Introductory notes for the "Introduction to Financial Mathematics" course

Victor Nistor

Contents

CHAPTER 1

Introduction

1. The goals of the course

This course is an introduction to the theory of "No Arbitrage Pricing," to some of the mathematical theories that it requires, and to some of the mathematical questions that it raises. The theory that will be presented in this course is useful and related to practice, but is often a great simplification of the "real thing." Nevertheless, the material that we will cover is an important stepping stone towards more complicated and more relevant theories.

Comentariu pentru participantii romani: aceste note sunt bazate pe niste note de curs pe care le-am folosit semestrul trecut. Unele din concepte din aceste note sunt mai elementare decat cele ce va fi acoperite in cursul propriu zis. Aceasta este mai ales cazul pentru sectiunea dedicata masurii. De exemplu, noi voi lucra cu masuri generale si vom folosi "continuous time" in a doua parte a cursului. Nu vom avea insa timp sa discutam "stochastic calculus" in acest curs. Voi presupune insa cunoscute majoritatea rezultatelor necesare de stochastic calculus. (Vedeti cursul domnului Profesor Lucian Beznea.) Vom incepe insa cu "discrete time," pentru ca ideile noi sunt mai usor de inteles in acest context.

Carti recomandate.

- Steve Shreve, Stochastic calculus for finance I. The binomial asset pricing model. Springer Verlag, 2004. xvi+187 pp. ISBN: 0-387-40100-8.
- J.B. Hunt and J.E. Kennedy, Financial Derivatives in Theory and Practice, Wiley, 2005.
- M. Baxter and A. Rennie, Financial Calculus: An introduction to Derivative Pricing (Cambridge, UK) 1996.
- Victor Goodman and Joseph Stampfli, The mathematics of finance: modeling and hedging. AMS 2001.
- John Hull, Options, Futures, and other derivatives.

CHAPTER 2

‘No-arbitrage pricing’ and financial markets

This is a very brief introduction to “No-arbitrage pricing” and financial markets.

1. Examples of financial instruments

To make things more concrete, let us start by looking at some examples of more often used financial instruments that will provide some basic examples for the theory developed in this class. In fact, **one of the main goals of this course** is to develop techniques to valuate financial instruments similar to the Futures and Options introduced below.

1.1. Forwards, Futures, and Options. The following is a rather informal and very concise presentation. The student may wish to consult additional sources for more information.

Forward contracts. A *forward contract* is a contract to buy a certain asset at the specified (future) time T for K units of currency (say USD). The time T is called *delivery date* or *exercise time*. The price K is called the *strike price* or *delivery price* and is fixed at the time the contract is signed.

The forward contracts are between two legal entities (banks, companies, private investors, ...). They are traded over the counter (not on exchanges).

Futures contracts. A *futures contract* with delivery date T and strike price K is similar to a forward contract with the same characteristics, the difference being that the futures contracts are traded on exchanges. This means that, typically, the buyer and seller do not know each other and, in any case, do not deal with each other directly.

Examples are *gold futures*, *SP500 futures*, ... It costs nothing to enter a futures or forward contract, but one needs to make margin payments (deposits) to enter a futures contract. Thus in practice the need to make margin payments will make the prices of futures and forward contracts to be different, but we shall ignore this issue and price futures as forward contracts. We shall therefore use Forward and Futures contracts interchangeably.

Options. A *European Call Option* is a contract that gives the buyer of the contract the **right** (but not the obligation) **to buy** a certain asset for the *exercise price* K at the *exercise time* T from the seller of the contract.

By replacing "buy" with "sell" in the definition of a European Call Option, one obtains a European Put Option. Thus, a *European Put Option* is a contract that gives the buyer of the contract the **right** (but not the obligation) **to sell** a certain asset for the *exercise price* K at the *exercise time* T to the seller of the contract.

The futures and options are examples of *derivative securities* (of *financial derivative*, or, simply, *derivatives*, or, yet, *contingent claims*) since their value depends on the value of other assets, from which these contracts *derive* their value. The exercise time is sometimes called *maturity time* or even *expiry time.*

Another important type of derivative securities are the **swaps**, which involve exchanges (swapping) cash flows. The swaps can theoretically be understood (or priced) in terms of futures and options, so they will be ignored in what follows. In practice, the swaps are very important, however.

The "American" Call and Put Options are defined similarly, but can be exercised at any time before the expiry time T or at the expiry time T. Thus for a European option, the exercise time t_{ex} and the expiry time T are the same, whereas for an American option we have $t_{ex} \leq T$.

For simplicity, in the following we shall usually assume that the assets (or *underlyings*) used to define our options are stocks.

A note on the notation. We shall denote typically by T the *expiry* or *maturity* or *final time.* By t we shall denote an arbitrary earlier time $t \leq T$, which is often the present time. When we want to distinguish the present time from the arbitrary time t, we shall denote by t_0 the present time. By $\tau = T - t$ we shall denote the *time to expiry.* Many equations become easier to grasp in the variable τ. About the notation S_t: this is the traditional notation. It may seem more natural to use the notation $S(t)$, in which case we would write $S_k(t)$ for the value of the asset k at time t. However, we regard our economy as a set of possible states $\omega \in \Omega$ of the economy. Then S_t will be a function on its own of the state of the economy. Thus $S_t(\omega)$ will denote the value of our asset (say, stock) at time t in the state ω. It is thus assumed that the knowledge of ω implies the knowledge of all the prices of all the assets at all times t. In fact, one can construct models in which ω exactly encodes all the prices at all times and nothing more. (This will be the case with the Cox-Ross-Rubinstein binomial model to be discussed later on.) While there will be some additional information on Ω (a filtration by σ-algebras), that information can often be derived from all the functions $S_{kt} : \Omega \to \mathbb{R}$ for all k and all t. (By "state of the economy" and "state of the world" we shall mean the same thing.)

