Mathematical Foundations for Finance

Exercise 1

Martin Stefanik
ETH Zurich

Which Exercise Class to Visit?

We would like to distribute students more or less evenly to the available exercise classes. Therefore

- Surnames starting with **A–G** $\longrightarrow$ Friday 8:00-10:00 HG D 7.1;
- Surnames starting with **H–O** $\longrightarrow$ Friday 8:00-10:00 LFW E 13;
- Surnames starting with **P–Z** $\longrightarrow$ Friday 10:00-12:00 LFW E 13.

Try to visit the exercise class to which you are assigned, if possible.

Organizational Notes I

- The exam will cover *all* material discussed during the *lectures* and during the *exercise classes*. Details will be reviewed towards the end of the semester.
- Old exams are available here, but you are highly discouraged from preparing from the old exams only.
- Presence on lectures and exercise classes is not obligatory but is highly recommended.
- Each class will have an exercise sheet, which will be uploaded to the course homepage on Tuesday *before* the corresponding class.

Organizational Notes II

- Your solutions need to be submitted to your assistant's box in front of HG G 53.2 by Tuesday 18:00 (in the week *after* the corresponding exercise).
- Handing in your solutions is not obligatory, but being able to solve the exercises independently goes a long way towards a good exam performance.
- The model solutions to the exercise sheets will be uploaded to the course homepage on Tuesdays as well (after your submission deadline).
- Regular question times (also called "Präsenz") will be held on *Mondays and Thursdays, 12:00-13:00* in HG G 32.6.

Learning Resources I

The lecture will closely follow the lecture notes that can be purchased before the beginning of the next lecture on September 24. We will also be selling these lecture notes during Präsenz hours. Other optional and additional sources are

- *Stochastic Finance: An Introduction in Discrete Time*, H. Föllmer, A. Schied, de Gruyter, 2011,
- *Introduction to Stochastic Calculus Applied to Finance*, D. Lamberton, B. Lapeyre, Chapman-Hall, 2008.

Learning Resources II

Especially for those who do not have the necessary background, it is also recommended to consult

- *Probability Essentials*, J. Jacod and P. Protter, Springer, 2003.

Another possibility is to also purchase the English version of the script used for the ETH course on Probability Theory by prof. Sznitman. This script will also be sold during Präsenz hours.

Another good resource is Mathematics Stack Exchange.

Sigma Algebra

Definition 1 (σ-algebra)

Let $\Omega \neq \emptyset$ be a set and let 2^Ω denote the power set (the set of all subsets) of Ω. $\mathcal{F} \subset 2^\Omega$ is called a *σ-algebra* if it satisfies the following:

1. $\Omega \in \mathcal{F}$,
2. $A \in \mathcal{F} \implies A^c = \Omega \setminus A \in \mathcal{F}$,
3. $A_n \in \mathcal{F}, n \in \mathbb{N} \implies \bigcup_{n=1}^\infty A_n \in \mathcal{F}$.

- The elements of $\mathcal{F}$ are called *measurable sets* or *events*.
- The "sigma" refers to the "countable" expressed in 3.
- De Morgan laws $\implies$ closedness under countable intersections.
- Why do we need σ-algebras and not always work with 2^Ω? We run into issues with defining natural *measures* on uncountable sets. Using σ-algebras of nice sets is enough and fixes the problem.

Probability Measure

Definition 2 (Probability measure)

A *probability measure* on a *measurable space* $(\Omega, \mathcal{F})$ is a mapping $P : \mathcal{F} \to [0,1]$ such that $P[\Omega] = 1$ and P is σ-additive, that is

$$P\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} P[A_n],$$

for $A_n \in \mathcal{F}, n \in \mathbb{N}$ such that $A_k \cap A_n = \emptyset$ if $k \neq n$. The triplet $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

The most basic properties:

- $P[\emptyset] = 0$,
- For $A \in \mathcal{F}$, $P[A^c] = 1 - P[A]$,
- For $A, B \in \mathcal{F}$, $A \subseteq B$, $P[A] \leq P[B]$,
- For $A, B \in \mathcal{F}$, $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

Random Variable

Definition 3 (Random variable)

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A map $X : \Omega \to \mathbb{R}$ is called a *(real-valued) random variable* if

$$X^{-1}(B) = \{X \in B\} = \{\omega \in \Omega \,|\, X(\omega) \in B\} \in \mathcal{F}$$

for all $B \in \mathcal{B}(\mathbb{R})$.

- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra on $\mathbb{R}$, i.e the smallest σ-algebra containing all open sets in $\mathbb{R}$.
- In words: A map is a random variable if the pre-images of (Borel) measurable sets on $\mathbb{R}$ are measurable sets.
- Note that this definition of pre-image works for any map, not just one-to-one maps.
- $\{X \in B\}$ and $X^{-1}(B)$ is just a notation for the set $\{\omega \in \Omega \,|\, X(\omega) \in B\}$.

Distribution of a Random Variable

Definition 4 (Distribution of a random variable)

Distribution or *law* of a random variable $X : \Omega \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is a measure P^X defined by

$$P^X[B] = P[X^{-1}(B)] = P[X \in B] = P[\{\omega \in \Omega \,|\, X(\omega) \in B\}]$$

for all $B \in \mathcal{B}(\mathbb{R})$.

Definition 5 (Distribution function)

(Cumulative) distribution function (cdf) of a real-valued random variable $X : \Omega \to \mathbb{R}$ is a function defined by

$$F_X(x) = P[X \leq x] = P^X\big[(-\infty, x]\big].$$

A More Specific Example I

Example 6

Let $\Omega = \{1, 2, 3\}$, $\mathcal{F} = 2^\Omega = \big\{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \Omega, \emptyset\}\big\}$ and $P[\omega] = 1/3$ for $\omega = 1, 2, 3$. The measure on the atoms determines the measure on all the other sets since they can be written as a finite union of the (disjoint) atoms. Let $X : \Omega \to \mathbb{R}$ be defined by $X(\omega) = 1$ for all $\omega \in \Omega$. Then we have for $B \in \mathcal{B}(\mathbb{R})$ that

$$P^X[B] = P[X^{-1}(B)] = \begin{cases} P[\Omega] = 1 & \text{if } \{1\} \subseteq B, \\ P[\emptyset] = 0 & \text{otherwise.} \end{cases}$$

One can see from the above example that we could have set $\mathcal{F} = \{\Omega, \emptyset\}$ and X would still be a random variable (a measurable map). Such a choice would, however, make it impossible to define other non-degenerate random variables on the same probability space.

A More Specific Example II

Example 7

Let $\Omega = (0,1), \mathcal{F} = \mathcal{B}\big((0,1)\big)$ and $P[A] = \mathcal{L}(A)$ for all $A \in \mathcal{F}$, where $\mathcal{L}$ denotes the Lebesgue measure. Define for a $\lambda > 0$ a random variable $X : \Omega \to \mathbb{R}$ by

$$X(\omega) = \frac{1}{\lambda} \log \left(\frac{1}{1-\omega} \right).$$

We then have that

$$\begin{aligned} F_X(x) &= P\left[X \leq x\right] = P\big[X^{-1}\big((-\infty, x]\big)\big] \\ &= P\left[\frac{1}{\lambda} \log \left(\frac{1}{1-\omega} \right) \leq x\right] = P\left[1 - \omega \geq e^{-\lambda x}\right] \\ &= P[\omega \leq 1 - e^{-\lambda x}] = \mathcal{L}\big[(0, 1 - e^{-\lambda x}]\big] = 1 - e^{-\lambda x}. \end{aligned}$$

This can be recognized as the cdf of the $\mathrm{Exp}(\lambda)$ distribution.

Sigma Algebra Generated by a Random Variable

The σ-algebra generated by a random variable X is the smallest σ-algebra such that X is measurable with respect to that σ-algebra. More formally we can define it as follows.

