathbb{E}}[e^{-rT}V(T)|\{\tilde{W}(s), 0 \leq s \leq t\}].$$

In the above formula, the expectation denoted by $\tilde{\mathbb{E}}$, is an expectation under the risk-neutral probability measure. This expectation is conditional on the flow of historical observations of the Brownian motion from time 0 up to time t (the uncertainty is resolved in the time interval $[0, t]$).

$X(t)$ is the capital that is needed in order to hedge the short position in the call with maturity date T. We can set $V(t) = X(t)$, which means that we are replicating the call.

Hence, this gives us

$$e^{-rt}V(t) = \tilde{\mathbb{E}}[e^{-rT}V(T)|\{\tilde{W}(s), 0 \leq s \leq t\}].$$

Wee can then divide the left and right hand sides by the discount term e^{-rt} and this yields the so-called *risk-neutral* pricing formula

$$V(t) = \tilde{\mathbb{E}}[e^{-r(T-t)}V(T)|\{\tilde{W}(s), 0 \leq s \leq t\}], \text{ for } t \in [0, T].$$

Since the Brownian motion has independent increments and is the only source of uncertainty in the geometric Brownian motion model with constant coefficients, the asset price $S(t)$ is Markovian. Consequently, we can rewrite the risk-neutral pricing formula as

$$\begin{aligned} V(t) &= \tilde{\mathbb{E}}[e^{-r(T-t)}V(T)|\{S(s), 0 \leq s \leq t\}] \\ &= \tilde{\mathbb{E}}[e^{-r(T-t)}V(T)|S(t)] \end{aligned}$$

Basically, exactly as in the binomial tree model, $V(t)$ is equal to the discounted expected payoff under the risk-neutral probability measure at time t.

Furthermore, in the case of a European call option, the payoff is given by

$$V(T) = (S(T) - K)^{+}.$$

It is actually fairly easy to derive from the risk-neutral pricing formula a closed-formula for the price of a European call. In the same manner, you can derive the price of a put and of other various European derivative securities and it is even possible to compute in a similar but slightly more complex way, the value of some specific exotic options such as Barrier or lookback options. Roughly, this formula works as long as the payoff can be written as a function of the maturity date T.

We used the so-called *martingale approach* to determine the price of a derivative security with payoff $V(T)$. This approach does not help determining the hedging strategy $\Delta(t)$. Within this framework, one can show that there exists a hedging strategy but is unable to derive it in closed-form.

In these notes, we use the martingale approach because of its clarity, effectiveness and simplicity. It is possible to derive this result by simply applying the *law of one price*, which is a consequence of the absence of arbitrage, the completeness of the market model and other hypotheses. After all, we are fully eliminating the risk here and a stochastic model is not absolutely necessary to treat this standard application to option pricing. For an overview of the law of one price approach, we refer to the book by D. Stefanica, *A primer for the mathematics of financial engineering*.

10.2.1 Derivation of the solution of Black-Scholes

Since the price of the call, $V(t)$, at time t, only depends on the time t and on the price of the stock at time t, but not on the stock price prior to t; by the Markov property, it can be shown that there exists a function $v(t, s)$ such that

$$V(t) = v(t, S(t)).$$

We rewrite the risk-neutral pricing formula as

$$v(t, s) = \tilde{\mathbb{E}}[e^{-r(T-t)}V(T)|S(t) = s],$$

where the whole history is summarized in the simple condition $S(t) = s$, stating that the current stock price at time t is given by s, and v is a deterministic function of the time t and the current stock price s, at time t.

In this section, we use the risk-neutral pricing formula to derive an analytical formula for the price of a European call with payoff $V(T) = (S(T) - K)^+$. The same approach can be used to compute the price of some other options. For instance, in the assignment, you will see the case of a Binary option. However, even when the risk-neutral pricing formula holds, it is not always possible to derive the price in closed form. For instance, the risk-neutral pricing formula is applicable to the Asian option but there is no analytical formula available for the price of an Asian option.

We recall that, under the risk-neutral measure, the stock price is

$$S(t) = S(0) \exp\{\sigma \tilde{W}(t) + (r - \frac{1}{2}\sigma^2)t\}.$$

Since we also have that

$$S(T) = S(0) \exp\{\sigma \tilde{W}(T) + (r - \frac{1}{2}\sigma^2)T\},$$

we find, by taking the ratio of these two formulas, that

$$S(T) = S(t) \exp\{\sigma(\tilde{W}(T) - \tilde{W}(t)) + (r - \frac{1}{2}\sigma^2)(T - t).\}$$

We then define the standard random variable

$$Y = -\frac{\tilde{W}(T) - \tilde{W}(t)}{\sqrt{T - t}}$$

and rewrite

$$S(T) = S(t) \exp\{-\sigma\sqrt{T - t}Y + (r - \frac{1}{2}\sigma^2)(T - t)\}.$$

To simplify, we will use from now on, the notation $s = S(t)$ for the current stock price. Thus, we have

$$
\begin{aligned}
&v(t,s) = \\
&\tilde{\mathbb{E}}\Big[\exp(-r(T-t)) \times \\
&\big(s\exp\{-\sigma\sqrt{T-t}Y + (r-\frac{1}{2}\sigma^2)(T-t)\} - K\big)^+\Big] \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp(-r(T-t)) \times \\
&\big(s\exp\{-\sigma\sqrt{T-t}y + (r-\frac{1}{2}\sigma^2)(T-t)\} - K\big)^+\exp(-\frac{1}{2}y^2)dy.
\end{aligned}
$$

Now, $\big(s\exp\{-\sigma\sqrt{T-t}y + (r-\frac{1}{2}\sigma^2)(T-t)\} - K\big)^+$ is positive if and only if

$$
s\exp\{-\sigma\sqrt{T-t}y + (r-\frac{1}{2}\sigma^2)(T-t)\} > K,
$$

which is equivalent to

$$
y < \frac{1}{\sigma\sqrt{T-t}}\big[\log\frac{s}{K} + (r-\frac{\sigma^2}{2})(T-t)\big].
$$

Consequently, we have

$$
\begin{aligned}
v(t,s) \quad = \quad & \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(T-t,s)}\exp(-r(T-t)) \times \\
& \big(s\exp\{-\sigma\sqrt{T-t}y + (r-\frac{1}{2}\sigma^2)(T-t)\} - K\big)\exp(-\frac{1}{2}y^2)dy
\end{aligned}
$$

where $d_-(T-t,s)$ is defined by

$$
d_-(\tau,s) = \frac{1}{\sigma\sqrt{\tau}}\big[\log\frac{s}{K} + (r-\frac{\sigma^2}{2})\tau\big].
$$

Next,

$$
\begin{aligned}
v(t,s) \quad = \quad & \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(T-t,s)} s\exp\{-\sigma\sqrt{T-t}y+-\frac{1}{2}\sigma^2(T-t)-\frac{1}{2}y^2\}dy \\
& -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(T-t,s)} \exp(-r(T-t))K\exp\{-\frac{1}{2}y^2\}dy \\
& = \frac{s}{\sqrt{2\pi}}\int_{-\infty}^{d_-(T-t,s)} \exp\{-\frac{1}{2}(y+\sigma\sqrt{(T-t)})^2\}dy \\
& -\exp(-r(T-t))KN(d_-(T-t,s)) \\
& = \frac{s}{\sqrt{2\pi}}\int_{-\infty}^{d_-(T-t,s)+\sigma\sqrt{T-t}} \exp\{-\frac{1}{2}z^2\}dz \\
& -\exp(-r(T-t))KN(d_-(T-t,s)) \\
& = sN(d_+(T-t,s))-\exp(-r(T-t))KN(d_-(T-t,s))
\end{aligned}
$$

where $d_+(T-t,s)$ is defined as

$$d_+(\tau,s)=\frac{1}{\sigma\sqrt{\tau}}\big[\log\frac{s}{K}+(r+\frac{\sigma^2}{2})\tau\big]$$

and N denotes the standard normal cumulative distribution function.

How about the price of a put?

The payoff of a European put is

$$V(T)=(K-S(T))^+.$$

The put-call parity formula tells us that the difference between a the price of a European call and the price of a European put is equal to the price of a forward contract

$$F(t,s)=s-e^{-r(T-t)}K=v(t,s)-p(t,s),$$

where $p(t,s)$ denotes the price of a put at time t and for the current stock price s.

Clearly, the above equation is true at $T=t$.

$$F(T,s)=v(T,s)-p(T,s).$$

Then, by the *law of one price*, this equality must still be true for every $0 \leq t < T$. Indeed, if this equation were wrong, one could either sell or buy the portfolio that is long a call, short a put and short a forward to realize a profit instantly, which goes against the assumption that there is no arbitrage in the Black-Scholes model. So the put-call parity must be true.

Consequently, the price of a European put is given by the Black-Scholes formula

$$p(t,s) = Ke^{-r(T-t)}N(-d_-(T-t,s)) - sN(-d_+(T-t,s)).$$

10.2.2 How about the hedging strategy? The PDE approach

One can derive the hedging strategy in closed form along with the Black-Scholes Partial Differential Equation by using the Partial Differential Equation approach.

First of all, we apply Ito's Lemma to the price $v(t, S(t))$ at time t of the European call:

$$\begin{aligned} dv(t,S(t) &= v_t(t,S(t))dt + v_s(t,S(t))dS(t) + \frac{1}{2}\sigma^2 v_{ss}(t,S(t))dS(t)dS(t) \\ &= [v_t(t,S(t)) + \mu S(t)v_s(t,S(t)) + \frac{1}{2}\sigma^2 S(t)^2 v_{ss}(t,S(t))]dt + \\ &\quad \sigma S(t)v_s(t)dW(t). \end{aligned}$$

Next, we consider an investor who is short a European call and needs to hedge her short position. We denote by $X(t)$ the value at time t of the hedging portfolio. We denote by $\Delta(t)$ the number of shares of underlying asset held at time t in the hedging portfolio. The evolution of the investor's self-financing portfolio can be written as

$$dX(t) = \Delta(t)dS(t) + r(X(t) - \Delta(t)S(t))dt.$$

Then, we compute the differential of the investor's position, which is long the hedging portfolio and short the call

$$
\begin{aligned}
d(X(t)-v(t,S(t)) &= \Delta(t)\mu S(t)dt+\Delta(t)\sigma S(t)dW(t)+r(X(t)-\Delta(t)S(t))dt- \\
&\quad [v_t(t,S(t))+\mu S(t)v_s(t,S(t))+\frac{1}{2}\sigma^2 S(t)^2 v_{ss}(t,S(t))]dt- \\
&\quad \sigma S(t)v_s(t)dW(t).
\end{aligned}
$$

The value of the hedging portfolio at time T, should be equal to the payoff of the call, i.e.

$$X(T)=v(T,S(T))=(S(T)-K)^+.$$

The initial capital $X(0)$ is provided by the price $v(0,S(0))$ of the call at time 0, i.e. $X(0)=v(0,S(0))$.