Assume that the underlying is a stock with price S_t at time t. The *pay-off* of a futures contract with strike price K and delivery date T is

eq.payoff.F

$$W_T = S_T - K. \tag{1}$$

If $S_T > K$, this is the amount to be gained by buying for K an asset that is traded at that time for S_T: buy for K and immediately sell for S_T. If $S_T < K$, the portfolio actually yields a loss.

Similarly, the pay-off of a European Call Option with strike K and maturity T is given as follows. Let us denote by $|x|_+ := \max\{x, 0\}$ and similarly, $|x|_- := \max\{-x, 0\}$. Then the pay-off of a European Call option with strike K, maturity T, and underlying S is $C_T^E = |S_t - K|_+$, that is

$$C_T = C_T^E = |S_T - K|_+ := \begin{cases} S_T - K & \text{if } S_T > K \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, the pay-off of a European Put Option with strike K, maturity T, and the same underlying S is

$$P_T = P_T^E = |S_T - K|_- := \begin{cases} -S_T + K & \text{if } S_T < K \\ 0 & \text{otherwise.} \end{cases}$$

(Here S is the price of the underlying.) Note that $|x|_+ - |x|_- = x$.

The above presentation (as well as the rest of this chapter) is an extremely simplified exposition. The student may want to read more about these derivatives from other sources (see the references provided for this course).

1.2. Portfolio value, pay-off, and fair price. The pay-offs considered in the previous subsections are particular instances of "fair prices." The pay-off is the fair price at the expiry (or maturity) and usually there is no ambiguity (or difficulty) in determining it. In general, however, the "fair value" is a theoretical concept. We agree that if these assets are traded, *and there is no arbitrage in the economy*, then their fair value is their market value. The assumption that there is no arbitrage in the economy is not satisfied exactly in practice, but is a good first order approximation. In our abstract models, we will always assume that there is no arbitrage in the economy. To explain rigorously what we mean by the assumption that "there is no arbitrage in the economy," we need a few more definitions.

Let us consider several assets (stocks, futures contracts, option contracts, ...) whose "fair value" is typically denoted S_{1t}, S_{2t}, ... S_{Nt}. By S_{0t} we shall sometimes denote the value of a bank account paying interest rate r and initial deposit of 1 unit of currency. Saying that $a_{0t} < 0$ would therefore mean that at time t we have borrowed a_{0t} units of currency. We assume the interest rate to be the same for deposits and loans, which is an abstract assumptions not valid in practice! For this reason, we shall eventually change the meaning of S_{0t} the value at time t of a bond with initial price 1. Since we shall assume that the interest rate is constant r, if T is the maturity date of the bond, then this bond is the promisse to pay at time T the ammount $(1 + r)^T$ (in discrete time).

A *portfolio* $\mathcal{P}$ is simply a collection a_{kt}, $0 \leq k \leq N$, where a_k is the position taken in each of these assets (how much is owned or owed of that asset). We agree that a_{0t} denotes how much cash is in that portfolio, and this cash will always be invested in the bank account. Alternatively, we denote by a_{0t} how many bonds (with initial value 1) we own at time t. If $a_{0t} < 0$ it means that we have sold bonds. We stress that unless otherwise stated, we normally make **no** assumption on the a_k's. In particular, a_{kt} may take on *negative* values. The meaning of this is that we allow the **shorting** of our assets. Saying that $a_{0t} < 0$ means that we have borrowed money from the "bank," and hence we owe money.

The value, or *pay-off*, of a portfolio $(a_0, a_1, \dots, a_N)$ at time t is simply

eq.def.port_val

$$W_t = \sum_{k=0}^{N} a_{kt} S_{kt}. \tag{2}$$

2. The "no-arbitrage" principle

The pricing methods developed in this course will be based on the "no arbitrage principle" and thus will be called "no arbitrage pricing."

2.1. No arbitrage pricing. The form of "no arbitrage principle" that we shall use is the following.

The no-arbitrage principle: Assume the pay-offs W_{1t} and W_{2t} of two **self-financing** portfolios $\mathcal{P}_1$ and $\mathcal{P}_2$ are such that $W_{1t} \geq W_{2t}$ under *all circumstances.* Then $W_{1t} \geq W_{2t}$ at all times $t \leq T$ (including the present time).

"Under all circumstances" means "in all states of the economy," and refers to the fact that W_{kt} may depend on certain circumstances that are outside our control. The mathematical term is that S_{kt} and hence also W_{jt} are *stochastic variables*, meaning, in particular, that they are not *deterministic variables* (their values in the future is not determined solely by their current values).

A self-financing portfolio is one that, informally, does not require additional infusions of cash, except the initial investment. In continuous time t, the precise definition of *self-financing portfolios* is more complicated (and involves stochastic calculus). It will be given in discrete time. Right now, if is enough to know that if the weights (or positions) a_k are constant (i.e. independent of time) then the resulting portfolio is self-financing.

In practice, it is obvious that the positions a_{jt} depend *only* on the information available up to time t (excluding t). In a theoretical framework, this assumption will have to be made explicit. Self-financing portfolios (or trading strategies) with this property are called *admissible.* No arbitrage has to be formulated using admissible strategies.

Note that by the no arbitrage principle, nobody will sell you an option for nothing. So the pay-off of an option sometimes includes the initial payment f.

A consequence of the no-arbitrage principle is that

$$W_{1t} = W_{2t} \text{ under all circumstances } \Rightarrow W_{1t} = W_{2t}, \quad t < T.$$

In case of a single portfolio, this translates into

eq.no.arb2
$$W_T = 0 \text{ under all circumstances } \Rightarrow W_t = 0, \quad t < T. \tag{3}$$

The "no-arbitrage principle" could also be called the "no free lunch principle." In other words, there is no way you can start with a debt and, no matter what, end up in positive territory without actually doing something (i.e. providing some goods or services).

