MATH 4210 FINANCIAL MATHEMATICS
SUPPLEMENTARY NOTES

BIN WANG

ABSTRACT. This is a set of supplementary notes for the course MATH 4210 Financial Mathematics taught by Professor Xiaolu Tan at the Chinese University of Hong Kong.

CONTENTS

Disclaimer: These notes are heavily based on the lectures given by Professor Xiaolu Tan for the course MATH 4210 Financial Mathematics at the Chinese University of Hong Kong, together with other sources that are listed in the References. There is no claim to any originality in the notes, but we hope — for the intended audience (mostly the undergraduate students taking this course) — they will provide a useful supplement. If the reader has found any errors, either typographical or conceptual, they are hugely encouraged to report the errors by sending an email to bwang@math.cuhk.edu.hk.

Acknowledgement

The author would like to thank Jiazhi Kang, Xiangying Pang and our supervisor Professor Xiaolu Tan, for their comments and proofreading.

1. Lecture 1 Sept 5th: Arbitrage and Probability Space

Definition 1.1. An **arbitrage** （套利行为） is a trading strategy such that

(1) it begins with zero money,
(2) has zero probability of losing money,
(3) and has a positive probability of earning money.

As you can see from this definition directly, a mathematical model of financial markets that admits an arbitrage cannot be used for analysis: **wealth can be generated from nothing and it does not run the risk of loss.**

In reality, markets sometimes exhibit arbitrage, but this is necessarily fleeting: once such an opportunity occurs, many investors would use it to make profits

Date: Sept 12th, 2022.

Email: bwang@math.cuhk.edu.hk.

without risks and hence remove the arbitrage. Therefore, from now on, in all of our subsequent analysis for financial markets, we will always assume that

there is no arbitrage.

Let us illustrate the idea through the following example.

Example 1.2. Consider three merchants who are willing to buy and sell bags containing apples and oranges. Their prices for buying and selling are as follows.

Merchants	Bag Content	Price
I	3 apples, 2 oranges	\$5
II	2 apples, 3 oranges	\$6
III	5 apples, 5 oranges	\$10

Now suppose a bright math student Wang walks into the market with no money and he sees his good friend Lan is wandering around, so he goes up, **borrows** \$10 from Lan, and promise to pay back in 5 minutes. Next, Wang buys a bag from Merchant III for \$10 and then sells the fruits to Merchants I and II; thus gaining \$11. Finally, he pays the money back to Lan and has a profit of \$1. Remember that Wang began with zero money and the trading takes no risk of loss: the worst he could do is to return the money to Lan and have zero profit, assuming that he did not charge Wang for interest rates.

Remark 1.3. This simple example is way too ideal that it almost cannot be applied in real life. Naturally, you would ask, what if Lan did not lend the money? what if the Merchants I and II were not willing to accept your fruits? These issues certainly exist; the example is only to help you understand what would happen if there is an arbitrage.

Since this is a course for modelling financial markets mathematically, let us briefly review the basic math tool that will be used throughout the rest of the course: probability theory. We will assume the reader is familiar with elementary probability theory at the level of the course MATH 3280 Introductory Probability at CUHK, or at the level of the textbook by Ross [4]; anything that is measure-theoretic will be kept to a minimum and the reader will be pointed to references for further study.

Definition 1.4. The set Ω of all possible outcomes of a particular experiment is called the **sample space** （样本空间） for the experiment. An **event** （事件） is a collection of possible outcomes of an experiment i.e. a subset of Ω.

Remark 1.5. To avoid mathematical triviality, we will always assume Ω is non-empty i.e. an experiment always has an outcome.

Example 1.6. For the experiment of tossing two coins, the sample space is just $\Omega = \{HH, HT, TH, TT\}$, where H represents "head" and T represents "tail". The event "the first coin is a head" is just the subset $\{HH, HT\}$.

As this is a 4000-level course, you probably have taken a course in elementary probability theory, so you know we have the need to do operations on events e.g. taking unions, intersections and complements. This is to help us define a probability measure and work with it consistently.

Definition 1.7. Consider an experiment and let Ω be its sample space. A **σ-algebra** on Ω is a collection of subsets of Ω such that

(1) $\emptyset \in \mathcal{F}$.
(2) If $A \in \mathcal{F}$, then $A^c := \Omega \setminus A \in \mathcal{F}$.
(3) If a sequence of subsets $A_1, A_2, \ldots$ is such that $A_i \in \mathcal{F}$ for all $i \in \mathbb{N}$, then their union $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Remark 1.8. If the definition is too abstract for you, then you may think of the elements of a σ-algebra as the events of the associated experiment.