Furthermore, we seek a hedging portfolio replicating the option, i.e. satisfying

$$X(t)=v(t,S(t)),\ \text{for all } 0\le t\le T.$$

The condition

$$X(T)=v(T,S(T))=(S(T)-K)^+.$$

is equivalent to

$$X(T)=v(T,S(T)),d((X(t)-v(t,S(t))=0.$$

Taking these condition into account, we obtain

$$\Delta(t)\mu S(t)dt+\Delta(t)\sigma S(t)dW(t)+r(v(t,S(t)-\Delta(t)S(t))dt-[v_t(t,S(t))+\mu v_s(t,S(t))+\frac{1}{2}\sigma^2 S(t)^2 v_{ss}(t,S(t))]dt-\sigma S(t)v_s(t)dW(t)=0.$$

Since our goal is to replicate the option payoff with certainty, we cancel the fluctuating Brownian terms by choosing the delta-hedging strategy

$$\Delta(t) = v_s(t, S(t).$$

Finally, the only terms left in the equation are the dt terms. Substituting the above hedging strategy into the equation, we obtain after simplifying

$$rv(t, S(t)) - v_t(t, S(t)) - rv_s(t, S(t)) - \frac{1}{2}\sigma^2 S(t)^2 v_{ss}(t, S(t)) = 0.$$

Note that, if instead of replicating the option, we had at any time t

$$X(t) > v(t, S(t)),$$

we could borrow the difference $X(t) - v(t, S(t))$ from the money market account paying the interest rate $r > 0$ to invest in the portfolio that is long the hedging portfolio $X(t)$ and short the European call $v(t, S(t))$, thus realizing a riskless profit. This would constitute an arbitrage opportunity!

Since the process $S(t)$ may take any nonnegative real value s, the option price $v(t, s)$ satisfies the Black-Scholes-Merton Partial Differential Equation (PDE in short) in the domain $[0, T) \times [0, +\infty)$.i.e.

$$v_t(t, s) + rsv_s(t, s) + \frac{1}{2}\sigma^2 s^2 v_{ss}(t, s) = rv(t, s), \text{ for all } t \in [0, T), s \geq 0.$$

Note that there is not any randomness in this PDE and that s is no longer a process but rather the spatial variable of the function v.

The above PDE must be coupled with the terminal condition

$$v(T, s) = (s - K)^+.$$

One can use the analytical formula previously derived and differentiate it in order to obtain the *delta* of the call analytically, one of the *Greeks*, representing the number of shares of the stock in the hedging Portfolio. It is given by

$$\Delta(t, s) = v_s(t, s) = N(d_+(T - t, s)).$$

10.3 Monte Carlo simulations

You can easily compute the price of a European call option by using Monte Carlo simulations. I describe this method in this section and you are asked to implement it on a computer for your weekly assignment.

We recall the closed-form formula for the stock price

$$S(T) = S(0)exp\{(r - \frac{1}{2}\sigma^2)T + \sigma\tilde{W}(T)\}.$$

We assume that this initial stock price is given and equal to s. We can rewrite the stock price at time T as

$$S(T) = sexp\{(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Y\}.$$

where Y is a standard normal random variable. We can simulate the stock price at the maturity date by using the above formula with Y drawn from the standard normal distribution. We will need to simulate many outcomes. So, for $i = 1, \cdots, n$, we can draw Y_i from the standard normal distribution, such that, $Y_1, Y_2, \cdots, Y_n$ are independent in order to obtain

$$S_i(T) = sexp\{(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Y_i\}, \text{ for } i = 1, \cdots, n.$$

We now present the Monte Carlo simulation algorithm below.

Algorithm

1. Start with $S(0) = s$.
2. Then, for every $i = 1 \cdots n$,

   2.1 Generate $S_i(T)$ by using the exact formula

$$S_i(T) = s\ \exp\{(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Y_i\}.$$

where Y_i is drawn from the standard normal random distribution and $Y_1, Y_2, \cdots Y_n$ are independent.

2.2 Set $C_i = e^{-rT}(S_i(T) - K)^+$.

3. Set $\hat{C}_n = \frac{1}{n}\sum_{i=1}^n C_i$.

In order to obtain some meaningful results, one need to choose a very large value for n, at least $n = 100,000$ or $n = 1,000,000$ because the rate of convergence of these Monte-Carlo simulations is only $\frac{1}{\sqrt{n}}$.

The convergence of the Monte Carlo simulations method is insured by the strong law of large numbers. Indeed, the estimator $\hat{C}_n = \frac{1}{n}\sum_{i=1}^n C_i$ converges to the price of the European call as n goes to ∞ by the strong law of large numbers. In addition, this estimator is unbiased in the sense that

$$\mathbb{E}[\hat{C}_n] = \mathbb{E}[e^{-rT}(S(T) - K)^+ | S(0) = s].$$

Consequently, the only source of error is the limited number of samples used for computing the expectation.

Furthermore, the central limit Theorem tells us what the distribution of the error is. Indeed, it is normal with a zero mean and a standard deviation $s_C/\sqrt{n}$ where s_C can be estimated by using the unbiased estimator

$$s_C = \sqrt{\frac{1}{n-1}\Sigma_{i=1}^n (C_i - \hat{C}_n)^2}.$$

Chapter 11

Module 11: A finite difference method for the Black-Scholes PDE

In practice, pure Finite Difference schemes are only useful in 1,2 or at most 3 spatial dimensions. They may be the only option for an optimal stochastic control problem. One of their merits is to be quite simple and easy to implement. They can also be potentially combined with Monte Carlo methods to solve a problem in higher dimensions.

For a basic introduction to Finite Difference methods for linear parabolic PDEs, I recommend the book by J.W. Thomas, *Numerical Partial Differential Equations*, edited by Springer.

11.1 Discretization of time and space

We pick a time step Δt and divide the interval $[0, T]$ into N subintervals, each being of length Δt, such that we have the relationship $N\Delta t = T$. In other words, we approximate the time interval $[0, T]$ by the collection of points $0, \Delta t, 2\Delta t, ..., n\Delta t = T$. As $\Delta t \to 0$ and $n \to +\infty$, the discrete grid converges to the whole time interval $[0, T]$.

Next, we can discretize the whole real line by dividing it into a infinite number of intervals. Our mesh is made of a collection of points $\{x_i = i\Delta x | i \in \mathbb{Z}\}$ where Δx is the mesh size. As $\Delta x \to 0$, the mesh converges to the whole real line. For Δx *small enough*, the mesh is a *good approximation* of the real line.

In what follows, we consider the mesh

$$\{(n\Delta t, i\Delta x) | n = 0..N, i \in \mathbb{Z}\},$$

where $(\Delta t, \Delta x)$ are the discretization steps and N is such that $N\Delta t = T$.

Our goal is to compute numerically an approximation of the option value on the mesh. We denote by u_i^n the approximation of the option value u at the point $(n\Delta t, i\Delta x)$, i.e.

$$u(n\Delta t, i\Delta x) \approx u_i^n.$$

11.2 The classical explicit and implicit schemes

First, let me recall the classic explicit and implicit schemes for the heat equation and discuss their properties

$$u_t - u_{xx} = 0 \text{ in } (0,T] \times I\!R. \tag{11.1}$$

$$u(0,x) = u_0(x) \tag{11.2}$$

We assume here that u_0 is continuous and bounded in $\mathbb{R}$.

We use the mesh $\{(n\Delta t, i\Delta x) | n = 0..N, i \in \mathbb{Z}\}$ where $N\Delta t = T$. We denote the approximation u of the above equation on the mesh by u_i^n.

11.2.1 The standard explicit scheme:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} - \frac{u_{i+1}^n + u_{i-1}^n - 2u_i^n}{\Delta x^2} = 0.$$

Since this scheme is explicit, it is very easy to compute at each time step $n+1$ the value of the approximation $\{u_i^{n+1} | i \in \mathbb{Z}\}$ from the value of the approximation at the time step n, namely $\{u_i^n | i \in \mathbb{Z}\}$.

$$u_i^{n+1} = u_i^n + \Delta t\{\frac{u_{i+1}^n + u_{i-1}^n - 2u_i^n}{\Delta x^2}\}.$$

The algorithm is initialized with the initial condition

$$u_i^0 = u_0(i\Delta x).$$

Next, let us discuss the properties of this scheme: clearly, it is **consistent** with the continuous equation since formally, the truncation error is of order two in space and order one in time. Let us recall how one can calculate the truncation error for a smooth function u with bounded partial derivatives. Simply write the Taylor expansions

$$u_{i+1}^n = u_i^n + u_x(n\Delta t, x_i)\Delta x + \frac{1}{2}u_{xx}(n\Delta t, x_i)\Delta x^2 + u_{xxx}\frac{1}{6}\Delta x^3 +$$

$$\frac{1}{24}u_{xxxx}\Delta x^4 + \Delta x^4 \epsilon(\Delta x).$$

$$u_{i-1}^n = u_i^n - u_x(n\Delta t, x_i)\Delta x + \frac{1}{2}u_{xx}(n\Delta t, x_i)\Delta x^2 - \frac{1}{6}u_{xxx}\Delta x^3 +$$

$$\frac{1}{24}u_{xxxx}\Delta x^4 + \Delta x^4 \epsilon(\Delta x)$$

where $\lim_{\Delta x \to 0} \epsilon(\Delta x) = 0$. Then, adding up the two expansions, subtracting $2u_i^n$ from the left- and right hand sides and dividing by Δx^2, one obtains

$$\frac{u_{i+1}^n + u_{i-1}^n - 2u_i^n}{\Delta x^2} = u_{xx} + \frac{1}{12}u_{xxxx}\Delta x^2 + \Delta x^2 \epsilon(\Delta x)$$

and thus the truncation error for this approximation of the second spatial derivative is of order 2. Similarly the expansion

$$u_i^{n+1} = u_i^n + u_t(n\Delta t, x_i)\Delta t + \frac{1}{2}u_{tt}(n\Delta t, x_i)\Delta t^2 + \Delta t^2 \epsilon(\Delta t)$$

where $\lim_{\Delta t \to 0} \epsilon(\Delta t) = 0$ yields

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = u_t(n\Delta t, x_i) + \frac{1}{2}u_{tt}\Delta t + \Delta t \epsilon(\Delta t).$$

The truncation error for the approximation of the first derivative in time is of order 1 only.

Furthermore, the approximation is stable and, hence the scheme is convergent if the following condition is satisfied:

$$(-1 + 2\frac{\Delta t}{\Delta X^2}) \leq 0$$

or equivalently

$$\Delta t \leq \frac{1}{2}\Delta X^2.$$

It is often called the CFL condition, where CFL is an abbreviation for Courant-Friedrichs-Lewy.

11.2.2 The standard implicit scheme

For many financial applications, the explicit scheme turns out to be very inaccurate because the CFL condition forces the time step to be so small that the rounding error dominates the total computational error (note that the computational error is equal to the sum of the rounding error and truncation error). Most of the time, an implicit scheme is preferred because it is unconditionally convergent, regardless of the size of the time step. We now evaluate the second derivative at time $(n+1)\Delta t$ instead of time $n\Delta t$,

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} - \frac{u_{i+1}^{n+1} + u_{i-1}^{n+1} - 2u_i^{n+1}}{\Delta x^2} = 0.$$

Implementing an algorithm allowing to compute the approximation is less obvious here. This discrete equation may be converted into a linear system of equations and the algorithm will then consist in solving this system. The truncation errors for smooth functions are the same as for the explicit scheme and the consistency follows from this analysis. Finally, it turns out that, for any choice of the time step, the implicit scheme is stable, and therefore convergent, which is an advantage.