2.2. The value of futures contracts. Let us use now the no-arbitrage principle to price a futures contact whose underlying is a no-dividend liquid stock (or some other liquid financial instrument). To that end, we need to take into account the "time value of money." We shall then denote by r the *interest rate*, which we assume to be fixed and to be the same for borrowing and for lending (depositing money). Then a portfolio with B cash at time t will be worth $Be^{r(T-t)}$ at time $T > t$.

Assume now that we want to price a futures contract. Denote by F_t its price at time t. Let K be the strike price and T the maturity time (=delivery date). Let us assume that we start with a portfolio consisting of B cash (or units of bank account) and the futures F and -1 of the stock S. The initial value of the portfolio is thus

eq.initial.F
$$W_{t_0} = B + F_{t_0} - S_{t_0}. \tag{4}$$

The value of this portfolio at time T is then

$$W_T = Be^{r(T-t_0)} + F_T - S_T,$$

since it is reasonable to assume that we will keep the cash invested in the bank account.

We know that $F_T = S_T - K$ is the value of the futures contract at the delivery date. Thus

$$W_T = Be^{r(T-t_0)} - K.$$

Let us choose B so that $W(T) = 0$, that is, $B = e^{-r(T-t)}K$. Then $W_t = 0$ for all $t \le T$, by the no arbitrage principle. Substituting into the Equation (4) eq.initial.F, we obtain

eq.initial.F2
$$W_{t_0} = e^{-r(T-t_0)}K + F_{t_0} - S_{t_0} = 0, \tag{5}$$

and hence

eq.Futures
$$F_t = S_t - e^{-r(T-t)}K, \tag{6}$$

for all times $t \le T$. Equation (6) gives the no-arbitrage value (or "fair price") of a futures contract with the specified parameters (K is the delivery price, S is the spot price of the underlying, r is the interest rate, which is assumed to be constant).

Let us also mention that the above pricing formulas do not apply to certain commodity futures (sometimes due to seasonal circumstances, lack of liquidity, storage costs, or other reasons).

Let us notice that in the above example $B_T = B_{t_0} e^{r(T-t_0)}$, the value of the bank account at time T, is a *deterministic variable* (so it is not "stochastic"), since the value of the bank account is completely determined by the initial deposit B_{t_0} (assuming the interest rate r to be constant).

(Alternative method using a replicating portfolio for the Forward contract–to be included.)

2.3. The "forward price" of a futures contracts. This is an example. Let t_0 denote the "present time." The *forward price* $F(t, T)$ of a forward (or futures) contracts is the value of the strike price K that makes the value of the forward contract to be zero at time t, which may, or may not be the present time. Given that $F_t = S_t - e^{-r(T-t)}K$ by Equation (6), we obtain $0 = S_t - e^{-r(T-t)}F(t, T)$, so

eq.ForwardPrice

$$F(t, T) = e^{r(T-t)} S_t. \tag{7}$$

It is reasonable then, when entering into a forward contract, to take K to be the forward price of the forward contract, unless one has additional insight into the market (which would amount to speculation).

Futures contracts are traded for $K \neq F(t, T)$ at time t. How can you make money risk free, thus contradicting arbitrage?

2.4. Put-call parity. Let us consider the following two portfolios. The first portfolio consists at the initial time t_0 of one share of the stock S and $-Ke^{-r(T-t_0)}$ in a bank account. The second portfolio consists of a call option (with weight one) and a put option (with weight one). That is, the second portfolio is "long one call option and short one put option." The options also have strike K and maturity T.

The pay-off of the first portfolio at time t is $W_{1t} = S_t - Ke^{-r(T-t)}$. Thus $W_{1T} = S(T) - K$. The pay-off of the second portfolio at time T is $|S(T) - K|_+ - |S(T) - K|_- = S(T) - K$. Thus the two portfolios have the same pay-offs at time T, and hence they have the same value at any other previous time t. Let C_t denote the value of the given Call option at time t and P_t the value of the given Put option at time t. This gives

eq.PutCall

$$C_t - P_t = S_t - Ke^{-r(T-t)}, \tag{8}$$

a relation usually referred to as the *Put-Call parity*, which is valid only for European options.

The quantity

$$D(t, T) := e^{-r(T-t)} \leq 1$$

used to compare the value of money at time T with the value of money at time t is called *discount factor*. The discount factor is defined also in the case when r is not constant, but then it is a stochastic variable, since its value will depend on future circumstances outside our control. Many of the above equations remain valid in this more general setting, provided that one uses the correct discount factor.

Determining the "fair price" (or arbitrage free value) of an option is a much more complicated problem and solving it (and related problems) will be one of the main motivations for the theory developed in this course.

2.5. Theoretical assumptions. We have already made several assumptions that are not realistic (not satisfied in practice). One is that that the interest rates for lending and borrowing are the same. Another one is that there are no transaction costs. We also assume that the bid-ask spread is zero (the bid and ask price are the same). There will be many such simplifying assumptions in the theory developed below. Some of these assumptions will be explicit, but some will also be implicit. Removing some of these assumptions will make the theory more realistic, but also much more complicated. There is extensive research devoted to fully developing more realistic theories, and most of it falls outside the scope of this class.

2.6. Why do we want to know the "fair price"? Here are some reasons why we may want to know the fair price: to price something that is not traded (new product or low liquidity product), to estimate the price of an asset under extreme situations (risk management), to find mispricings in the market and to take advantage of them.

2.7. Exercises and examples. Let us consider a forward contract F with strike price K and delivery date T. We assume that the underlying is a liquid asset (say a stock) S. We denote by F_t the fair price of F at time t. By S_t we denote the market price of S at time t. Let us denote by r the interest rate, which is assumed to be constant.
1. Determine a portfolio $\mathcal{P}_1$ that consists of cash and the asset S that has the same pay-off W_{1T} (at time T) as the forward contract F.
Solution. The portfolio $\mathcal{P}_1$ that has the same payoff as the Forward contract consists at time t of the underlying S and $-ke^{-r(T-t)}$ cash. In components, $\mathcal{P}_1 = (-ke^{-r(T-t)}, 1)$.