Definition 8 (σ-algebra generated by a collection of sets)

Let Ω be a non-empty set and $\mathcal{A}$ a collection of subsets of Ω. The *σ-algebra generated by $\mathcal{A}$*, denoted $\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$, that is

$$\sigma(\mathcal{A}) = \{B \subseteq \Omega \,|\, B \in \mathcal{F} \text{ for any } \sigma\text{-algebra } \mathcal{F} \text{ on } \Omega \text{ with } \mathcal{A} \subseteq \mathcal{F}\}.$$

Definition 9 (σ-algebra generated by a random variable)

The *σ-algebra generated by a random variable $X : \Omega \to \mathbb{R}$* is the σ-algebra generated by the the collection of sets of the form $\{X \in B\}, B \in \mathcal{B}(\mathbb{R})$.

Almost Surely

Definition 10 (Almost surely)

Let $(\Omega, \mathcal{F}, P)$ be a probability space. We say that an event $B \in \mathcal{F}$ happens *P-almost surely* if $P[B] = 1$.

- This equivalently means that $P[B^c] = 0$, i.e. the probability of B not happening is zero.
- We often use the abbreviation P-a.s., or simply a.s. when the probability measure in question is clear from the context.
- For instance, if one says that $X \overset{\text{a.s.}}{=} Y$ it means that $P\big[\{\omega \in \Omega \,|\, X(\omega) = Y(\omega)\}\big] = 1$. Similarly for other properties.
- Saying that $X = Y$ is thus stronger than saying that $X \overset{\text{a.s.}}{=} Y$, since $X = Y$ really means that $X(\omega) = Y(\omega)$ pointwise for *every* $\omega \in \Omega$.

Expectation

Definition 11 (Expectation)

The *expectation* of a random variable X on $(\Omega, \mathcal{F}, P)$ with $\int_\Omega |X(\omega)| dP(\omega) < \infty$ is defined as

$$E[X] = \int_\Omega X(\omega) dP(\omega).$$

- The expectation is just a (Lebesgue) integral.
- The set of all random variables X with $\int_\Omega |X|^p dP < \infty$, $p \geq 1$ will be denoted $L^p(P)$ (or L^p if the P in question is clear from the context).
- Useful properties:
  - For $a \in \mathbb{R}$, $E[a] = a$.
  - For $a \in \mathbb{R}$ and r.v.'s $X, Y \in L^1(P)$, $E[aX + Y] = aE[X] + E[Y]$.
  - For $B \in \mathcal{B}(\mathbb{R})$, $P[X \in B] = E[\mathbb{1}_{\{X \in B\}}]$.
  - *Jensen's inequality*: for a convex function $\varphi : \mathbb{R} \to \mathbb{R}$ and $X, \varphi(X) \in L^1(P)$ we have that $\varphi(E[X]) \leq E[\varphi(X)]$.

Monotone Convergence Theorem

Theorem 12 (Monotone convergence theorem)

Let X_n be a non-decreasing sequence of non-negative random variables with $X_n \xrightarrow{P\text{-a.s.}} X$, then

$$\lim_{n\to\infty} E[X_n] = E[X].$$

Note that we do not have any integrability assumption in here. This is because we assume that $X_n \geq 0$ for all $n \in \mathbb{N}$ and there are no problems with defining the integral for any non-negative measurable function (random variable).

Why is it useful?

- Obviously can be used to prove some asymptotic behavior of a sequence of random variables.
- Since $P[B] = E[\mathbb{1}_B]$ we can often compute $P[B]$ by computing simpler $P[B_n]$ for a sequence of sets such that $B_n \subseteq B_{n+1}$ for all $n \in \mathbb{N}$ and using the fact that $\mathbb{1}_{B_n}$ forms an non-decreasing sequence of non-negative functions.