11.3 The Black-Scholes-Merton PDE

The price of a European call $u(t, x)$ satisfies the degenerate linear PDE

$$
\begin{aligned}
&-u_t + ru - \frac{1}{2}\sigma^2 x^2 u_{xx} - rxu_x = 0 \text{ in } (0,T] \times [0,+\infty) \\
&u(T,x) = (x-K)^+.
\end{aligned}
$$

Note that, in this example, the solution is no longer bounded but is nonnegative and grows at most linearly for large values of x. The framework can be slightly modified to accommodate the linear growth of the value function at infinity. In the next subsection, we study the behavior of the solution at the edge $x = 0$ because we will need this information later for solving numerically the Black-Scholes PDE.

11.3.1 Boundary condition: what happens at x=0?

$x = 0$ is a *boundary*, i.e. it is at the edge of the domain. In BS, the only boundary in space is at $x = 0$. The PDE actually holds at $x = 0$ and can be solved explicitly, leading to the solution $u(t, 0) = 0$. To see this, plug $x = 0$ into the PDE. You obtain

$$
u_t(t,0) = ru(t,0),
$$

which leads to the explicit solution $u(t, 0) = e^{rt}u(T, 0)$. Since, in addition, $u(T, 0) = (0 - K)^+ = 0$, then $u(t, 0) = 0$.

At this point, for a sake of convenience, we perform a change of time variable because our scheme will have to be initialized with the terminal condition. We use the variable $\tau = T - t$, which represents the time to expiration, instead of the real time t and we define

$$
v(\tau,x) = u(t,x).
$$

Then, since $v_\tau = -u_t$, the function v satisfies the PDE

$$
\begin{aligned}
&v_\tau + rv - \frac{1}{2}\sigma^2 x^2 v_{xx} - rxv_x = 0 \text{ in } [0,T) \times [0,+\infty) \\
&v(0,x) = (x-K)^+.
\end{aligned}
$$

We actually choose to rewrite the above PDE with our standard notation $u(t, x)$ instead of the notation $v(\tau, x)$ (note that the following PDE is exactly the same as the one above).

$$
\begin{aligned}
&u_t + ru - \frac{1}{2}\sigma^2 x^2 u_{xx} - rxu_x = 0 \text{ in } [0,T) \times [0,+\infty) \\
&u(0,x) = (x-K)^+.
\end{aligned}
$$

The Black-Scholes-Merton PDE is linear and its Elliptic operator is degenerate. The degeneracy makes it somewhat difficult to handle mathematically. However, we can get rid of the degeneracy by using a change of variable.

11.3.2 Getting rid of the degeneracy in BS

It is actually possible to get rid of the degeneracy in BS by using the change of variable $y = \log x$. We define the price v in the new variable y by setting

$$v(t,y) = u(t,x).$$

Since

$$v_t = u_t, v_y = xu_x, x^2 u_{xx} = v_{yy} - v_y,$$

the new BS PDE satisfied by $v(t,y)$ is

$$
\begin{aligned}
&v_t + rv - \frac{1}{2}\sigma^2 v_{yy} - (r - \frac{1}{2}\sigma^2)v_y = 0 \text{ in } (0,T] \times \mathbb{R} \\
&v(0,y) = (e^y - K)^+.
\end{aligned}
$$

11.3.3 Scheme for the Black-Scholes PDE

First of all, we introduce the mesh

$$\{(n\Delta t, i\Delta y) | n = 0 \cdots N, i \in \mathbb{Z}\},$$

and the notation

$$y_i = i\Delta y.$$

We can approximate the second spatial derivative and the first derivative with respect to the time with the same finite differences as for the heat equation. Next, we need to approximate v_y. This can be done for instance by using the centered Finite Difference

$$-(r - \frac{1}{2}\sigma^2)v_y \approx -(r - \frac{1}{2}\sigma^2)\frac{u_{i+1}^{n+1} - u_{i-1}^{n+1}}{2\Delta y}.$$

Moreover we can easily approximate the zeroth order term

$$rv(t,y) \approx ru_i^{n+1},$$

and this yields the implicit scheme

$$\begin{aligned}
&\frac{u_i^{n+1} - u_i^n}{\Delta t} + ru_i^{n+1} - \frac{1}{2}\sigma^2\frac{u_{i+1}^{n+1} + u_{i-1}^{n+1} - 2u_i^{n+1}}{\Delta y^2} - \\
&(r - \frac{1}{2}\sigma^2)\frac{u_{i+1}^{n+1} - u_{i-1}^{n+1}}{2\Delta y} = 0, \\
&u_i^0 = (e^{y_i} - K)^+.
\end{aligned}$$

This scheme is unconditionally convergent for every pair $(\Delta t, \Delta y)$.

The next question is: how do we invert this scheme? We can rewrite the above scheme in matrix form and solve the system of linear of equations we obtain.

11.3.4 The implicit scheme in matrix form

To this end, we introduce the notation $h = (\Delta t, \Delta y)$ and we define the vector u_h^n and the matrix A_h

$$u_h^n = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ u_{i-1}^n \\ u_i^n \\ u_{i+1}^n \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}, A_h = \begin{bmatrix} \cdot & \cdot & \cdot & & & & & & \\ & \cdot & \cdot & \cdot & & & & & \\ & & a & b & c & & & & \\ & & & & \cdot & \cdot & \cdot & & \\ & & & & & \cdot & \cdot & \cdot & \\ & & & & & & \cdot & \cdot & \cdot \\ & & & & & & & & \end{bmatrix}.$$

where the coefficients a, b, c are defined by the particular finite-difference scheme we are implementing.

$$\begin{aligned} a &= -\frac{1}{2}\Delta t\frac{\sigma^2}{\Delta y^2} + \Delta t\frac{(r - \frac{1}{2}\sigma^2)}{2\Delta y} \\ b &= 1 + \Delta t(r + \frac{\sigma^2}{\Delta y^2}) \\ c &= \Delta t(-\frac{(r - \frac{1}{2}\sigma^2)}{2\Delta y} - \frac{1}{2}\frac{\sigma^2}{\Delta y^2}). \end{aligned}$$

In matrix form, the scheme then reads

$$A_h u_h^{n+1} = u_h^n.$$

and its solution is given by

$$u_h^{n+1} = A_h^{-1} u_h^n.$$

Since the matrix A_h is tridiagonal, solving the above system can be done efficiently. You should not use the Matlab operator $inv(A)$, but rather the code presented at the end of the notes.

11.3.5 Boundary conditions

In practice, one cannot work on the unbounded domain $(-\infty, +\infty)$. One must work instead on a large but bounded domain $[-R, R]$ where R is large enough.

One must impose reasonable boundary conditions at both $y = -R$ and $y = R$ since one cannot use the scheme on the boundary (the point at the edge R does not have a neighbor to the right and the point at the edge $-R$ does not have a neighbor to the left).

For the Black-Scholes model, the correct boundary condition at $y = R$ is

$$v(t, R) \approx e^R - e^{-rt}K,$$

where t is the time to expiration.

We also know approximately the solution at $y = -R$. Indeed, $y = -R$, for R large, does correspond to $x = 0$. So we have

$$v(t, -R) \approx 0.$$

We are now working on the finite mesh

$$\{(n\Delta t, i\Delta y) | n = 0...N, i = -M \cdots 0 \cdots M\},$$

where N, M are such that $N\Delta t = T$ and $M\Delta y = R$.

The next relevant question is: does our scheme do the right job at $y = -R$ and $y = R$? Or do we need to modify it in order to take into account these boundary conditions?

First of all, the general scheme in matrix form holds only in the interior of the interval, i.e. for $-(M-1) \leq i \leq M-1$. The first line of the matrix then corresponds to the case $i = -(M-1)$ and it has only 2 coefficients since a is missing. It does not matter since a would be multiplied by u^n_{-M}, which should be equal to 0 anyway. In other words, we can safely ignore the left neighbor u^n_{-M} in the scheme for $i = -(M-1)$. So our scheme works well the way it is for $i = -(M-1)$.

Of course, the scheme in matrix form must be coupled with the Dirichlet boundary condition

$$u^n_{-M} = 0,$$

for $i = M$.

However, at $y = R$, there is currently a term missing in the scheme. Indeed, we have to impose at $y = R$,

$$u_M^n = e^{M\Delta y} - e^{-r(n\Delta t)}K,$$

and we need to make our scheme consistent with this condition.

For $i = M - 1$, the coefficient c is missing in the matrix and this is the coefficient that would multiply the term u_M^n to which the Dirichlet condition applies. We must incorporate the Dirichlet condition in the scheme.

We can do so by modifying the right hand side of the equation in matrix form. For $i = M - 1$, the last component of the right-hand side should be

$$u_{M-1}^n + \Delta t\{\frac{1}{2}\frac{\sigma^2}{\Delta y^2} + \frac{(r - \frac{1}{2}\sigma^2)}{2\Delta y}\} \times (e^{M\Delta y} - e^{-r((n+1)\Delta t)}K).$$

Finally we are going to solve the equation

$$A_h u_h^{n+1} = B_h^n.$$

where A_h is the $(2M - 1) \times (2M - 1)$ matrix

$$A_h = \begin{bmatrix} b & c & & & & & & \\ . & . & . & & & & & \\ & . & . & . & & & & \\ & & a & b & c & & & \\ & & & . & . & . & & \\ & & & & . & . & . & \\ & & & & & . & . & . \\ & & & & & & a & b \end{bmatrix}$$

and

$$B_h^n = \begin{bmatrix} u^n_{-(M-1)} \\ . \\ . \\ . \\ u^n_{i-1} \\ u^n_i \\ u^n_{i+1} \\ . \\ . \\ . \\ u^n_{M-1} + \Delta t\{\frac{1}{2}\frac{\sigma^2}{\Delta y^2} + \frac{(r-\frac{1}{2}\sigma^2)}{2\Delta y}\} \times (e^{M\Delta y} - e^{-r((n+1)\Delta t)}K \end{bmatrix}.$$

11.4 Implementation

In the assignment, you are asked to implement the implicit scheme for the BS PDE in Matlab. You may use the efficient algorithm for solving the linear system with a tridiagonal matrix that I posted on Blackboard. Note that, in your code, you could define 3 vectors containing the coefficients a, b, c, rather than define a whole $(2M-1)\times(2M-1)$ matrix to represent the very sparse matrix A_h (this would consume too much memory). Alternately, you may use the backslash operator in Matlab, which, in recent versions, is able to recognize tridiagonal matrices and solve this type of system efficiently. However, this requires creating a $(2M-1)\times(2M-1)$ matrix, which might consume much more memory than defining 3 vectors. So, if you use the backslash operator, you should at lest create a *sparse matrix* in Matlab, rather than a regular matrix full of zeros.

warning: In any case, you should never use the *inv* operator because it is both too slow and possibly inaccurate.

Here is below how the algorithm for solving a tridiagonal system.

1. Set the parameters T, K, r, σ, the steps $\Delta t, \Delta y$ the integers N, M.
2. Define the vectors a,b,c that constitute the 3 diagonals of A_h or the matrix A_h.

3. Initialize $u_i^0 = (e^{i\Delta y} - K)^+$ for $i = -M, \cdots, M$.

4. For $n = 0 \cdots N-1$ (time loop)

   4.1 Define B_h^n

   4.2 Given $(u_i^n)_{(i=-M\cdots M)}$, solve the system $A_h u_h^{n+1} = B_h^n$ to compute u_i^{n+1} for $i = -(M-1)\cdots M-1$.