2. For what value of K is $W_{1t} = 0$? Denote by $F(t, T)$ the value of K for which $W_{1t} = 0$.
Solution. We have $W_{1t} = -Ke^{-r(T-t)} + S_t$. If $W_{1t} = 0$ then $-Ke^{-r(T-t)} + S_t = 0$. Therefore $K = e^{r(T-t)} S_t$.

3. Assuming that $K > F(t, T)$ and futures contracts are sold at time t for the strike K and delivery date T. Assume the futures behave like the forwards (in particular, it costs nothing to enter into a futures contract). Show how you can make risk free money in this situation.
Solution. Assume $K > F(t, T) := e^{r(T-t)} S_t$, that is, one can sign Forward contracts for a value of K that is larger than what we believe is the fair price. (We believe that the fair price is $F(t, T)$, and we are in the process of showing this). The Forward contract with strike K can be entered into without any cost. Its payoff is $S_t - K < S_t - F(t, T)$. The replicating portfolio $\mathcal{P}_1$ constructed above will also cost nothing to enter into for strike $F(t, T)$ and will have payoff $S_t - F(t, T)$, which is more than what can be obtained on the market. So the Forward contract on the market has a smaller payoff. So the Forward contract on the market is overvalued. Hence we will sell (more precisely write) a Forward contract (which promises to sell one share of S for the price

K) and we will buy the replicating portfolio $\mathcal{P}_1$. That is, we borrow S_t to buy one share of S.

At time T we will owe $S_t e^{r(T-t)} = F(t,T)$, but we will own a share of S which we will sell for $K > F(t,T)$ (using our Forward contract). We end up with a risk free profit of $K > F(t,T)$. (This contradicts no arbitrage.)

4. Assume the numbers in the example discussed in class (Example 1.1.1 in the book, that is $S_0 = 4$, $u = 2$, $d = 1/2$, $r = 1/4$). Assume the Call (European option contract) is traded for 2 at initial time (time $t = 0$). Describe how you can make a risk free profit.
Solution. The European Call contract is priced more than its fair value. This means that it is relatively expensive.

Therefore, at initial time $t = 0$, sell the Call (that is, we write a Call contract with the specified parameters) and receive \$2, the price for which this Call is traded at that time. At the same time we buy the replicating portfolio (for "hedging purposes"). This means (see the calculations from class or from the book) that we buy 1/2 shares of stock for \$2. (In this case, we are left with no cash and we have no dept. In general, however, we would deposit the remaining cash, if any, in order to accrue interest. In case we had to borrow money, we would have to pay interest.)

At time $t = 1$ we proceed as follows. If we toss "Heads," the buyer of our Call Option will exercise it. We therefore buy an additional 1/2 shares of stock for \$4. Then we would own a full share of the stock and we would give it to the buyer of the option, receiving K=\$5 from the buyer. We are left in this way with \$1 profit (=\$5 - \$4).

On the other hand, if we toss "Tails," the buyer will not exercise. Then we sell our 1/2 share for \$1, again to be left with a profit of \$1.

5. Explain how $1 + r \geq u > d > 0$ leads to arbitrage opportunities.
Solution. Let us assume $1 + r \geq u > d$. At time $t = 0$ we then sell (more precisely "short") one share of the stock for S_0 and invest the proceeds. At time $t = 1$ we buy the stock to return it to the lender. The payoff of this trading strategy is as follows. In the case "Tails," we have to pay dS_0 for the stock. We thus make a sure profit since our cash deposit has accrued to $S_0(1+r) > dS_0$. Our profit is $S_0(1+r-d) > 0$. In case we toss "Heads," the payoff is $S_0(1 + r - u) \geq 0$. Thus we also make a profit if $1 + r > u$. In case $1 + r = u$, we break even in the case "Heads." This is a trading strategy that starts with $W_0 = 0$ and ends up with $W_1 \geq 0$, $W_{1t} > 0$. That is an example of (weak) arbitrage.

6. Try to find an estimate (inequality) for the prices C_t and P_t of a European Call and Put by comparing their pay-off with that of a Forward contract.

CHAPTER 3

Measure and Probability

Comentariu pentru participantii romani. Toti cei cu o pregatire matematica (universitate, ...) nu vor gasi mai nimic nou in aceasta sectiune. Incercati insa sa lucrati exemplele care se refera la martingale in cazul general a doua σ-algebre.

This is a very brief introduction to measure spaces and probability. To a large extent, we shall use only "finite probability spaces." Nevertheless, the modern theory of finance relies heavily on the language and results of Probability theory, so a quick foray into measure spaces and probability will be useful.

This will serve also as a reference dictionary that you can consult in the future.

Some of the definitions and frameworks are more general than the ones considered in the textbook.

1. Measure and probability spaces

Here is a long list of notations and definitions. They will be discussed in more detail in class.

1.1. Measure spaces.

By $\mathcal{P}(X)$ we shall denote the set of all subsets of the set X. (Example when $X = \{0, 1\}$.)

DEFINITION 3.1. Let X be a set. A σ-algebra in X is a subset $\mathcal{M} \subset \mathcal{P}(X)$ with the following three properties:

(1) $X \in \mathcal{M}$,
(2) If $S \in \mathcal{M}$, then $S^c := X \smallsetminus S \in \mathcal{M}$,
(3) If $S_1, S_2, \ldots, S_n, \ldots$ is a sequence of subsets of X and each S_j is in $\mathcal{M}$, then $\cup S_n \in \mathcal{M}$.

The pair $(X, \mathcal{M})$, where $\mathcal{M}$ is a σ-algebra in X is called a *measurable space*. Examples will be given shortly.

1.2. Positive measures.

DEFINITION 3.2. A *positive measure* μ on a measurable space $(X, \mathcal{M})$ is a function $\mu : \mathcal{M} \to [0, \infty]$ with the property that $\mu(\cup S_n) = \sum_n \mu(S_n)$ for any sequence of disjoint subsets $S_n \in \mathcal{M}$.