Example 1.9. Consider our coin-tossing experiment in the previous example. The set

$$\mathcal{F}_1 = \{\emptyset, \{HH, HT\}\}$$

is **not** a σ-algebra because the complement of the empty set $\emptyset$ which is just Ω itself, is not in $\mathcal{F}_1$.

Exercise 1.10. Write down the σ-algebra for our coin-tossing experiment. If you are not sure about how to do this, you can go to the tutorial and ask the TA Jiazhi Kang.

Definition 1.11. A **probability measure** （**概率测度**） on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

(1) $\mathbb{P}(\Omega) = 1$ and
(2) Whenever $A_1, A_2, \ldots$ is a sequence of disjoint events i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$, we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

The number $\mathbb{P}(A)$ is called the probability of the event $A \in \mathcal{F}$ and the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space** （**概率空间**）.

Example 1.12. The simplest example is that of tossing just one coin, in which the sample space is $\Omega = \{H, T\}$ and the σ-algebra is $\mathcal{F} = \{\emptyset, \Omega, A_1, A_2\}$, where

$$A_1 = \{H\} \equiv \text{“ the outcome is a head ”}$$
$$A_2 = \{T\} \equiv \text{“ the outcome is a tail ”}$$

We may **assign** the probability measure as

$$\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(\Omega) = 1,\ \mathbb{P}(A_1) = \frac{1}{2},\ \mathbb{P}(A_2) = \frac{1}{2}$$

.

Why do we use the word “ assign ”? Isn't this the most obvious thing in the world? Let us now point out the possibility of equipping $(\Omega, \mathcal{F})$ with a different probability measure.

Suppose we are not tossing the coin, but instead we are looking at the price of a stock, its value will either go up or down on the next day, which we can represent as a head (if it goes up) and a tail (if it goes down). Thus we have the same sample space and σ-algebra. However, this time, the stocks are of Lan's company and Lan has told Wang some private information about his upcoming products, Wang realizes that the price will definitely go up the next day. Hence, in this case, we have

$$\mathbb{P}(A_1) = 1,\ \mathbb{P}(A_2) = 0$$

This is how we are going to model financial markets using probability theory in the rest of this course. Also, we could illustrate this idea by tossing a coin whose mass is not uniformly distributed so that the probability of obtaining a head might be greater than $\frac{1}{2}$.

Remark 1.13. It is due to our request of real-life applications that we have to make a well-defined probability measure, which gives us the probability of an event; in order to do so, we have to first create a space in which we can validly operate on events of our experiment. This is why we defined $\mathcal{F}$ and $\mathbb{P}$ in the way they are.

For those readers who are interested in the theoretical foundations, they may want to take a course in measure theory and/or a course in probability theory e.g. MATH 5011 and/or STAT 5005 at CUHK. The standard textbooks in the fields are [5] and [2].

Remark 1.14. The coin-tossing experiment is probably the simplest and the deepest model in mathematics. It can not only model the trend of financial markets, but also the spin of an electron. That is, by understanding how to toss a coin, you can both unravel the mysteries of our universe and enter the financial industry to make a lot of money. How cool is that!

2. Lecture 2 Sept 7th & Lecture 3 Sept 14th: Random Variable and Conditional Expectation

Definition 2.1. A **random variable** （**随机变量**） is a function $X : \Omega \to \mathbb{R}$ with the property that

$$\{X \leq c\} := \{\omega \in \Omega : X(\omega) \leq c\} \in \mathcal{F}$$

for all $c \in \mathbb{R}$.

Remark 2.2. We need this property so that the probability $\mathbb{P}(X \leq c)$ is well-defined. Moreover, the usual definition of a random variable requires that $\{X \in B\} \in \mathcal{F}$ for all Borel sets B in $\mathbb{R}$. In fact, these two definitions are equivalent; we use the above one just because it is easier to interpret its physical meaning. In other words, you can see why we would need to consider the event that the value of X is smaller than c, but it would be a lot harder to understand what it means to have the value of X falling within an arbitrary (Borel) set $B \subseteq \mathbb{R}$. As for the question of what is a Borel set, please see [5, p. 12].

Definition 2.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose X is a random variable.