   4.3 Set $u_{-M}^{n+1} = 0$, $U_M^{n+1} = e^{M\Delta y} - e^{-r((n+1)\Delta t)}K$.

11.4.1 Implementation in MATLAB of the tridiagonal system solver

We show below the Matlab code for the tridiagonal solver mentioned earlier.

function x = TDMAsolver(a,b,c,d)

%a, b, c are the column vectors for the compressed tridiagonal matrix, d is the right vector

$n = length(d)$; % n is the number of rows

% Modify the first-row coefficients

$c(1) = c(1)/b(1)$; % Division by zero risk.

$d(1) = d(1)/b(1)$;

for $i = 2 : n-1$

$temp = b(i) - a(i) * c(i-1)$;

$c(i) = c(i)/temp$;

$d(i) = (d(i) - a(i) * d(i-1))/temp$;

end

$d(n) = (d(n) - a(n) * d(n-1))/(b(n) - a(n) * c(n-1))$;

% Now back substitute.

$x(n) = d(n);$

for $i = n - 1 : -1 : 1$

$x(i) = d(i) - c(i) * x(i + 1);$

end

end

Chapter 12

Module 12: one period investment models

12.1 Single period consumption and investment models

In this lecture, I draw heavily from the book by S. R. Pliska, *Introduction to mathematical Finance, discrete time models*, in particular chapters 1 and 2. Any mistake is the sole responsibility of the author of these notes.

We consider the following model of an experiment leading to a finite number of possible outcomes, i.e.

$$\Omega = \{\omega_1, \cdots, \omega_K\}.$$

In a one-period model, we have two times, i.e., $t = 0$, corresponding to the beginning of the period, and $t = 1$, corresponding to the end of the period. The agent invests her initial wealth $w_0 = w$ at time 0 and by time $t = 1$, she has a wealth equal to w_1. The aim is to compute the optimal investment strategy. We can use a utility function to measure the performance for instance or alternately a mean-variance objective. I start with a utility function in the first model and at the end, I present the mean-variance criterion.

A utility function is a function $u : \mathbb{R} \times \Omega \to \mathbb{R}$, such that $w \to u(w, \omega)$ is differentiable, concave, and strictly increasing for each $\omega \in \Omega$. More precisely, if w is the wealth at time $t = 1$ and ω is the outcome of the random experiment, $u(w, \omega)$ represents the *utility* derived from the amount w. Our measure of performance can then be the expected utility of terminal wealth

$$\mathbb{E}[u(w_1)] = \sum_{\omega \in \Omega} \mathbb{P}(\omega) u(w_1(\omega), \omega).$$

The investor invests in $N + 1$ assets and we denote the trading strategy by $(H_0, H_1, \cdots, H_N) \in \mathbb{H} = \mathbb{R}^{N+1}$.

Our goal is to solve the optimal portfolio problem

$$\max_{(H_0, \cdots, H_N) \in \mathbb{H}} \mathbb{E}[u(w_1)], \text{ subject to } w_0 = w.$$

Next, we assume that the first asset is a bank account and that its value at time 0 is $B_0 = 1$, and at time 1 is $B_1 \geq 1$. In general we also assume that $B = (B_0, B_1)$ is a random variable. However, it will often be taken deterministic for many applications. The other assets are risky and the value of one share of the nth asset is denoted, at time t, by $S_n(t)$. In addition, we will use the notation

$$\Delta S_n = S_n(1) - S_n(0).$$

Furthermore, we make the bank account the *numeraire*, i.e. we define the discounted price process $(S_1^*(t), \cdots, S_N^*(t))$ by setting

$$S_n^*(t) = S_n(t)/B(t).$$

The initial wealth can be expressed as

$$w_0 = H_0 + H_1 S_1^*(0) + \cdots + H_N S_N^*(0).$$

Note that H_n represents the number of shares of the nth asset that are held in the portfolio.

The terminal wealth is

$$w_1 = B_1(H_0 + H_1 S_1^*(1) + \cdots + H_N S_N^*(1)).$$

We keep in mind that the initial wealth is given by $w_0 = w$. Then H_0 can be replaced, in the above formula, by

$$H_0 = w - H_1 S_1^*(0) - \cdots - H_N S_N^*(0).$$

Consequently, we can rewrite the objective function as

$$\max \mathbb{E}[u(B_1\{w + H_1\Delta S_1^* + \cdots H_N\Delta S_N^*\})].$$

Next, I claim that, if there exists an arbitrage opportunity, there cannot exist a solution to the above optimization problem.

In order to be able to see this, we denote by $\hat{H}$ the solution of this optimization problem and by H^a an arbitrage opportunity. We then apply the strategy

$$H = \hat{H} + H^a.$$

The corresponding discounted wealth at time 1 is

$$w + \sum_{n=1}^{N} H\Delta S_n^* = w + \sum_{n=1}^{N} \hat{H}\Delta S_n^* + \sum_{n=1}^{N} H^a\Delta S_n^* \geq w + \sum_{n=1}^{N} \hat{H}\Delta S_n^*$$

because H^a is an arbitrage opportunity. In addition, there is at least one outcome ω, for which this inequality is strictly positive. Then, we just exhibited a strategy that yields a better outcome than the optimal strategy, which is contradictory.

If there is a solution to the optimization problem, then there is no arbitrage opportunity in this model. Consequently, if there is a solution to the optimization problem, then, there is a risk-neutral probability measure. Next, we derive a relationship between the optimal solution and the risk-neutral probability measure.

12.1.1 Risk-neutral probability measure

We rewrite the objective function as

$$\sum_{\omega\in\Omega} \mathbb{P}(\omega)u(B_1(\omega)\{w + \sum_{n=1}^{N} H_n\Delta S_n^*(\omega)\}, \omega).$$

Then, at the optimum, the gradient of the objective function with respect to the variable $(H_1, H_2, \cdots, H_N)$ is equal to 0 (*first order necessary optimality condition*).

$$\sum_{\omega\in\Omega} \mathbb{P}(\omega)u'(B_1(\omega)\{w + \sum_{n=1}^{N} H_n\Delta S_n^*(\omega)\}, \omega)B_1(\omega)\Delta S_n^*(\omega) = 0,$$

where $n = 1, \cdots, N$. It can be interpreted as

$$\mathbb{E}[B_1 u'(w_1)\Delta S_n^*] = 0, \; n = 1, \cdots, N,$$

where u' is the first partial derivative of u, with respect to its first variable.

Besides, if $\tilde{\mathbb{P}}$ is a risk neutral probability measure, the discounted risky assets should satisfy the martingale property

$$\mathbb{E}^{\tilde{\mathbb{P}}}[\Delta S_n^*] = \sum_{\omega \in \Omega} \tilde{\mathbb{P}}(\omega)\Delta S_n^*(\omega) = 0, \; n = 1, \cdots, N.$$

By comparing this property with the previous equation, we see that we can obtain a risk neutral probability measure by setting

$$\tilde{\mathbb{P}}(\omega) = \mathbb{P}(\omega) B_1(\omega) u'(w_1(\omega), \omega),$$

at least up to a normalization constant. Indeed, we need to normalize the above result, to obtain a probability measure. We finally set

$$\tilde{\mathbb{P}}(\omega) = \frac{\mathbb{P}(\omega) B_1(\omega) u'(w_1(\omega), \omega)}{\mathbb{E}[B_1 u'(w_1)]}.$$

This way, we have

$$\sum_{k=1}^{K} \tilde{\mathbb{P}}(\omega_k) = 1.$$

We can also deduce from the first order optimality conditions and the definition of the risk neutral probability measure, that the discounted wealth process is a martingale, under the risk neutral probability measure.

To see this, we write

$$\begin{aligned}\mathbb{E}^{\tilde{\mathbb{P}}}[w_1/B_1] &= \sum_{\omega \in \Omega} \tilde{\mathbb{P}}(\omega) w_1(\omega)/B_1(\omega) \\ &= \sum_{\omega \in \Omega} \tilde{\mathbb{P}}(\omega) B_1(\omega)(w + \sum_n H_n \Delta S_n^*(\omega))/B_1(\omega) \\ &= w + \sum_n H_n \mathbb{E}^{\tilde{\mathbb{P}}}[\Delta S_n^*] = w = w_0.\end{aligned}$$

The next natural question is: if there is a risk neutral probability measure, does there always exist a solution to the optimization problem? The answer is no, not necessarily.

However, if there is a risk neutral probability measure, then there is a utility function u and an initial wealth w, such that the corresponding optimization problem has a solution. Besides, the market model is said to be complete if every contingent claim can be generated by some trading strategy. If this is not the case, the market model is said to be incomplete. One can show that the market model is complete if and only if the risk-neutral probability measure is unique.

12.1.2 How do you solve the optimization problem?

This problem is a standard convex optimization problem and can be solved by using standard techniques. However, the system of N nonlinear equations with N variables, corresponding to the first order optimality condition can be hard to solve in closed form in practice.

Instead, we can use an alternate technique based on the risk neutral probability measure. This is a two-step process: first we compute the optimal terminal value v_1 and secondly, we compute the trading strategy that generates this particular value of the terminal wealth.

As a first step, we solve the subproblem

$$\max_{v \in \mathbb{V}_w} \mathbb{E}[u(v)],$$

where $\mathbb{V}_w$ is given by

$$\mathbb{V}_w = \{v \in \mathbb{R}^K : \mathbb{E}^{\tilde{\mathbb{P}}}[v/B_1] = w\}.$$

This means that the discounted wealth process must be a martingale, under the risk-neutral portability measure. We can solve this subproblem by using a Lagrange multiplier, that is, we solve

$$\max_{v} \mathbb{E}u(v) - \lambda \mathbb{E}^{\tilde{\mathbb{P}}}[v/B_1],$$

where λ is such that the solution of the above optimization problem satisfies

$$\mathbb{E}^{\tilde{\mathbb{P}}}[v/B_1] = w.$$

n	$S_n^*(0)$	$S_n^*(1,\omega_1)$	$S_n^*(1,\omega_2)$	$S_n^*(1,\omega_3)$
1	6	6	8	4
2	10	13	9	8

We rewrite the objective problem as

$$\begin{aligned}&\mathbb{E}u(v) - \lambda \mathbb{E}[\tilde{\mathbb{P}}/\mathbb{P}v/B_1] = \mathbb{E}[u(v) - \lambda(\tilde{\mathbb{P}}/\mathbb{P})v/B_1] \\ &= \sum_\omega \mathbb{P}(\omega)[u(v(\omega)) - \lambda(\tilde{\mathbb{P}}(\omega)/\mathbb{P}(\omega))v(\omega)/B_1(\omega)].\end{aligned}$$

We use the first order optimality condition: if v is the maximum, then v must satisfy

$$u'(v(\omega)) = \lambda\tilde{\mathbb{P}}(\omega)/(\mathbb{P}(\omega)B_1(\omega)), \text{ for all } \omega \in \Omega.$$

This, together with the definition of $\tilde{\mathbb{P}}$, yields

$$\lambda = \mathbb{E}[B_1 u'(w_1)],$$

where w_1 is the optimal solution

Also, the optimal solution w_1 verifies

$$w_1(\omega) = (u')^{-1}(\lambda(\tilde{\mathbb{P}}(\omega)/\mathbb{P}(\omega))/B_1(\omega)),$$

where $(u')^{-1}$ denotes the inverse function of u'. We use the last result to derive a condition on λ:

$$w = \mathbb{E}^{\tilde{\mathbb{P}}}[w_1/B_1] = \mathbb{E}^{\tilde{\mathbb{P}}}[(u')^{-1}(\lambda(\tilde{\mathbb{P}}/\mathbb{P})/B_1)/B_1].$$

The second step is easy: you just compute the trading strategy H that replicates the contingent claim w_1, that is, the contingent claim that will have at time $t = 0$, the price w.