The property $\mu(\cup S_n) = \sum_n \mu(S_n)$ that a measure must satisfy is called *countable additivity*. The triple $(X, \mathcal{M}, \mu)$ is called a *measure space*.

In particular, we have

$$\mu(S \cup S') = \mu(S) + \mu(S') \tag{9}$$

for any two disjoint, $\mathcal{M}$–measurable sets. Similarly,

$$\mu(\cup_{k=1}^N S_k) = \mu(S_1 \cup S_2 \cup \ldots \cup S_N) = \sum_{k=1}^N \mu(S_k) = \mu(S_1) + \mu(S_2) + \ldots + \mu(S_N). \tag{10}$$

1.3. Measurable functions. Let $(X, \mathcal{M})$ be a measurable space. A function $f : X \to \mathbb{R}$ is called *measurable* if, for any real number a, the set $\{x, f(x) < a\}$ is in $\mathcal{M}$ (that is, it is measurable).

2. Examples

We now discuss the main examples we are interested in.

2.1. Finite measure spaces. Let $X = \{x_1, x_2, \ldots, x_N\}$ be a finite set. We take $\mathcal{M} = \mathcal{P}(X)$, so all sets are measurable. Similarly, all functions $f : X \to \mathbb{R}$ are measurable. A positive measure μ on X will be uniquely determined by $p_j = \mu(\{x_j\})$. Then for any other set $S = \{p_{i_1}, p_{i_2}, \ldots, p_{i_k}\}$, $i_1 < i_2 < \ldots < i_k$, we have

$$\mu(S) = \sum_{x \in S} \mu(\{x\}) = p_{i_1} + p_{i_2} + \ldots + p_{i_k}.$$

So a finite measure on X is completely determined by the weights $p_1, p_2, \ldots, p_N \in [0, \infty]$.

2.2. Finite probability spaces. A positive measure μ is called a *probability measure* if $\mu(X) = 1$. The triple $(X, \mathcal{M}, \mu)$ will then be called a *probability space*. From now on, all our measure spaces will be, in fact, probability spaces.

A finite probability space $(X, \mathcal{P}(X), \mu)$ is then completely determined by the weights $\mu(\{p\})$ for $p \in X$. That is, if $X = \{x_1, x_2, \ldots, x_N\}$ and $p_j = \mu(\{x_j\})$ as above, then the additional condition is

$$\sum_{j=1}^{N} p_j = p_1 + p_2 + \ldots + p_N = 1.$$

In particular, the range of our measure satisfies $\mu(S) \in [0, 1]$ for any $S \in \mathcal{M}$.

2.3. Finite partition probability spaces. A useful generalization of the finite probability example is the example of a finite partition probability space. This example can be skipped at first reading, since it will not be needed for the first chapter of the textbook.

Let us assume that we are given a set X that is partitioned into N disjoint sets $P_1, P_2, \ldots, P_N$:

$$P_1 \cup P_2 \cup \ldots \cup P_N = X.$$

Then to this partition we associate the σ-algebra

$$\mathcal{M} = \{S \subset X,\ S = P_{i_1} \cup P_{i_2} \cup \ldots \cup P_{i_N},\ \text{such that } i_1 < i_2 < \ldots < i_k\},$$

that is, $\mathcal{M}$ consists of all possible unions of some sets among the P_j's. A probability measure $\mu : \mathcal{M} \to [0, 1]$ will again be determined by $p_j = \mu(P_j)$ because if $S = P_{i_1} \cup P_{i_2} \cup \ldots \cup P_{i_N},$, with $i_1 < i_2 < \ldots < i_k$, then $\mu(S) = \mu(P_{i_1}) + \mu(P_{i_2}) + \ldots + \mu(P_{i_N}) = p_{i_1} + p_{i_2} + \ldots + p_{i_N}$.

A finite probability space is, in particular, a finite partition probability space by taking $P_j = \{x_j\}$. More examples will be discussed later.

A function $f : X \to \mathbb{R}$ will be $\mathcal{M}$–measurable if, and only if, it is constant on each of the sets P_j.

(In this lecture we also discussed the example of the set Ω_N.)

2.4. Countable partition probability spaces. This example is similar to the one above, except that X is a countable union of disjoint sets P_n, $n = 1, 2, \ldots$. Also, the condition that we have a probability measure is that $p_j \geq 0$ and $\sum p_j = \sum_{j=1}^{\infty} = 1$, this time the sum being a series. (Recall that the sum of a series with positive terms always makes sense, but it may be infinite.)

A finite partition probability space (and hence also a finite probability space) is a countable partition probability space.

A function $f : X \to \mathbb{R}$ is $\mathcal{M}$ measurable if, and only if, it is constant on all the sets P_j.

3. Expectation and integral

Let $(X, \mathcal{M}, \mu)$ be a probability space. A measurable function $f : X \to \mathbb{R}$ is also called a *random variable*. An important quantity associated to a random variable f is its expectation $\mathbb{E}(f)$. The expectation is defined for any probability space and random variable f under some mild conditions on f.

3.1. The countable case. Let X be a countable partition probability space. and let $f : X \to \mathbb{R}$ be a $\mathcal{M}$ measurable function. Then f is constant on each of the sets P_j. Let f_j be that constant (the value of f at any of the points of P_j). Then we define

$$\mathbb{E}(f) = \int_X f d\mu = \sum_{j=1}^{\infty} f_j p_j, \tag{11}$$

provided that the series is absolutely convergent ($\sum_j |f_j| p_j < \infty$).

For a finite probability space $X = \{x_1, x_2, \ldots, x_N\}$ we thus have

$$\mathbb{E}(f) = \sum_{j=1}^{N} f(x_j) p_j = f(x_1)p_1 + f(x_2)p_2 + \ldots + f(x_N)p_N. \tag{12}$$

In particular, we do not have to worry about the convergence of any series.