- A sub σ-algebra of $\mathcal{F}$ is a subcollection $\mathcal{G}$ of subsets of Ω (i.e. $\mathcal{G} \subseteq \mathcal{F}$) such that $\mathcal{G}$ itself is a σ-algebra.
- The σ-algebra generated by X, denoted by $\sigma(X)$, is the smallest σ-algebra containing all sets of the form $\{X \leq c\}$, where $c \in \mathbb{R}$.
- Let $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ be two sub σ-algebras. We say $\mathcal{G}$ and $\mathcal{H}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ for all $A \in \mathcal{G}$ and all $B \in \mathcal{H}$.
- We say two random variables X, Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.
- We say X and $\mathcal{G}$ are independent if $\sigma(X)$ and $\mathcal{G}$ are independent.

Definition 2.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a random variable with $\mathbb{E}[|X|] < \infty$. Suppose $\mathcal{G} \subseteq \mathcal{F}$ is a sub σ-algebra of $\mathcal{F}$. The **conditional**

expectation （条件期望） of X given $\mathcal{G}$ is **any** random variable Z with $\mathbb{E}[|Z|] < \infty$ that satisfies

(1) Z is $\mathcal{G}$-measurable, and
(2) for all $\mathcal{G}$-measurable bounded random variable Y, we have

$$\mathbb{E}[XY] = \mathbb{E}[ZY] \tag{2.5}$$

Remark 2.6. Property (2) may be replaced with

$$\int_A Z(\omega)\ d\mathbb{P}(\omega) = \int_A X(\omega)\ d\mathbb{P}(\omega) \quad \forall\ A \in \mathcal{G}. \tag{2.7}$$

The thing is, if we use this definition, we will have to define the integral $\int_A X(\omega)\, d\mathbb{P}(\omega)$ first, which requires measure theory; we refer the ambitious reader to [2] for details.

Note that in this definition, we said " any random variable" that satisfies the above two properties is a qualified conditional expectation. Let us now briefly show this is a well-defined concept using (2.7).

Suppose Y and Z are two random variables satisfying (1) and (2). By (1), the random variable $Y - Z$ is $\mathcal{G}$-measurable and so the set $A := \{\omega \in \Omega : (Y - Z)(\omega) > 0\}$ is in $\mathcal{G}$. Now, apply (2) we obtain

$$\int_A Y(\omega)\ d\mathbb{P}(\omega) = \int_A X(\omega)\ d\mathbb{P}(\omega) = \int_A Z(\omega)\ d\mathbb{P}(\omega).$$

Equivalently, we have

$$\int_A [Y(\omega) - Z(\omega)]\ d\mathbb{P}(\omega) = 0.$$

However, the integrand is strictly positive on the set A, so the only way for the integral to be zero is that A has measure zero i.e. $\mathbb{P}(A) = 0$. Hence, we have $Y \leq Z$ everywhere except on a set of measure zero. Repeating the same argument for $B := \{\omega \in \Omega : (Z - Y)(\omega) > 0\}$, we obtain that $Y = Z$ everywhere except on a set of measure zero. Therefore, somehow we have proved that a conditional expectation is unique. For the reason why we treat two random variables that are equal almost everywhere the same, please see a textbook in measure theory.

Definition 2.8. If an event $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, then we say A occurs **almost surely**.

Now, having proved uniqueness, the next natural question is that does there exist a conditional expectation for each random variable with finite expectation? The answer is a yes, please see the **Radon-Nikodym theorem** (again) in measure theory [2, section 4.1]; as the proof is outside the scope of this course.

Notation 2.9. Once we have found **one** qualified conditional expectation for X given $\mathcal{G}$, we will denote it as $\mathbb{E}[X|\mathcal{G}]$.

Remark 2.10. The ordinary expectation $\mathbb{E}[X]$ of a random variable is a number, but the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ is a randon variable. Please note the huge difference.

You may view $\mathbb{E}[X|\mathcal{G}]$ as an estimator for X based on information provided by $\mathcal{G}$. Property (1) means that the value of the estimator $\mathbb{E}[X|\mathcal{G}]$ can be obtained from $\mathcal{G}$; while property (2.7) ensures that $\mathbb{E}[X|\mathcal{G}]$ is a reasonably good estimator for X, as they result in the same average value over events in $\mathcal{G}$.

Notation 2.11. If $\mathcal{G} = \sigma(Y)$ is the σ-algebra generated by some other random variable Y, then we write $\mathbb{E}[X|Y]$ instead of $\mathbb{E}[X|\sigma(Y)]$.