12.1.3 Example:

Take $N = 2$, $K = 3$, a constant interest rate $r = 1/9$. Take $\mathbb{P}(\omega_1) = 1/2, \mathbb{P}(\omega_2) = 1/4, \mathbb{P}(\omega_3) = 1/4$. Note that here, $B_1 = (1 + r)$. The price process is given by Take $u(v) = -\exp(-v)$. So $(u')^{-1}(y) = -\ln y$.

Also, we use the notation $L = \tilde{\mathbb{P}}/\mathbb{P}$. In order to compute $\tilde{\mathbb{P}}$, we solve the system of equations characterizing $\tilde{\mathbb{P}}$

$$\mathbb{E}^{\tilde{\mathbb{P}}}[\Delta S_n^*] = \sum_{\omega \in \Omega} \tilde{\mathbb{P}}(\omega)\Delta S_n^*(\omega) = 0,\ n = 1, \cdots, N.$$

The previous system of equations is equivalent to

$$\begin{aligned}
&0\tilde{\mathbb{P}}(\omega_1) + 2\tilde{\mathbb{P}}(\omega_2) - 2\tilde{\mathbb{P}}(\omega_3) = 0\\
&3\tilde{\mathbb{P}}(\omega_1) - 1\tilde{\mathbb{P}}(\omega_2) - 2\tilde{\mathbb{P}}(\omega_3) = 0\\
&\tilde{\mathbb{P}}(\omega_1) + \tilde{\mathbb{P}}(\omega_2) + \tilde{\mathbb{P}}(\omega_3) = 1.
\end{aligned}$$

The unique solution is $(1/3, 1/3, 1/3)$. We deduce

$$L(\omega_1) = 2/3, L(\omega_2) = L(\omega_3) = 4/3.$$

We then use the formulae derived above for the optimal solution

$$w_1(\omega) = -\ln(\lambda L/B_1).$$

Moreover λ is characterized by

$$w = \mathbb{E}^{\tilde{\mathbb{P}}}[-\ln(\lambda(\tilde{\mathbb{P}}/\mathbb{P})/B_1)/B_1] = -\ln(\lambda)\mathbb{E}^{\tilde{\mathbb{P}}}[1/B_1] - \mathbb{E}^{\tilde{\mathbb{P}}}[\ln(L/B_1)/B_1].$$

This yields

$$\lambda = \exp\{\frac{-w - \mathbb{E}^{\tilde{\mathbb{P}}}[\ln(L/B_1)/B_1]}{\mathbb{E}^{\tilde{\mathbb{P}}}[1/B_1]}\}.$$

We can then deduce

$$w_1 = \frac{w + \mathbb{E}^{\tilde{\mathbb{P}}}[\ln(L/B_1)/B_1]}{\mathbb{E}^{\tilde{\mathbb{P}}}[1/B_1]} - \ln(L/B_1).$$

We can substitute the value of w_1 found above into the definition of the objective function and hence derive the optimal value of the objective function:

$$u(w_1) = -\lambda L/B_1.$$

This yields

$$\mathbb{E}[u(w_1)] = -\lambda\mathbb{E}[L/B_1] = -\lambda\mathbb{E}^{\tilde{\mathbb{P}}}[1/B_1].$$

Furthermore, note that $B_1 = (1 + r) = 10/9$.

Using the numerical values, we find

$$\mathbb{E}^{\tilde{\mathbb{P}}}[\ln(L/B_1)] = -0.04873.$$

The optimal wealth is given by

$$w_1 = w(1+r) + \mathbb{E}^{\tilde{\mathbb{P}}}[\ln(L/B_1)] - \ln(L/B_1).$$

So the optimal wealth w_1 is $w(10/9) + 0.46209$ for $\omega{=}\omega_1$ and is $w(10/9) - 0.23105$ for $\omega = \omega_2$ or ω_3. The optimal objective value is

$$-\lambda/(1+r) = -9/10\lambda.$$

We simply need to find the trading strategy replicating w_1. It solves

$$w_1/B_1 = w + H_1 \Delta S_1^* + H_2 \Delta S_2^*.$$

H_1, H_2 solves the system of 3 equations

$$\begin{aligned} 0.41590 &= 0H_1 + 3H_2 \\ -0.20795 &= 2H_1 - H_2 \\ -0.20795 &= -2H_1 - 2H_2 \end{aligned}$$

The solution is $H_1 = -0.03466$, $H_2 = 0.13863$. We then compute H_0 which verifies

$$w = H_0 + 6H_1 + 10H_2.$$

This yields

$$H_0 = w - 1.17834.$$

12.2 Risk and return

Let $\mathbb{Q}$ be a risk-neutral probability measure and $\omega \in \Omega$. The ratio $\mathbb{Q}(\omega)/B_1(\omega)$ is often called the state price of ω and the ratio $L = \mathbb{Q}(\omega)/\mathbb{P}(\omega)$ is called the *state price density*. We have seen that the return for the bank account is

$$R_0 = \frac{B_1 - B_0}{B_0}.$$

The return for the risk asset number n is

$$R_n = \frac{S_n(1) - S_n(0)}{S_n(0)},\ n = 1, \cdots, N.$$

Moreover, we have

$$\begin{aligned}
S_n^*(1) - S_n^*(0) &= \frac{S_n(1) - B_1 S_n(0)}{B_1} \\
&= \frac{[1 + R_n]S_n(0) - [1 + R_0]S_n(0)}{1 + R_0} \\
&= S_n(0)\Big(\frac{R_n - R_0}{1 + R_0}\Big).
\end{aligned}$$

If $\mathbb{Q}$ is such that $\mathbb{Q}(\omega) > 0$, for all $\omega \in \Omega$, then $\mathbb{Q}$ is a risk-neutral probability measure if and only if

$$\mathbb{E}^{\mathbb{Q}}\Big[\frac{R_n - R_0}{1 + R_0}\Big] = 0,\ n = 1, \cdots, N.$$

When $R_0 = r$ is deterministic, then the above equality becomes

$$\mathbb{E}^{\mathbb{Q}}[R_n] = r,\ n = 1, \cdots, N.$$

We define the mean return for the asset n:

$$\overline{R}_n = \mathbb{E}[R_n].$$

We can see that

$$cov(R_n, L) = \mathbb{E}[R_n L] - \mathbb{E}[R_n]\mathbb{E}[L] = \mathbb{E}^{\mathbb{Q}}[R_n] - \mathbb{E}[R_n] = r - \overline{R}_n.$$

The quantity $\overline{R}_n - r$ is called the risk premium for the security.

Consider now the return R of a portfolio corresponding to the trading strategy $(H_0, H_1, \cdots, H_N)$. This is

$$R = \frac{v_1 - v_0}{v_0}.$$

Note the the portfolio's gain $G = v_1 - v_0$ can be written as

$$G = H_0(1 + r) - H_0 + \sum_{n=1}^{N} H_n \Delta S_n = H_0 r + \sum_{n=1}^{N} H_n S_n(0) R_n.$$

So R can also be expressed as

$$R = \frac{H_0}{v_0} r + \sum_{n=1}^{N} [\frac{H_n S_N(0)}{v_0}] R_n.$$

We can then check that

$$\overline{R} - r = -cov(R, L),$$

where

$$\overline{R} = \mathbb{E}[R].$$

Consider the contingent claim $a + bL$, where $a, b \neq 0$ are two real numbers. Assume that there is a trading strategy H' that can replicate the claim, i.e., there is a trading strategy H', such that

$$v_1' = a + bL,$$

where v_1' is the wealth at the end of the period when the strategy H' is applied. We also have

$$v_0'(1 + R') = a + bL,$$

where R' is the return of the portfolio corresponding to the strategy H'.

Consider now the return R corresponding to an arbitrary trading strategy H. It is then easy to show that

$$cov(R, L) = \frac{v_0'}{b} cov(R, R').$$

Indeed, on the one hand we have

$$cov(a + bL, R) = bcov(R, L),$$

and on the other hand,

$$cov(a + bL, R) = cov(v_0'(1 + R'), R)) = v_0' cov(R, R').$$

We deduce that

$$\overline{R} - r = -\frac{v_0'}{b} cov(R, R').$$

In particular, when $H = H'$, we obtain

$$\overline{R}' - r = -\frac{v_0'}{b} cov(R', R') = -\frac{v_0'}{b} var(R').$$

We deduce from the last 2 equalities

$$\overline{R} - r = \frac{cov(R, R')}{var(R')}(\overline{R}' - r).$$

We call the ratio

$$\beta = \frac{cov(R, R')}{var(R')}$$

the beta of the trading strategy H with respect to the trading strategy H'. The interpretation of last slide's result is: the risk premium of H is proportional to the risk premium of H' and the constant is given by the beta. It looks like the Capital Asset Pricing Model, except that here, H' corresponds to the contingent claim $a + bL$ instead of the market portfolio.

12.3 The mean-variance portfolio problem

The mean-variance portfolio problem is a very well-known problem. In this section, we assume that the interest rate r, such that $B_1 = 1 + r$, is deterministic. We assume that there is no arbitrage opportunity. We finally assume that there exists a portfolio, whose expected rate of return is different from the interest rate r. The rate of return of the portfolio is denoted by R.

A classic mean-variance problem is

$$\min var(R)$$

subject to

$$\mathbb{E}[R] = \rho,$$

where $\rho \geq r$ is specified.

If $\rho = r$, the optimal value of the objective function is 0. Indeed, the optimum portfolio is entirely invested in the non risky asset, whose rate of return is r and deterministic. Hence, the variance of the portfolio is 0, and its rate of return is r.

Now, in general, if $\rho \geq r$, the optimization problem is well defined. If $\rho > r$, the solution will be a finite positive number. The return of the portfolio can be written as

$$R = \frac{H_0(1+r) - H_0 + \sum_{n=1}^{N} H_n(S_n(1) - S_n(0))}{v_0}$$
$$= \frac{H_0}{v_0} r + \sum_{n=1}^{N} \frac{H_n S_n(0)}{v_0} R_n,$$

where R_n is the rate of return of the asset number n.

In particular, one can see from the previous formula, that for any $\rho > r$, one can construct a feasible portfolio. You just need to pick the right combination of the riskless asset and of a portfolio with a given return $R \neq r$. We denote by S the value of the portfolio with return R. You need to choose (H_0, H) such that

$$\rho = \frac{H_0}{v_0} r + \frac{HS(0)}{v_0} \mathbb{E}R,$$

where $v_0 = H_0 + HS(0)$. We are going to solve this problem, by using the *martingale approach* once more. To this end, we replace the original problem by an equivalent problem.

We look at the minimization problem

$$\min var(v_1)$$

subject to

$$\mathbb{E}[v_1] = v(1+\rho)$$
$$v_0 = v$$

R is a solution of the first problem if and only if $v_1 = v(1+R)$ is a solution of the second problem.

Proof: Let v_1 be a solution of the second problem and consider the portfolio $R = (v_1 - v)/v$. Then, first of all, R satisfies the constraint of the first problem

$$\mathbb{E}[R] = \mathbb{E}[(v_1 - v)/v] = \frac{1}{v}\mathbb{E}[v_1] - 1 = \frac{v(1+\rho)}{v} - 1 = \rho.$$

Moreover, consider any other return μ, that satisfies the constraint $\mathbb{E}[\mu] = \rho$. Then, $v_1^\mu = v(1+\mu)$ is a feasible solution of the second problem. Indeed, we have

$$\mathbb{E}[v_1^\mu] = \mathbb{E}[(1+\mu)v] = v(1+\rho).$$

Next, we have the estimate

$$var(R) = \frac{1}{v^2} var[v_1 - v] = \frac{1}{v^2} var[v_1] \leq \frac{1}{v^2} var[v_1^\mu] = var[\mu],$$

which shows that R is the solution of the first problem. The converse statement can be shown in a very similar way.