The following notation is also sometimes use for the expectation of f: $\mathbb{E}[f] = \mathbb{E}f = \int_X f d\mu = \mathbb{E}(f)$.

3.2. Properties of the expectation. The following properties of the expectation may be useful. For simplicity, we shall assume that we are in the case of a finite partition probability space. (In general, one has to assume that the random variables that appear in the formulas are "integrable.")

- $\mathbb{E}(af) = a\mathbb{E}(f)$ where f is a random variable (or measurable function) and $a \in \mathbb{R}$ is a constant. The random variable af is defined by the formula $(af)(x) = af(x)$.
- $\mathbb{E}(f) \geq 0$ if $f \geq 0$.
- If f is constant, $f(p_j) = a$, then $\mathbb{E}(f) = a$.
- If f and g are measurable, then $\mathbb{E}(f + g) = \mathbb{E}(f) + \mathbb{E}(g)$.
- More generally, if t_j are real constants and f_j are measurable functions, $j = 1, \ldots, N$, then

$$\mathbb{E}(\sum_{k=1}^{N} t_j f_j) = \sum_{k=1}^{N} t_j \mathbb{E}(f_j). \tag{13}$$

- If f is a random variable such that $|f(x)| \leq M$, then $|\mathbb{E}(f)| \leq M$.
- If f_n is a sequence of random variable such that $f_n(x) \to f(x)|$ is monotone increasing for any x, then $\mathbb{E}(f_n)| \to \mathbb{E}(f)$ (Monotone convergence theorem).
- If f_n is a sequence of random variable such that $f_n(x) \to f(x)|$ and there is a random variable $g \geq 0$, $\mathbb{E}(g) < \infty$, such that $|f_n(x)| \leq g(x)$, then $\mathbb{E}(f_n)| \to \mathbb{E}(f)$ (Bounded convergence theorem).

3.3. Variation. Let $f : X \to \mathbb{R}$ be a random variable (or measurable function). Let us denote $\overline{X} = \mathbb{E}(X)$. We then define the *variance* $Var(X)$ of X by the formula

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

Notice that, by the properties of the expectation, the quantity $Var(X)$ is always ≥ 0. The quantity $\sigma := Var(X)^{1/2}$ is usually called the *standard deviation* of X.

In statistics, the *sample standard deviation* is a multiple of our standard deviation for the finite probability space with equal weights. (To discuss.)

Let us assume that f and g are random variables such that $\mathbb{E}(f^i g^j) = \mathbb{E}(f^i)\mathbb{E}(g^j)$ (this happens, for instance, if f and g are independent random variables). Then $Var(f+g) = Var(f) + Var(g)$.

4. States of the economy

This will be discussed in full detail sometimes later, but you may want to read as a motivation for the above discussion. This model is useful in the **binomial asset pricing model**.

Let N be a horizon ($N = 1, 2, \dots, \infty$). We consider the set Ω_N of sequences $(\omega_1, \omega_2, \dots, \omega_N)$ with $\omega_j \in \{0, 1\}$. Thus, if $N = \infty$ we consider infinite sequences of 0 and 1. We think of ω_n as telling us if the price of a stock that we are interested goes up ($\omega_n = 1$) or down ($\omega_n = 0$) on the nth day. We assume that the price always moves up or down by just one tick.

If $k < N$, we have a natural map $\Omega_N \to \Omega_k$ given by truncation (we keep the first k terms of an N term sequence). For each k, we thus obtain a partition of Ω_N into as many disjoint sets as the elements of Ω_k. The elements of this partition are formed by sets P_j such that all the sequences $a \in P_j \subset \Omega_N$ begin with the same k terms. The set Ω_k has the meaning of how much information we have on the kth day (we know the past behavior of the stock, but not the future one). The resulting partition σ-algebra on Ω_N will be denoted $\mathcal{F}_k$ and will have the same meaning. In this way we obtain a sequence[1] of σ-algebras

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \tag{14}$$

This model applies to discrete time finance. For continuous time finance we will need to use σ-algebras $\mathcal{F}_t$, $0 \le t$, such that $\mathcal{F}_t \subset \mathcal{F}_s$ for $t < s$. This falls outside the scope of this course, however.

In the book, the spaces Ω is called the *coin toss space* and the set $\{0, 1\}$ is replaced with $\{H, T\}$ (Heads or Tails). For us, the space Ω will represent the space of possible states of the Economy.

At this time, start reading from the book, beginning with Chapter 1.

[1] this sequence may be finite if N is finite

5. Conditional expectations: Definition and some examples

5.1. Definition. Let $(\Omega, \mathbb{P})$ be a finite probability space (i.e Ω is a finite set and $\mathbb{P}$ is a probability measure for which every subset of Ω is measurable). Let us assume we are given also a partition $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$ of Ω and let $\mathcal{F}$ be the partition σ-algebra associated to $\mathcal{P}$ (i.e. $\mathcal{F}$ consists of all possible unions of the sets P_j.)

Let us denote by $\mathbb{R}^\Omega$ the set of all functions $f : \Omega \to \mathbb{R}$ and by $\mathbb{R}^\mathcal{P}$ the set of all $\mathcal{F}$-measurable functions $g : \Omega \to \mathbb{R}$. Since a function g is $\mathcal{F}$-measurable if, and only if, it is constant on each P_j, g is really a function $\mathcal{P} \to \mathbb{R}$, which justifies the notation $\mathbb{R}^\mathcal{P}$ for the set of $\mathcal{F}$ measurable functions. For simplicity, we shall assume that $\mathbb{P}(P_j) > 0$ for all j.

Definition 3.3. The *conditional expectation with respect to* $\mathbb{P}$ *given* $\mathcal{F}$ (or *conditioned by* $\mathcal{F}$) is the map $\mathbb{E}^\mathbb{P}_\mathcal{F} : \mathbb{R}^\Omega \to \mathbb{R}^\mathcal{P}$ defined by

$$\mathbb{E}^\mathbb{P}_\mathcal{F}[f](\omega) = \mathbb{E}^\mathbb{P}[f|\mathcal{F}](\omega) = \frac{1}{\mathbb{P}(P_j)} \sum_{\omega' \in P_j} \mathbb{P}(\omega') f(\omega'),$$

if $\omega \in P_j$.