Proposition 2.12. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X, Y be random variables with $\mathbb{E}[|X|], \mathbb{E}[|Y|] < \infty$. Suppose $\mathcal{G} \subseteq \mathcal{F}$ is a sub σ-algebra of $\mathcal{F}$. We have

(1) (Tower Property) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.
(2) If X is $\mathcal{G}$-measurable, then $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$
(3) If X is independent of Y, then $\mathbb{E}[X|Y] = \mathbb{E}[X]$.
(4) (Linearity) $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$.

Remark 2.13. We will first explain intuitions of why these statements are true and then for completeness, we shall present a brief mathematical verification for each statement. For the mathematical proofs, the reader is **not required** to understand them in full details because we are assuming they have not taken a course in measure theory. At this stage, they are only required to convince themselves that the statements are naturally true.

Proof.

(1) This equality is just saying that $\mathbb{E}[X|\mathcal{G}]$ is an unbiased estimator for X and it can be proved by taking $A = \Omega$ in (2.7). Note that $\mathbb{E}[X]$ is a number so it is $\mathcal{G}$-measurable.
(2) Intuitively, if the information about X is already contained in $\mathcal{G}$, then we only need information about Y to estimate XY. Thus the estimator for XY reduces to an estimator for Y. Now we present a brief measure theoretic proof. Note again that you are not required to fully understand it.

Since X is $\mathcal{G}$-measurable and $\mathbb{E}[Y|\mathcal{G}]$ is also $\mathcal{G}$-measurable (by the definition of a conditional expectation!), the product $X\mathbb{E}[Y|\mathcal{G}]$ is $\mathcal{G}$-measurable. It remains to verify (2.7). We claim that it suffices to do so for the special case when $X = \mathbf{1}_B$ is the indicator function, where $B \in \mathcal{G}$. The reasoning goes as follows. First, any random variable X can be written as the sum of two non-negative random variables $X = X^+ - X^-$, where

$$X^+ := \max\{X, 0\}, \quad X^- := \max\{-X, 0\}.$$

By measure theory, each non-negative random variable can be approximated by a sequence of non-negative simple functions i.e. there exists a sequence $X_i = \sum_{j=1}^m a_{ij}\mathbf{1}_{B_{ij}}$ such that $\lim_{i\to\infty} X_i(\omega) = X(\omega)$ for every $\omega \in \Omega$. By linearity, if the statement is true for $X = \mathbf{1}_B$, then it can be passed to general X. Indeed, we have

$$\begin{aligned}\int_A X\mathbb{E}[Y|\mathcal{G}]\ d\mathbb{P}(\omega) = \int_A \mathbf{1}_B\mathbb{E}[Y|\mathcal{G}]\ d\mathbb{P}(\omega) &= \int_{A\cap B} \mathbb{E}[Y|\mathcal{G}]\ d\mathbb{P}(\omega) \\ &= \int_{A\cap B} Y(\omega)\ d\mathbb{P}(\omega) \quad \text{by definition of } \mathbb{E}[Y|\mathcal{G}] \\ &= \int_A \mathbf{1}_B Y(\omega)\ d\mathbb{P}(\omega) \\ &= \int_A X(\omega)Y(\omega)\ d\mathbb{P}(\omega)\end{aligned}$$

(3) If X is independent of Y, then the information generated by Y is not going to help us determine X. Thus the best estimate we could get is its expectation value.

For a proof, as commented in the proof of (1), $\mathbb{E}[X]$ is $\sigma(Y)$-measurable; and as we have done for (2), we verify (2.7) only for $X = \mathbf{1}_B$, where B is independent of $\sigma(Y)$. Indeed,

$$\begin{aligned}
\int_A X(\omega)\, d\mathbb{P}(\omega) = \int_A \mathbf{1}_B\, d\mathbb{P}(\omega) = \int_{A\cap B} d\mathbb{P}(\omega) \\
&= \mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B) \quad \text{by independence} \\
&= \mathbb{P}(A)\cdot\mathbb{E}[\mathbf{1}_B] \\
&= \int_A \mathbb{E}[\mathbf{1}_B]\, d\mathbb{P}(\omega) \\
&= \int_A \mathbb{E}[X]\, d\mathbb{P}(\omega)
\end{aligned}$$

(4) Trivial: because linearity preserves measurability and the rest follows from linearity of integration.

□

Digression 2.14. There is a geometric interpretation for the conditional expectation, however, in order to make this notation precise, we would need a result from functional analysis: Let H be a Hilbert space and let F be a closed subspace. Then for all $x \in H$ there exists a unique $z \in F$ such that $\langle x - z, y\rangle = 0$ for all $y \in F$. This y is called the **orthogonal projection** of x onto F.