We can solve the second problem by using a Lagrange multiplier. This consists in solving

$$\min var[v_1] - \beta \mathbb{E}[v_1],$$

subject to the constraint $v_0 = v$, where β is such that $\mathbb{E}[v_1] = v(1+\rho)$. We actually replace the above problem by the equivalent problem

$$\max \mathbb{E}[-\frac{1}{2} v_1^2 + \beta v_1],$$

subject to

$$v_0 = v,$$

which is of the same form as the problem solved earlier by using the martingale approach.

Indeed, the particular utility function here is given by

$$u(v_1) = -\frac{1}{2} v_1^2 + \beta v_1,$$

and is differentiable and concave. Note that is not strictly increasing and is therefore not often used as a utility function in the applications, but the methodology we saw still applies. We need to verify that the 2 problems are equivalent and to give the solution of this problem. It is easy to apply the previous results and show that the solution of the optimization problem is given by

$$\hat{v_1} = \frac{\beta}{\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}]} (\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] - \tilde{\mathbb{P}}/\mathbb{P}) + v(1+r) \frac{\tilde{\mathbb{P}}/\mathbb{P}}{\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}]}.$$

Indeed, since $(u')^{-1}(y) = \beta - y$, the optimal wealth is given by

$$\hat{v_1} = \beta - \lambda L/(1+r).$$

Moreover, the condition that λ must verify is

$$\mathbb{E}^{\tilde{\mathbb{P}}}[(\beta - \lambda L/B_1)/B_1] = v_0.$$

Hence

$$\lambda = (1+r)(-v_0(1+r)+\beta)/\mathbb{E}^{\tilde{\mathbb{P}}}[L].$$

This gives us

$$\hat{v_1} = \beta - L(-v_0(1+r)+\beta)/\mathbb{E}^{\tilde{\mathbb{P}}}[L],$$

which is equivalent to the formula given in the previous slide. Furthermore,

$$\mathbb{E}[\hat{v_1}] = \beta(\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] - 1)/\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] + v(1+r)/\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}].$$

Then, if $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are not identical, we have

$$\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] > 1.$$

We deduce that $\mathbb{E}[\hat{v_1}] = v(1+\rho)$ if and only if

$$\beta = \frac{v[(1+\rho)\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] - (1+r)]}{\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] - 1}.$$

For the value of β above, v_1 is feasible for the mean-variance problem.

Furthermore, what we can show is that, for this value of β, the optimization problem we just solved is equivalent to the mean-variance optimization problem.

We still need to provide a complete argument for the above statement. Consider the solution of the modified problem we just computed, $\hat{v_1}$, and any other random variable v_1, such that $\mathbb{E}[v_1] = v(1+\rho)$. Since $\mathbb{E}[v_1] = \mathbb{E}[\hat{v_1}]$ and

$$\mathbb{E}[-\frac{1}{2}v_1^2 + \beta v_1] \leq \mathbb{E}[-\frac{1}{2}\hat{v_1}^2 + \beta\hat{v_1}],$$

then

$$var(\hat{v_1}) \leq var(v_1),$$

and consequently, $\hat{v_1}$ is the solution of the mean-variance optimization problem.

Conversely, suppose that $\hat{v_1}$ is solution of the mean-variance problem with β related to ρ as in previous slide. Consider another random variable v_1 that is feasible for the mean-variance problem. Then,

$$\mathbb{E}[-\frac{1}{2}\hat{v_1}^2 + \beta\hat{v_1}] \geq \mathbb{E}[-\frac{1}{2}v_1^2 + \beta v_1],$$

and since any solution of the modified problem must be feasible for the mean variance problem, $\hat{v_1}$ must be the optimal solution of the modified problem.

12.3.1 Establishing CAPM

Substituting the value of β into the formula for the optimal solution $\hat{v_1}$, we compute the return $\hat{R}$ corresponding to $\hat{v_1}$:

$$\hat{R} = \frac{\rho\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] - r}{\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] - 1} - \frac{\rho - r}{\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\mathbb{P}}/\mathbb{P}] - 1}\tilde{\mathbb{P}}/\mathbb{P}.$$

The computation leading to the above formula is not particularly tricky but really lengthy and I skip it. The optimal solution $\hat{R}$ of the mean-variance portfolio problem is an affine function of the state price density $L = \tilde{\mathbb{P}}/\mathbb{P}$. Furthermore, in regard of the section on *risk and return*, this establishes the Capital Asset Pricing Model theory.

CAPM: If $\hat{R}$ is the solution of the mean-variance portfolio problem for $\rho \geq r$ and if R is the return of an arbitrary portfolio, then

$$\mathbb{E}[R] - r = \frac{cov(R, \hat{R})}{var(\hat{R})}(\mathbb{E}[\hat{R}] - r).$$

The solution of the mean-variance problem can be a stock index whose mean return can be estimated. Then the CAPM gives us a way of computing the mean return of any arbitrary portfolio. Consider a fixed portfolio whose return is a solution of the mean variance problem for some return $\hat{\rho} > r$, typically this would be a*mutual fund*. Then the solution of the mean variance problem can be achieved for any other mean return by a portfolio consisting of investments in the risk less security and the fixed portfolio (i.e. *mutual fund*). This is a very nice results but be aware it only applies to investors who have a quadratic utility function.

Proof: An investor puts a fraction λ in the risk less security and the balance $1-\lambda$ in the mutual fund, where

$$\lambda = \frac{\hat{\rho} - \rho}{\hat{\rho} - r},$$

where $\rho \geq r$. This portfolio's return is

$$R = \lambda r + (1-\lambda)\hat{R}',$$

and consequently

$$\mathbb{E}[R] = \rho.$$

Chapter 13

Module 13: Complements in Option Pricing: The American option, dividends, the Greeks and gamma neutral hedging strategy

In these notes, I draw from Andrew Papanicolaou's set of slides, and the articles by Casagna, Mercurio, and by Raju.

13.1 American options

An American call option has the additional feature of being exercised at any time prior to T. Let $C_A(t)$ denote the price of the call at time t with strike K and maturity T. At any time $t \leq T$, given the current value of the underlying asset, we have to decide whether to exercise the option at time t, or continue holding it, and the price of the American call option can be thought of as

$$C_A(t) = \max(S(t) - K, C_{cont}(t)),$$

where $C_{cont}(t)$ represents the continuation value.

Since we have the option of continuing until the end, the value of the American call option is greater or equal to the value of the European call option.

The following heuristic argument may help see this: consider a time t very close to the maturity date $T > 0$. Then, either you exercise at time t, or you wait until the maturity date T. Thus,

$$C_A(t) = \max(S(t) - K, C_{cont}(t)) \geq C_{cont}(t) = C_E(t),$$

where $C_E(t)$ is the price of the European call. You can iterate this argument backward in time to convince yourself that the price of the American option is greater or equal to the price of the European option.

Next, we are going to show intuitively that it is never optimal to exercise an American call option without dividend early.

13.1.1 The case of the American call without dividend

We let $t < T$. By put-call parity for European options,

$$\begin{aligned}
C_A(t) &= \max(S(t) - K, C_{cont}(t)) \\
&\geq C_E(t) \\
&= P_E(t) + S(t) - Ke^{-r(T-t)} \\
&> S(t) - Ke^{-r(T-t)} \\
&\geq S(t) - K,
\end{aligned}$$

which represents the value of early exercise at time t.

Hence,

$$\max(S(t) - K, C_{cont}(t)) > S(t) - K, \text{ or all } t < T$$

so that

$$C_A(t) = C_{cont}(t) = C_E(t),$$

for all $t \leq T$, and it is never optimal to exercise early a call, on a non-dividend-paying asset.

Example to illustrate the above argument

Suppose that you are long an *in the money* American call (i.e. $S(t) > K$) and that you predict that the value of the asset will go down and leave the call *out of the money*, i.e. such that $S(T) < K$. In this situation, you should not exercise. You could short the asset, and hedge your short position by holding the call.

For example, suppose that the asset is trading at \$105, you are long a 3-month American call with $K = \$100$, and the risk-free rate is $r = 0.05$. We consider two strategies:

1. You exercise the option immediately at time t, and invest the proceeds as follows
    - Pay $K = \$100$ for one share of the asset and sell the asset at \$105.
    - Invest proceeds of \$5 in a 3-month maturity bond.
    - In 3 months, your cash position is \$5.06.
2. Short the asset, hold the option, and invest the proceeds:
    - At time t, short the asset at \$105 and invest the proceeds in a 3-month maturity bond with rate $r = 0.05$.
    - In 3 months: sell the bond to collect \$106.32 in your bank account; you owe no more than \$100 to cover the short position. You end the period with a cash position of at least \$6.32.

The second strategy has at least an edge of \$1.26.

Alternately, one could simply sell the call, which would also be better than exercising early but you would be giving up the time-value of holding it.

13.1.2 Dividends

We saw earlier the put-call party in the case without dividend.

Consider now an asset that pays a dividend $D(T)$, at time T, to a cash account.

At time T, the put-call parity relation reads

$$(K - S(T))^+ + S(T) = (S(T) - K)^+ + K.$$

Taking the discounted expectation under the risk-neutral probability measure of the left- and right-hand sides in the above equality, yields

$$e^{-rT}\tilde{\mathbb{E}}[(K - S(T)^+] + e^{-rT}\tilde{\mathbb{E}}[S(T)] = e^{-rT}\tilde{\mathbb{E}}[(S(T) - K)^+] + e^{-rT}K.$$

In the case with dividend, $e^{-rt}(S(t) + D(t))$ is a martingale under the risk-neutral probability measure.

Since, by the martingale property,

$$e^{-rT}\tilde{\mathbb{E}}[S(T)] = S(0) + D(0) - e^{-rT}\tilde{\mathbb{E}}[D(T)],$$

we have the following put-call parity relation, in the case with dividends

$$P_E(0) + S(0) + D(0) - e^{-rT}\tilde{\mathbb{E}}[D(T)] = C_E(0) + e^{-rT}K,$$

where $P(0)$ and $C(0)$ denote the respective prices of the put and call options at time 0 and the dividends are held in the cash account.

Furthermore, we can specify the asset price model. For instance, we assume that the dividends are distributed continuously at the rate $\delta > 0$.

The Black-Scholes continuous-time asset price model becomes in this case

$$dS(t) = (r - \delta)S(t)dt + \sigma S(t)d\tilde{W}(t),$$

under the risk-neutral probability measure.

If the dividend is invested in the cash account, the Stochastic Differential Equation describing the evolution of the dividend process $D(t)$ is given by

$$dD(t) = (rD(t) + \delta S(t))dt, D(0) = 0.$$

Note that we assumed that $D(0) = 0$.