In case $\mathbb{P}(P_j) = 0$, we define $\mathbb{E}^\mathbb{P}_\mathcal{F}[f](\omega)$ arbitrarily for $\omega \in P_j$, but agree to identify two functions that differ only on a subset of measure zero of Ω. For simplicity, we shall ignore this issue in what follows.

5.2. Product measure. Let (A, μ_A) and (B, μ_B) be two finite probability spaces. We shall denote by $\Omega = A \times B$ the set of pairs (a, b), where $a \in A$ and $b \in B$, as usual. We define $\mu = \mu_A \times \mu_B$, the product measure, by $\mathbb{P}(a, b) = \mu_A(a)\mu_B(b)$. For each $a \in A$, we define

$$P_a = \{a\} \times B \subset A \times B. \tag{15}$$

This defines a partition $\mathcal{P}$ of $\Omega = A \times B$

$$\mathcal{P} = \{P_a,\, a \in A\}. \tag{16}$$

Let $\mathcal{F}$ be the corresponding finite partition σ-algebra in Ω. We want to understand the conditional expectation given $\mathcal{F}$.

Let $\omega \in P_a$. That is, $\omega = (a, b)$. First, we need to compute $\mathbb{P}(P_a)$:

$$\mathbb{P}(P_a) = \sum_{b \in B} \mathbb{P}(a, b) = \sum_{b \in B} \mu_A(a)\mu_B(b) = \mu_A(a) \sum_{b \in B} \mu_B(b) = \mu_A(a).$$

We then have for $\omega = (a,b) \in \Omega$ given and $\omega' = (a,b') \in \Omega$, with b' arbitrary

$$\mathbb{E}^{\mathbb{P}}[f|\mathcal{F}](\omega) = \mathbb{E}^{\mathbb{P}}[f|\mathcal{F}](a,b) = \frac{1}{\mathbb{P}(P_a)} \sum_{\omega' \in P_a} \mathbb{P}(\omega') f(\omega')$$

$$= \frac{1}{\mu_A(a)} \sum_{b' \in B} \mu_A(a)\mu_B(b') f(a,b') = \sum_{b' \in B} \mu_B(b') f(a,b') = \int_B f(a,b') d\mu_B(b').$$

The last integral is, of course, the partial integral (or expectation) with respect to the last variable.

Let $\mathbb{P}_1$ be a measure on $\Omega_1 := \{H,T\}$. Then we can define inductively $\mathbb{P}_n = \mathbb{P}_{n-1} \times \mathbb{P}_1$, which is a measure on

$$\Omega_n := \{H,T\}^n = \Omega_{n-1} \times \Omega_1.$$

Then $\mathbb{P}_{n+m} = \mathbb{P}_n \times \mathbb{P}_m$ as measures on $\Omega_{n+m} = \Omega_n \times \Omega_m$ and we can apply to this product decomposition the results about conditional expectations for product spaces.

More specifically, we are interested in the decomposition

eq.prod
$$\Omega_N = \Omega_n \times \Omega_{N-n}. \tag{17}$$

Using the measure $\mathbb{P}_N$ and the product structure, we define the corresponding conditional expectation. In particular, using the notation of the above example, we have $\mathcal{F} = \mathcal{F}_n$, which was introduced earlier in the course (recall that the $\mathcal{F}_n$ measurable functions are the functions $f : \Omega_N \to \mathbb{R}$ that depend only on the first n variables). The conditional expectation

$$\mathbb{E}_n[f] := \mathbb{E}^{\mathbb{P}}[f|\mathcal{F}_n] \tag{18}$$

is simply the integration with respect to the second variable in the product decomposition (17). You should compare now this approach with the alternative one considered in the book.

Then we can say that the measure $\mathbb{P} = \mathbb{P}_N$ is the risk neutral measure if, and only if, $S_n = \mathbb{E}_n(\frac{S_{n+1}}{1+r})$. If $\mathbb{P}$ is the risk neutral mesure, then the fair prices of a security V (or contingent claim) will satisfy $V_n = \mathbb{E}_n(\frac{V_{n+1}}{1+r})$. Again, one should compare our approach with that in the book.

Let $\mathbb{P}_1$ be the risk neutral measure for a period one economy with the same parameters u, d, and r (that is, $\mathbb{P}(H) = (1+r-d)/(u-d)$ and $\mathbb{P}(T) = (u-1-r)/(u-d)$). The assumptions that $S_{n+1}(\omega H) = uS_n(\omega)$

and $S_{n+1}(\omega T) = dS_n(\omega)$ then gurantee that that P_N is the unique risk neutral measure (here we use also $d < 1 + r < u$).

We will come back to these issues later on in the framework of Martingales. In particular, we will define more general conditional expectations $\mathbb{E}_n$.

6. Properties of conditional expectations.

We now list some of the main properties of the conditional expectations. In the following, $(\Omega, \mathbb{P})$ will be a finite probability space. Also, let $\mathcal{P} = \{P_1, P_2, \dots, P_k\}$ be a finite partition of Ω and $\mathcal{F}$ be the associated σ-algebra. We shall denote by $\mathbb{R}^\Omega$ the set of functions $f : \Omega \to \mathbb{R}$ and by $\mathbb{R}^\mathcal{P}$ the subset of $\mathbb{R}^\Omega$ consisting of $\mathcal{F}$-measurable functions. Recall that a function $g : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable if, and only if, g is constant on all sets P_j of our partition $\mathcal{P}$. We shall denote, in particular, by $g(P_j)$ the value of such a g to an arbitrary element of P_j.