Dominated Convergence Theorem

Theorem 13 (Dominated convergence theorem)

Let X, Y and X_n for $n \in \mathbb{N}$ be random variables with $Y \in L^1$, $X_n \xrightarrow{P\text{-a.s.}} X$ and $|X_n| \leq Y$ for all $n \in \mathbb{N}$, then

$$\lim_{n \to \infty} E[X_n] = E[X].$$

- Notice the integrability condition due to allowing for random variables that are not necessarily non-negative.
- The theorem has applications similar to monotone convergence theorem, but allows for a wider class of sequences.

Conditional Expectation I

Conditional expectation is one of the most important concepts in the course. This is because relevant concepts such as *martingales* (will be discussed later) are defined using the notion of conditional expectation. Prices (and price processes) of derivative products can be conveniently expressed using conditional expectation as well, which we will see later.

Definition 14 (Conditional expectation)

Let $(\Omega, \mathcal{F}, P)$ be a probability space and fix a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. A *conditional expectation* of a random variable X given $\mathcal{G}$ is a *random variable* Y with the following two properties:

1. Y is $\mathcal{G}$-measurable (i.e. $Y^{-1}(B) \in \mathcal{G}$ for all $B \in \mathcal{B}(\mathbb{R})$),
2. $E[X\mathbb{1}_A] = E[Y\mathbb{1}_A]$ for all $A \in \mathcal{G}$.

Any such random variable Y will be denoted $E[X \mid \mathcal{G}]$.

Conditional Expectation II

- Unlike the classical expectation, the conditional expectation is a *random variable*. We sometimes explicitly write $E[X \mid \mathcal{G}](\omega)$.
- It can be shown that $E[X \mid \mathcal{G}]$ is the best $\mathcal{G}$-measurable approximation of (a generally $\mathcal{G}$-non-measurable) random variable X. The word "best" is always tied to a criterion – in this case, the best random variable $E[X \mid \mathcal{G}]$ minimizes the distance given by

$$d(X, Y) := E[(X - Y)^2]$$

between our X and all $\mathcal{G}$-measurable and square-integrable random variables Y.

Conditional Expectation III

The most important and the most applicable properties of the conditional expectation include the following:

- For $a \in \mathbb{R}$ and r.v.'s $X, Y \in L^1(P)$, $E[aX + Y \,|\, \mathcal{G}] = aE[X \,|\, \mathcal{G}] + E[Y \,|\, \mathcal{G}]$.
- $E[X \,|\, \mathcal{G}] = X$ if X is $\mathcal{G}$-measurable.
- $E\big[E[X \,|\, \mathcal{G}]\big] = E[X]$.
- Tower property: for a σ-algebra $\mathcal{H} \subseteq \mathcal{G}$, $E\big[E[X \,|\, \mathcal{G}] \,|\, \mathcal{H}\big] = E[X \,|\, \mathcal{H}]$.
- Jensen's inequality: For a convex function $\varphi : \mathbb{R} \to \mathbb{R}$ and $X, \varphi(X) \in L^1(P)$ we have that $\varphi(E[X \,|\, \mathcal{G}]) \leq E[\varphi(X) \,|\, \mathcal{G}]$ P-a.s.
- Monotone convergence theorem: $\lim_{n\to\infty} E[X_n \,|\, \mathcal{G}] = E[X \,|\, \mathcal{G}]$ P-a.s. for a non-decreasing sequence of non-negative random variables with $X_n \xrightarrow{P\text{-a.s.}} X$.
- Dominated convergence theorem: $\lim_{n\to\infty} E[X_n \,|\, \mathcal{G}] = E[X \,|\, \mathcal{G}]$ P-a.s. for a sequence of random variables with $X_n \xrightarrow{P\text{-a.s.}} X$ and $|X_n| \leq Y$ for a $Y \in L^1(P)$.

Stochastic Processes

Definition 15 (Stochastic process)

Let $\mathcal{T}$ be an index set. A *stochastic process* is a collection of random variables X_t, $t \in \mathcal{T}$ defined on a common probability space $(\Omega, \mathcal{F}, P)$.