We can deduce the SDE for $e^{-rt}D(t)$:

$$d(e^{-rt}D(t)) = \delta e^{-rt}S(t)dt.$$

We integrate the above SDE from 0 to t

$$e^{-rt}D(t) = D(0) + \delta \int_0^t e^{-ru}S(u)du,$$

and take the expectation of the left- and right-hand sides:

$$\tilde{\mathbb{E}}[e^{-rt}D(t)] = D(0) + \delta \int_0^t \tilde{\mathbb{E}}[e^{-ru}S(u)]du = \delta \int_0^t (S(0) - e^{-ru}\tilde{\mathbb{E}}[D(u)]du.$$

Then, we let $m(t) = e^{-rt}\tilde{\mathbb{E}}[D(t)]$ and differentiate the above equality to derive the ODE satisfied by m:

$$m'(t) = \delta S(0) - \delta m(t), m(0) = 0.$$

This gives us the result

$$m(t) = e^{-rt}\tilde{\mathbb{E}}[D(t)] = S(0)(1 - e^{-\delta t}).$$

The put-call parity becomes in this case

$$S(0)e^{-\delta t} + P_E(t) = C_E(t) + e^{-rt}K.$$

13.1.3 American call option with dividend

It may be optimal to exercise early an American call option on an dividend paying asset. To see this, we write the martingale property

$$\tilde{\mathbb{E}}[e^{-r(T-t)}(D(T) + S(T)|\{W(s), t \leq s \leq T\}] = D(t) + S(t),$$

and generalize the previous put-call parity equation to

$$P_E(t)+S(t)+D(t)-e^{-r(T-t)}\tilde{\mathbb{E}}[D(T)|\{W(s), 0 \leq s \leq t\}] = C_E(t)+e^{-r(T-t)}K,$$

to estimate $C_A(t)$

$$\begin{aligned} C_A(t) &\geq C_E(t) = \\ & \quad P_E(t) + S(t) + D(t) - Ke^{-r(T-t)} - e^{-r(T-t)}\tilde{\mathbb{E}}[D(T)|\{\tilde{W}(s), 0 \leq s \leq t\}] \\ &> S(t) - K + D(t) - e^{-r(T-t)}\tilde{\mathbb{E}}[D(T)|\{\tilde{W}(s), 0 \leq s \leq t\}], \end{aligned}$$

but it might be that $C_E(t) < S(t) - K$ if the dividend rate increases on average between t and T (submartingale).

13.1.4 American put option

If the interest rate is positive, it may be optimal to exercise early an American put option on a non-dividend-paying asset.

We argue by contradiction. We suppose that the above statement is false. Denoting by $P_A(t)$ and $P_E(t)$ the respective values at time t of the American and European puts, we would have

$$P_A(t) = \max(K - S(t), P_{cont}(t)) = P_E(t), \text{ for all } t \geq 0.$$

However, for $r > 0$, suppose that, for some t, $S(t) < K$. Then, we may have $P_E(t) < K - S(t)$ because $\tau \to e^{-r(\tau-t)}(K - S(\tau))^+$ is not necessarily a submartingale.

The above strict inequality contradicts the equality $\max(K - S(t), P_{cont}(t)) = P_E(t)$. Hence,

$$P_A(t) = \max(K - S(t), P_{cont}(t)) > P_E(t).$$

The standard Black-Scholes theory does not work for pricing American puts; pricing American puts is rather complex and requires an extra layer of technicalities.

13.2 The Greeks

Generally, they measure the sensitivity of an option price to changes in the variables and parameters. Here, we focus on three Greeks, the Delta, Gamma and Theta. We denote by $c(t, s)$ the price of a call option at time t, and for an underlying security's price s and the Greeks are functions of the time and of the price of the underlying defined as

$$\begin{aligned}
\Delta(t, s) &= c_s(t, s) > 0, \\
\Gamma(t, s) &= c_{ss}(t, s) > 0, \\
\theta(t, s) &= -c_t(t, s).
\end{aligned}$$

We can rewrite the Black-Scholes PDE

$$c_t + \frac{1}{2}\sigma^2 s^2 c_{ss} + rsc_s - rc = 0,$$

in terms of the above three Greeks,

$$\frac{1}{2}\sigma^2 s^2 \Gamma = \theta + r(c - s\Delta).$$

Furthermore, the Greeks can be interpreted in the following manner. First of all, $\Delta = c_s$ is the hedge, as seen in the lecture on continuous-time derivative's pricing, and measures the sensitivity of the option price to changes in the underlying asset's price. Secondly, $\Gamma = c_{ss}$ measure the sensitivity of the hedge to changes in the underlying asset's price. Thirdly, $\theta = -c_t$ is the time value.

Usually, $\theta > 0$, and in particular, for a call option, $\theta > 0$. Exceptionally, θ may be negative, in particular for a low-strike put. Furthermore, θ can also be seen as the premium earned by the hedge compared to the benchmark given by the risk-free rate. In order to see this, consider a portfolio that is long the call option and short the Δ hedge portfolio. The time-evolution of the hedging portfolio with value $X(t)$, at time t, is

$$dX(t) = \Delta(t, S(t))dS(t) + r(X(t) - \Delta(t, S(t))S(t))dt,$$

and the value of the portfolio that is long the call option and short the hedge is

$$
\begin{aligned}
d(c(t,S(t)) - X(t)) &= dc(t,S(t)) - \Delta(t,S(t))dS(t) - r(X(t) - \Delta(t,S(t))S(t))dt, \\
&= (-\theta(t,S(t)) + \frac{1}{2}\sigma^2 S(t)^2 \Gamma(t,S(t)))dt + \Delta(t,S(t))dS(t) \\
&- \Delta(t,S(t))dS(t) - r(X(t) - \Delta(t,S(t))S(t))dt, \\
&= \left(-\theta(t,S(t)) - r(X(t) - \Delta(t,S(t))S(t)) + \frac{1}{2}\sigma^2 S(t)^2 \Gamma(t,S(t))\right) dt.
\end{aligned}
$$

13.2.1 The Black-Scholes Greeks

The three Greeks introduced above can be computed analytically for a European call, from the Black-Scholes analytical formula. We obtain

$$
\begin{aligned}
\Delta(t,s) &= N(d_+(T-t,s)), \\
\Gamma(t,s) &= N'(d_+(T-t,s))/(s\sigma\sqrt{T-t}) \\
\theta &= \frac{1}{2}s\sigma N'(d_+(T-t,s))/\sqrt{T-t} + rKe^{-r(T-t)}N(d_-(T-t,s))
\end{aligned}
$$

Similarly, for a European put, these Greeks are equal to

$$
\begin{aligned}
\Delta(t,s) &= N(d_+(T-t,s) - 1, \\
\Gamma(t,s) &= N'(d_+(T-t,s))/(s\sigma\sqrt{T-t}) \\
\theta &= \frac{1}{2}s\sigma N'(d_+(T-t,s))/\sqrt{T-t} - rKe^{-r(T-t)}N(-d_-(T-t,s))
\end{aligned}
$$

Furthermore, we deduce from the formulae above that, for a European call, $0 \leq \Delta \leq 1$ whereas for a put, $-1 \leq \Delta \leq 0$. For both the European call and put, $\Gamma > 0$.

13.3 Hedging

In this section, we explore various aspects of hedging, such as the effect of time discretization on Delta hedging and the need for Delta-Gamma hedging.

13.3.1 The effect of time discretization

In practice, we apply the hedging strategy in discrete time and the time-discretization is going to produce an error, and thus introduce some risk in the hedging portfolio.

We recall the dynamics of hedging a call option

$$d(c(t,S(t))-X(t)) \quad = \quad dc(t,S(t))-\Delta(t,S(t))dS(t)-r(X(t)-\Delta(t,S(t))S(t))dt.$$

After discretizing the time with a step $\Delta t>0$, the above equation is approximated as

$$\begin{aligned}&c(t+\Delta t,S(t+\Delta t))-X(t+\Delta t)-c(t,S(t))+X(t)\approx\\&c(t+\Delta t,S(t+\Delta t))-c(t,S(t))-\Delta(t,S(t))(S(t+\Delta t)-S(t))-r(X(t)-\Delta(t,S(t))S(t))\Delta t\end{aligned}$$

As we can see, even if we apply the Delta hedge $\Delta=c_s$, we cannot completely eliminate risk because of the truncation error.

13.3.2 Delta neutral

We can compute the Greeks of any given portfolio. In particular one can determine the Delta of any portfolio. Consider the portfolio which is long a call option and short the Delta hedge. We denote its value at time t by $V(t)$. We also denote the hedging process by $\Delta_c(t)$. We have

$$V(t)=c(t,S(t))-\Delta_c(t)S(t).$$

We also define the function v such that $v(t,s)=V(t)$ when $S(t)=s$

The Delta of this portfolio is computed by taking the derivative of its value with respect to the price of the underlying. We obtain

$$\Delta_V(t)=c_s(t,S(t))-\Delta_c(t)=0.$$

Since the Delta of this portfolio is equal to 0, this portfolio is said to be delta-neutral. The above delta-neutral portfolio is sensitive to changes in the underlying asset, especially when the Gamma of the call option is large, which occurs when the call option is at the money.

When the Gamma is large, the theoretical hedging portfolio undergoes large moves and, in practice, the hedging portfolio must be rebalanced frequently.

Ideally, it would be better to construct a hedging portfolio, which is both delta neutral and gamma neutral. However, this is not possible in the current basic Black-Scholes framework that we are using as this would require the introduction of an additional call option.

13.3.3 Delta neutral and gamma neutral

We introduce another call option on the same underlying to hedge with. Both options have the same maturity date $T > 0$ and different strike prices K, and K'. The two option prices are respectively denoted by $c(t, s)$ and $c'(t, s)$.

This time, we hold in the hedging portfolio α shares of underlying asset and β European calls with strike K'. As before, purely formally, and without a proof, we equate the evolution of the hedging portfolio and of the European call with strike K:

$$\begin{aligned} dc(t, S(t)) - dX(t) &= c_t dt + c_s dS(t) + \frac{1}{2} c_{ss} dS(t) dS(t) \\ &- \alpha dS(t) - \beta dc'(t) - r(X(t) - \alpha S(t) - \beta c')dt \\ &= c_t dt + c_s dS(t) + \frac{1}{2} c_{ss} dS(t) dS(t) \\ &- \alpha dS(t) - \beta c'_t dt - \beta c'_s dS(t) - \beta \frac{1}{2} c'_{ss} dS(t) dS(t) \\ &- r(X(t) - \alpha S(t) - \beta c')dt \end{aligned}$$

We cancel the second derivatives by setting

$$\beta = c_{ss}/c'_{ss} = \Gamma/\Gamma',$$

and we choose

$$\alpha = c_s - \beta c'_s,$$

in order to cancel the Brownian terms (dS terms).

This improves the hedge by making it less sensitive to the fluctuations of the process S:

$$\begin{aligned} dc(t,S(t)) - dX(t) &= c_t dt - \beta c'_t dt - r(X(t) - \alpha S(t) - \beta c')dt \\ &= (-\theta + \beta\theta' - r(X(t) - \alpha S(t) - \beta c'))dt. \end{aligned}$$

The above equation does not contain any Γ term!

Finally, consider the portfolio which is long the call with strike K and short the hedge

$$H(t) = c(t,S(t)) - \alpha S(t) - \beta c'(t,S(t)),$$

and compute its Delta and Gamma, respectively denoted by Δ_H and Γ_H, by differentiating H twice with respect to the underlying S

$$\Delta_H(t) = c_s - \alpha - \beta c'_s = 0, \Gamma_H(t) = c_s s - \beta c'_{ss} = 0.$$

The above portfolio is delta and gamma neutral!