We denote as before by $\mathbb{E}^\mathbb{P}_\mathcal{F} : \mathbb{R}^\Omega \to \mathbb{R}^\mathcal{P}$ the conditional expectation map. Recall that we also denote by $\mathbb{E}^\mathbb{P}[f|\mathcal{F}] = \mathbb{E}^\mathbb{P}_\mathcal{F}(f)$. We now list the properties of $\mathbb{E}^\mathbb{P}_\mathcal{F}$.

1. For $a, b \in \mathbb{R}$ and $f, g \in \mathbb{R}^\Omega$, let us denote by $h = af + bg$ the function $h(\omega) = af(\omega) + bg(\omega)$. Then we have that the conditional expectation $\mathbb{E}^\mathbb{P}_\mathcal{F}$ is linear:

$$\mathbb{E}^\mathbb{P}_\mathcal{F}(af + bg) = a\mathbb{E}^\mathbb{P}_\mathcal{F}(f) + b\mathbb{E}^\mathbb{P}_\mathcal{F}(g). \tag{19}$$

2. $\mathbb{E}^\mathbb{P}_\mathcal{F}(1) = 1$, the constant function equal to 1.

3. If $\mathcal{F} = \{\emptyset, \Omega\}$, that is, if $\mathcal{P}$ consists of the set Ω alone, then $\mathbb{E}^\mathbb{P}_\mathcal{F}(f) = \mathbb{E}^\mathbb{P}(f)$, that is, the constant function equal to the expectation of f.

4. If $\mathcal{F} = \mathcal{P}(\Omega)$, that is, if $\mathcal{P}$ consists of all the single element subsets of Ω, then $\mathbb{E}^\mathbb{P}_\mathcal{F}(f) = f$.

5. *The projection property:* Let $h = \mathbb{E}^\mathbb{P}_\mathcal{F}(f)$. Then $\mathbb{E}^\mathbb{P}(fg) = \mathbb{E}^\mathbb{P}(hg)$ for any $g \in \mathbb{R}^\mathcal{P}$ (that is, for any $\mathcal{F}$-measurable function g).

6. Conversely, let h be an $\mathcal{F}$-measurable function such that $\mathbb{E}^\mathbb{P}(fg) = \mathbb{E}^\mathbb{P}(hg)$ for any $g \in \mathbb{R}^\mathcal{P}$, then $h = \mathbb{E}^\mathbb{P}_\mathcal{F}(f)$.

7. *Taking out what is known:* Assume g is $\mathcal{F}$-measurable, then $\mathbb{E}^\mathbb{P}_\mathcal{F}(gf) = g\mathbb{E}^\mathbb{P}_\mathcal{F}(f)$.

8. *Iterated projection property:* Let $\mathcal{F}' \subset \mathcal{F}$ be a smaller σ-algebra. Then $\mathbb{E}^\mathbb{P}_{\mathcal{F}'}(\mathbb{E}^\mathbb{P}_\mathcal{F}(f)) = \mathbb{E}^\mathbb{P}_{\mathcal{F}'}(f)$.

Let us comment on these properties. Properties 2, 3, 4, and 5 follow from the definition using a direct calculation. Let us denote by $(u, v) = \mathbb{E}^{\mathbb{P}}(uv)$. Then (u, v) is an inner product on $\mathbb{R}^{\Omega}$ (assuming that $\mathbb{P}(\omega) > 0$ for any $\omega \in \Omega$). The subspace of $\mathcal{F}$-measurable functions forms a linear subspace $\mathbb{R}^{\mathcal{P}}$ of $\mathbb{R}^{\Omega}$. Property 5 means that $\mathbb{E}^{\mathbb{P}}_{\mathcal{F}}$ is the orthogonal projection (with respect to this inner product) of $\mathbb{R}^{\Omega}$ onto $\mathbb{R}^{\mathcal{P}}$. This orthogonal projection is uniquely determined and is a linear map (prove this as an exercise). Property 1 is then a consequence of the linearity of the orthogonal projection. Properties 6, 7, and 8 are consequences of the uniqueness of the orthogonal projection.

Property 8 is in fact a general property of orthogonal projections. Indeed, let $W' \subset W \subset V$ be linear subspaces of an inner product space V. Let us denote by P_W and $P_{W'}$ the orthogonal projections onto these subspaces. Then $P_{W'}P_W = P_{W'}$ (meaning $P_{W'}(P_W u) = P_{W'}u$.

We will not use the following discussion, but it will help make Property 8 clearer. Since Ω is a finite set, any σ-algebra $\mathcal{F}' \subset \mathcal{P}(\Omega)$ is generated by a partition $\mathcal{P}' = \{P'_1, P'_2, \ldots, P'_q\}$. The assumption that $\mathcal{F}' \subset \mathcal{F}$ implies that each of the sets $P'_j \subset \mathcal{F}$, and hence each of the sets P'_j is a union of the sets P_i defining the partition $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$ that, in turn, defines $\mathcal{F}$.

Here are two exercices.

1. Show that $\mathbb{E}^{\mathbb{P}}_{\mathcal{F}}(f) \geq 0$ if $f \geq 0$. (More precisely, $\mathbb{E}^{\mathbb{P}}_{\mathcal{F}}(f)(\omega) \geq 0$ for any $\omega \in \Omega$.)

2. Let g be a convex function (if g is twice differentiable, this is equivalent to saying that $g'' \geq 0$). Then

$$g(\mathbb{E}^{\mathbb{P}}_{\mathcal{F}}(f)) \leq \mathbb{E}^{\mathbb{P}}_{\mathcal{F}}(g \circ f). \tag{20}$$

Examples of g are $g(x) = e^x$, $g(x) = -\ln x$, and $g(x) = x^p$, $p > 1$. (We assume $x > 0$ in the last two examples.) This then gives

$$e^{\mathbb{E}^{\mathbb{P}}_{\mathcal{F}}(f)} \leq \mathbb{E}^{\mathbb{P}}_{\mathcal{F}}(e^f). \tag{21}$$

This is *Jensen's inequality*. Prove it and write the resulting Jensen's inequality for the other two functions.