Now, we will accept the following facts without proofs:

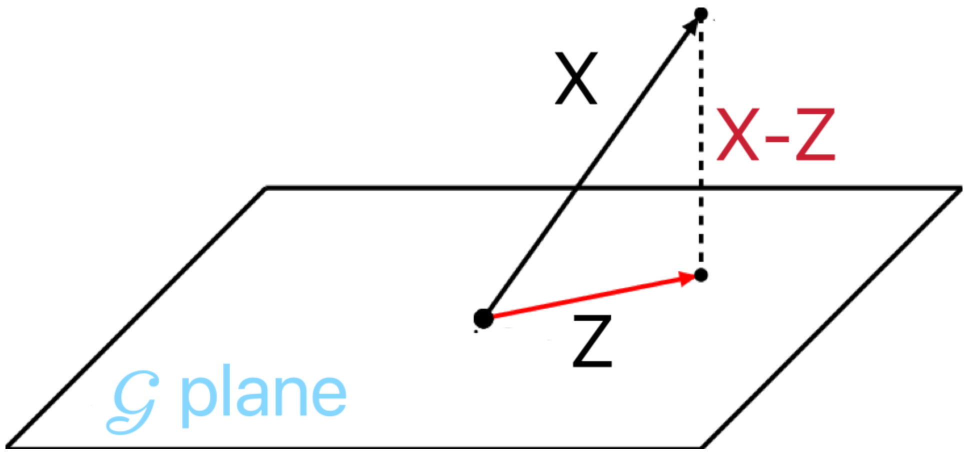
(1) $H(\mathcal{F}) := \{X : X : \mathcal{F} \to \mathbb{R}$ is a random variable with $\mathbb{E}[X^2] < \infty\}$ is a Hilbert space with the inner product $\langle X, Y\rangle := \mathbb{E}[XY]$.
(2) Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub σ-algebra. Then $H(\mathcal{G})$ is a closed subspace of $H(\mathcal{F})$.

Therefore, for every random variable $X \in H(\mathcal{F})$, there exists a unique random variable $Z \in H(\mathcal{G})$ such that $\langle X - Z, Y\rangle = 0$ for all $Y \in H(\mathcal{G})$. By linearity of the inner product, we have $\langle X, Y\rangle = \langle Z, Y\rangle$ for all $Y \in H(\mathcal{G})$. Moreover, by the definition of this inner product, we have $\mathbb{E}[XY] = \mathbb{E}[ZY]$ for all $Y \in H(\mathcal{G})$.

You can not only see the origin of the defining property (2.5) and also you can see how the existence and uniqueness of Z have been elegantly entailed by using the language of functional analysis.

As you may have been taught in a linear algebra class, inner product equating zero can have a rough sense of orthogonality (not rough in the Euclidean space). Thus, we may interpret the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ as the random variable such that its difference with X is perpendicular to all random variable Y lying on the information plane of $\mathcal{G}$ i.e. $\langle X - Z, Y\rangle = 0$.

Consider the following diagram:



We can view X as an arbitrary vector in space and we would like to estimate its length. As you can see from this diagram, the projection of X onto the $\mathcal{G}$-plane, the vector Z, its length though is smaller than the true length of X, it provides a reasonably good estimate for it. Moreover, the vector $X - Z$ is perpendicular to all vectors lying on the $\mathcal{G}$-plane; its existence and uniqueness can also be deduced from propositions in classical Euclidean geometry which you all have learned in middle school[1].

This is exactly like inferring about your height based on your shadow projected by a street lamp on the ground and you should now see why we have said $\mathbb{E}[X|\mathcal{G}]$ is the random variable that, in a sense, best approximates our X.

References

[1] R. H. Chan, Y. Z. Guo, S. T. Lee and X. Li. Financial mathematics, derivatives and structured products. Springer, Singapore, 2019.

[2] R. Durrett, Probability: Theory and Examples. Fifth edition. Cambridge University Press, Cambridge, 2019. xii+419 pp.

[3] V. Paulsen, Advanced Financial Mathematics, online lecture notes available at `https://www.uni-muenster.de/imperia/md/content/Stochastik/financial_mathematics.pdf`.

[4] S. Ross, A First Course in Probability, 8th edition. Pearson Education Inc. 2010.

[5] W. Rudin, Real and complex analysis. Third edition. McGraw-Hill Book Co., New York, 1987. xiv+416 pp.

[1] So far we have provided three interpretations about the conditional expectation: measure-theoretic, functional analytic and Euclidean geometric. The reader may pick his/her favourite one.