13.3.4 Profit and loss

In this subsection, we assess the profit and loss resulting from holding a call option and implementing the delta or delta-gamma hedging strategy. We consider the holder of a European call option.

Suppose that the holder of the option does not hedge his long position. His profit and loss, at time T, is then the payoff minus the cost of buying the option

$$P\&L(T) = (S(T) - K)^+ - e^{rT}C(0,S(0)).$$

Now, if the holder of the option hedges his position, his profit and loss changes to

$$P\&L(T) = (S(T) - K)^+ - X(T),$$

where $X(T)$ represents the value at time T of the hedging portfolio.

Delta hedge

As seen earlier, the profit and loss incurred by applying a Delta hedge evolves according to

$$dc(t,S(t)) - dX(t) = \left(-\theta + \frac{1}{2}\Gamma\sigma^2 S^2 - r(X(t) - \Delta S(t))\right) dt.$$

We say that the option holder is long $\Gamma = c_{ss} > 0$ by convexity of the call option and that the hedger is long $\theta = -c_t$.

Delta-gamma hedge

Going back to an earlier calculation, the profit and loss incurred by applying a delta-gamma hedge evolves according to

$$\begin{aligned}
dc(t,S(t)) - dX(t) &= c_t dt + c_s dS(t) + \frac{1}{2} c_{ss} dS(t) dS(t) \\
&- \alpha dS(t) - \beta c'_t dt - \beta c'_s dS(t) - \beta \frac{1}{2} c'_{ss} dS(t) dS(t) \\
&- r(X(t) - \alpha S(t) - \beta c')dt \\
&= (-\theta_c + \beta\theta_{c'} - r(X(t) - \alpha S(t) - \beta c'))dt,
\end{aligned}$$

where θ_c and θ_c' denote respectively the Theta of the call with strike K and the call with strike K'. The premium over investing in the risk free account is given by the quantity $-\theta_c + \beta\theta_{c'}$.

A numerical example: hedging with calls

Consider the holder of $1,000$ shares of stock whose price is \$100, at time 0. She can hedge her position with puts on the same stock, with strike $K = 100$ and time to maturity 0.2493 (91 days). The interest rate is equal to 5%, and the volatility to 0.2. The parameters, put and call prices and Greeks are reported in Table 13.1.

The value of the hedging portfolio at time 0 is given by

$$H(0) = 1,000S(0) + MP(0),$$

where M must be determined.

In order to make this portfolio delta neutral, we must solve for M the equation

$$1,000 + M\Delta_p(0) = 0.$$

Since the Delta of the put at time 0 is equal to -0.4306, she needs about $2,322$ puts to hedge her position.

Furthermore, we can compute the Gamma of this portfolio

$$\Gamma_H = 2322 \times .0393 = 91 > 0,$$

and its Theta

$$\theta_H = -2322 \times 5.5471 = -12880.36.$$

The profit and loss over the time interval $[0, \Delta t]$ is given by

$$\begin{aligned}
P\&L(\Delta t) &= 2322(P(\Delta t) - P(0)) + 1000(S(\Delta t) - 100) \\
&= H(\Delta t) - H(0) \\
&\approx H_t \Delta t + H_s \Delta S + \frac{1}{2} H_{ss} \Delta S^2 \\
&= -12880.36x\Delta t + \frac{1}{2} 91(S(\Delta t) - 100)^2.
\end{aligned}$$

At the maturity date (for $\Delta t = 0.2493$), the worst case scenario is when $S(\Delta t) = \$100$. In that case, we obtain

$$P\&L(T) = 2322(K - S(T))^+ - P(0)) + 1000(S(T) - 100) = 2322 \times (-3.37) = -7825,$$

which corresponds to a 7.8% loss. It is relatively expensive but a good long term hedge because $\Gamma_H > 0$.

Hedging with puts

Suppose we hedge our position with puts instead of calls. The value of the hedging portfolio at time 0 is now given by

$$H(0) = 1,000S(0) + MC(0),$$

where M must be determined.

In order to make this portfolio delta neutral, we must solve for M the equation

$$1,000 + M\Delta_c(0) = 0.$$

Since the Delta of the call at time 0 is equal to 0.5694, she needs about -1756 puts to hedge her position.

Furthermore, we can compute the Gamma of this portfolio

$$\Gamma_H = -1756 \times .0393 = -69 < 0,$$

and its Theta

$$\theta_H = 1756 \times 10.4852 = 18412.01.$$

The profit and loss over the time interval $[0, \Delta t]$ is given by

$$\begin{aligned} P\&L(\Delta t) &= -1756(C(\Delta t) - C(0)) + 1000(S(\Delta t) - 100) \\ &\approx 18412.01\Delta t - \frac{1}{2}69(S(\Delta t) - 100)^2. \end{aligned}$$

It is a riskier hedge because $\Gamma_H < 0$ and as $S(\Delta t)$ moves away from \$100, the profit and loss deteriorates.

How about delta-gamma hedging with puts?

Now, we hedge with two different puts. The first put has the same parameters as the ones listed in the numerical example whereas the second put has a strike price of \$110, the other parameters being equal to those of the first put (we refer to Table 13.1 for a list a parameters). We denote the respective prices of the puts P_1 and P_2, and we consider their respective Deltas, Δ_{p_1} and Δ_{p_2} and their respective Gammas, Γ_{p_1} and Γ_{p_2}. The agent holds M_1 puts of type 1 and M_2 puts of type 2.

The value of the portfolio at time 0 is given by

$$H(0) = 1000xS(0) + M_1P_1(0) + M_2P_2(0).$$

In order to make this position Delta-Gamma neutral, we set

$$\begin{aligned}
\Delta_H &= 1000 + M_1\Delta_{p_1} + M_2\Delta_{p_2} = 0, \\
\Gamma_H &= M_1\Gamma_{p_1} + M_2\Gamma_{p_2} = 0
\end{aligned}$$

or

$$M_1 \approx -1630 \text{ and } M_2 \approx 2176.$$

Secondly, we introduce the speed p_{sss} of a put with price $p(t, S(t))$. For the benchmark parameters that have been set in the subsection with the numerical example, the speeds for the two puts are respectively -0.011 and 0.020

Since Delta-Gamma hedging cancels the first and second derivatives with respect to the underlying asset, we can write the profit and loss over the time interval $[0, \Delta t]$ in terms of the third derivatives, also called speeds:

$$P\&L(\Delta t) \approx (-M_1\theta_1 - M_2\theta_2)\Delta t + \frac{1}{6}(M_1 \times (\text{speed}_1) + M_2 \times (\text{speed}_2))(S(\Delta t) - 100)^3,$$

where θ_1 and θ_2 are the Thetas of the respective puts, and $\text{speed}_1, \text{speed}_2$ are their respective speeds.

Next, we investigate hedging with two different calls instead of puts. We denote the respective prices of the calls C_1 and C_2, and we consider their respective Deltas, Δ_{c_1} and Δ_{c_2} and their respective Gammas, Γ_{c_1} and Γ_{c_2}. The agent holds M_1 calls of type 1 and M_2 calls of type 2.

The value of the portfolio at time 0 is given by

Option 1	Option 2	
Stock price	\$100	\$100
Exercise price	\$100	\$110
Risk-free rate	0.05	0.05
Time to maturity	0.2493 (91 days)	0.2493
Volatility	0.20	0.20
Call price	\$4.61	\$1.19
Put price	\$3.37	\$9.82
Δ_c	0.5694	0.2178
Γ_c	0.0393	0.0295
θ_c	10.4852	6.9255
Δ_p	-0.4306	-0.7822
Γ_p	0.0393	0.0295
θ_p	5.5471	1.4936

Table 13.1: Parameters of the numerical example

$$H(0) = 1000 \times S(0) + M_1 C_1(0) + C_2 P_2(0).$$

In order to make this position Delta-Gamma neutral, we set

$$\begin{aligned} \Delta_H &= 1000 + M_1 \Delta_{c_1} + M_2 \Delta_{c_2} = 0, \\ \Gamma_H &= M_1 \Gamma_{c_1} + M_2 \Gamma_{c_2} = 0 \end{aligned}$$

or

$$M_1 \approx -3588 \text{ and } M_2 \approx 4789.$$

Finally, we write the profit and loss over the time interval $[0, \Delta t]$ in terms of the speeds:

$$P\&L(\Delta t) \approx (-M_1\theta_1 - M_2\theta_2)\Delta t + \frac{1}{6}(M_1 \times (\text{speed}_1) + M_2 \times (\text{speed}_2))(S(\Delta t) - 100)^3,$$

where θ_1 and θ_2 are the Thetas of the respective calls, and $\text{speed}_1, \text{speed}_2$ are their respective speeds.

13.3.5 Other Greeks

The Vega measure the call price's sensitivity to misspecification of σ,

$$v = c_\sigma.$$

The Vega can be computed from the Black-Scholes analytical formula

$$v = sN'(d_+(T-t,s))\sqrt{T} = s^2\sigma\Gamma T,$$

where, as usual, $s = S(t)$. We see that the Vega is proportional to the Gamma.

Differentiating the option price a second time with respect to σ yields the Volga,

$$c_{\sigma\sigma},$$

whereas differentiating a second time with respect to the underlying defines the Vanna,

$$c_{\sigma s}.$$

Finally, we define the Rho which measures the sensitivity to fluctuations in the interest rate, by differentiating one time with respect to r, i.e.

$$\rho = c_r.$$

Relating Vega to Gamma in a stochastic volatility model

Now, we let the volatility fluctuate randomly over time. The stochastic volatility process is denoted by $\sigma(t)$. For instance, the stochastic volatility or local volatility model could be calibrated from market option prices. In that case, it would be called *implied volatility.*

We apply Ito's formula to differentiate the call option price, which now depends on the additional variable $\sigma(t)$:

$$\begin{aligned} dc(t,S(t),\sigma(t)) &= c_t dt + c_s dS(t) + c_\sigma d\sigma(t) + \frac{1}{2}c_{ss} dS(t) dS(t) \\ &+ \frac{1}{2}c_{\sigma\sigma} d\sigma(t) d\sigma(t) + c_{s\sigma} dS(t) d\sigma(t). \end{aligned}$$

We also add another option with price c', corresponding to a different strike K', and assume that c and c' share the same volatility process $\sigma(t)$. If we construct a delta-gamma neutral hedge in the presence of $\sigma(t)$, the hedging portfolio will also be vega neutral. Indeed,the Vega of the portfolio which is long the call and short the hedge, is given by

$$\begin{aligned} c_\sigma - \beta c'_\sigma &= s^2\sigma\Gamma_c T - \beta s^2 \sigma \Gamma_{c'} T \\ &= 0 \end{aligned}$$

because $\beta = \Gamma_c / \Gamma_{c'}$.

Finally, we compute the dynamics of the profit and loss of the portfolio which is long the call option with price c and short the delta-gamma hedging portfolio using the call option with strike K'

$$\begin{aligned} dP\&L(t) &= (-\theta_c + \beta\theta_{c'} - r(X(t) - \alpha S(t) - \beta c'))dt + (c_{\sigma s} - \beta c'_{\sigma s})dS(t)d\sigma(t) \\ &+ \frac{1}{2}(c_{\sigma\sigma} - \beta c'_{\sigma\sigma})(d\sigma(t))^2. \end{aligned}$$

It is worth mentioning that, similarly, with two additional call options, we could construct a delta neutral strategy that would also be vanna and volga neutral.

Acknowledgement: I thank Andrew Papanicolaou and Monty Essid for their helpful suggestions