- In the first part of the lecture will be mostly dealing with the case $\mathcal{T} = \{0, 1, \dots, T\}$, $T \in \mathbb{N}$ and finite Ω, while in the second part we will mostly have $\mathcal{T} = [0, T]$, $T \in (0, \infty)$ and uncountable Ω.
- There is one an additional view on stochastic processes that we will use – a collection of *sample paths* (or a *trajectories*) of the process indexed by $\omega \in \Omega$, i.e.

$$\begin{aligned} X(\omega, \cdot) : \mathcal{T} &\to \mathbb{R} \\ t &\mapsto X_t(\omega). \end{aligned}$$

Filtration

Filtration is the key concept required for formalizing some of the most important types of behavior of stochastic processes over time.

Definition 16 (Filtration)

Let $\mathcal{T} \subseteq [0, \infty)$. A *filtration* $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{T}}$ on a measurable space $(\Omega, \mathcal{F})$ is a family of σ-algebras $\mathcal{F}_t \subseteq \mathcal{F}$ which is increasing in the sense that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$.

- A *filtered probability space* $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is simply a probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{T}}$.
- One can easily show that $\mathcal{F}_t := \sigma(X_s, 0 \leq s \leq t)$ forms a filtration. Since this is the most natural filtration for a probability space where we are dealing with a single stochastic process X_t, it is also referred to as the *natural filtration* for X_t.

Filtration as the Flow of Information

One can often hear that filtration models the evolution of information over time.

- The richer a σ-algebra is, the more events can be assigned a probability and are thus observable.
- For instance, if $\Omega = \{1, \ldots, 6\}$ and $X(\omega) = \omega$ corresponds to a value on a die, then the σ-algebra one needs to use (so that the corresponding r.v. is measurable) when she is told the precise value thrown on the die ($\mathcal{F} = 2^\Omega$) is larger than the σ-algebra she needs to use when she is only told whether the number is even or odd $\left(\bar{\mathcal{F}} = \big\{\{1,3,5\}, \{2,4,6\}, \Omega, \emptyset\big\}\right)$. Larger σ-algebras thus convey more information.
- A filtration $\mathbb{F}$ can therefore be interpreted as the evolution of information over time, and the requirement that $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ as information not being lost over time, which is reasonable.

Adapted Processes

Definition 17 (Adapted process)

A stochastic process X_t on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}}, P)$ is *adapted* to $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{T}}$ if X_t is $\mathcal{F}_t$-measurable for all $t \in \mathcal{T}$.

An adapted process simply represents a process whose values can be observed at time t (regardless of which of the possible values is taken).

Example 18

A good example in the financial context: Online brokerage firms often provide their clients with data delayed by ϵ unless the clients pay for the real-time prices. If we denote the real-time prices by X_t and the delayed prices by Y_t, we clearly have that $Y_t = X_{t-\epsilon}$. In this case X_t will not be adapted to the filtration representing the information that we have available at time t from the price process that we observe, $\mathcal{F}_t = \sigma(Y_s, 0 \leq s \leq t)$. However, this process would of course be adapted to $\mathcal{G}_t = \sigma(X_s, 0 \leq s \leq t)$, which is why it might seem that processes which are not adapted are not natural.

Predictable Processes in Discrete Time

Definition 19 (Predictable process)

A *discrete time* stochastic process $(X_t)_{t=0,1,\dots,T}$, $T \in \mathbb{N}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,1,\dots,T}, P)$ is *predictable* with respect to $\mathbb{F} = (\mathcal{F}_t)_{t=0,1,\dots,T}$ if X_t is $\mathcal{F}_{t-1}$-measurable for all $t = 1, \dots T$.

- The definition is a bit more delicate for continuous time processes.
- One requires *trading strategies* to be predictable processes. This is because in order to be able to take advantage of price movement from the current time t to the future time $t+1$, one must make an investment at time t.
- Conveniently, stochastic integral, which, as will be seen later, naturally represents the gains of a trading strategy, can be defined for predictable processes in general.

Thank you for your attention!