Stochastic Calculus and Option Pricing
Week 1: Infinite Probability Spaces, σ-algebras, Conditional Expectations

Roza Galeeva

September 9, 2020

Outline of the lecture

Most of the material for this Lecture is based on the textbook by Steven Shreve *Stochastic Calculus for Finance, II*, chapters 1,2

The main Learning Goals of the Lecture

1. Notion of σ-algebra and information
2. Probability Space
3. Random Variable
4. Conditional Expectation and Independence

Outline of the lecture:

- Elementary Probability Theory
- Infinite Probability Spaces
- σ-algebra, examples
- Conditional Expectations

Brief Recall of Elementary Probability Theory

We consider an experiment where all possible outcomes can be described by a finite number of events:

$$\Omega = \{\omega_1, \omega_2.., \omega_N\}$$

Ω is a probability space, and ω_i is an elementary event. We can also consider subsets of Ω, which we call events. If we consider a set of events $\mathcal{A}_0 \subset \Omega$, then with operations like $\cup$, $\cap$ and complement, we can construct new sets, which are also events, Adding to the set "impossible" event $\emptyset$ and the "sure" event Ω, we will get a system of sets $\mathcal{A}$, which is *algebra*, which means:

i $\Omega \in \mathcal{A}$, $\emptyset \in \mathcal{A}$

ii if $A \in \mathcal{A}$ and $B \in \mathcal{A}$, then $A \cup B$, $A \cap B$, $A \setminus B$ also belong to $\mathcal{A}$.

Probability Space

For each elementary event $\omega_i \in \Omega$ we assign a "weight" $p(\omega_i)$, which we call a probability of event ω_i with

(a) $0 \leq p(\omega_i) \leq 1$ (non-negativity)

(b) $p(\omega_1) + ... + p(\omega_N) = 1$

Then for any $A \in \mathcal{A}$

$$\mathcal{P}(A) = \sum_{i:\omega_i \in A} p(\omega_i)$$

The triple

$$(\Omega, \mathcal{A}, \mathcal{P})$$

with $\mathcal{A}$ being an algebra of Ω and $\mathcal{P}$ gives a probability model, defines a probability space with finite number of events.

Infinite Probability Spaces

An infinite probability space is used to model a random experiment with infinitely many possible outcomes. For example:

(i) choose a number from the unit interval $[0, 1]$

(ii) toss a coin infinitely many times.

For (i) our sample space is the unit interval, for (ii) we define Ω_∞= the set of infinite sequences of Hs and Ts (or 1 and 0). It is well known that any number $a \in [0, 1]$ can be uniquely represented as

$$a = \frac{a_1}{2} + \frac{a_2}{2^2} ... (a_i = 0, 1)$$

Therefore, the set Ω_∞ is *uncountable*.

σ-algebra

Definition Let Ω be a nonempty set, and let $\mathcal{F}$ be a collection of subsets of Ω. We say $\mathcal{F}$ is a σ-algebra (or σ-field) provided that:

(i) the empty set $\emptyset$ belongs to $\mathcal{F}$

(ii) whenever a set A belongs to $\mathcal{F}$, its compliment A^c also belongs to $\mathcal{F}$

(iii) whenever a sequence of sets $A_1, A_2, \ldots$ belongs to $\mathcal{F}$, their union $\cup_{n=1}^{\infty} A_n$ also belongs to $\mathcal{F}$

Examples:

$$\mathcal{F}_1 = \{\emptyset, \Omega\}$$

$$\mathcal{F}_2 = \{\emptyset, \Omega, A, A^c\} \; \} \text{ for some } A \neq \emptyset \text{ and } A \neq \Omega$$

$$\mathcal{F}_3 = \mathcal{P}(\Omega) = \{A : A \subset \Omega\}$$

$$\Omega = \{a, b, c, d\} \; , \; \mathcal{F} = ..$$

$\mathcal{F}_1$ is the smallest σ-field on Ω and $\mathcal{P}(\Omega)$ is the biggest one, as it contains all possible subsets of Ω.

Probability Measure

Definition

Let Ω be a non-empty space, and let $\mathcal{F}$ be a σ-algebra of subsets of Ω. A probability measure $\mathcal{P}$ is a function that, to every set $A \in \mathcal{F}$, assigns a number in $[0, 1]$, called the probability of A and written $\mathcal{P}(A)$. It is required that:

(i) $\mathcal{P}(\Omega) = 1$, and

(ii) whenever $A_1, A_2, ..$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$\mathcal{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{n=1}^{\infty} \mathcal{P}(A_n)$$

The triple $(\Omega, \mathcal{F}, \mathcal{P})$ is called a probability space.
We have $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$,
For two disjoint sets A, B : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, and $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

Lebesgue measure and Borel sets

Model to choose at random from the unit interval $[0,1]$, so that the probability is distributed uniformly over the interval:

$$\mathcal{P}([a,b]) = b - a, 0 \leq a \leq b \leq 1$$

This particular measure on $[0,1]$ is called *Lebesgue* measure, denoted by $\mathcal{L}$. It is defined on collection of *Borel sets* on $[0,1]$ denoted by $\mathcal{B}[0,1]$ and defined as the smallest σ-field of subsets on $[0,1]$ containing all intervals $(a,b]$, $a,b \in [0,1]$.
In the same way, we can define Borel sets on the whole $\mathbb{R}$ and $a = -\infty, b = \infty$ and σ-field $\mathcal{B}(\mathbb{R})$. It contains

all closed intervals

all open intervals

all intervals $[a,b)$, $a,b \in \mathbb{R}$

For instance,

$$(a,b) = \cup_{n=1}^{\infty} (a, b - 1/n]$$

Random Variable

Definition

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. A random variable is a real-valued function X defined on Ω with the property that for every Borel subset B of $\mathbb{R}$ the subset of Ω given by

$$\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\}$$

is in the σ-algebra $\mathcal{F}$.

Example: (binomial model for stock prices). Let consider a classical example of infinite coin-toss. We define

$$S_0 = 4, S_1(H) = 8, S_1(T) = 2, S_2(HH) = 16, S_2(TT) = 1, S_2(HT) = ..$$

$$S_2(TH) = 4...$$

We call a r.v. X *integrable* if the expectation $\mathbb{E}(|X|) < \infty$.

Examples of σ-fields

For a given collection $\mathcal{C}$ of subset Ω, there exists a smallest σ- field $\sigma(\mathcal{C})$ on Ω containing $\mathcal{C}$. We call $\sigma(\mathcal{C})$ the σ-field generated by $\mathcal{C}$

Example 1: Using previous examples, $\mathcal{F}_i = \sigma(\mathcal{C}_i)$, where $\mathcal{C}_1 = \{\emptyset\}$, $\mathcal{C}_2 = \{A\}$

Example 2 The σ-field generated by a discrete random variable: We consider a discrete random variable Y with distinct values y_i and defined by subsets $A_i = \{\omega : Y(\omega) = y_i\}$ which constitute a joint partition of Ω. Choose

$$\mathcal{C} = \{A_1, A_2 ....\}$$

$\sigma(\mathcal{C})$ must contain all sets of the form: $A = \cup_{i \in I} A_i$ where I is any subset of $\mathbb{N} = \{1, 2, ...\}$, including $I = \emptyset$ (giving $A = \emptyset$ and $I = \mathbb{N}$ (giving $A = \Omega$)

Note that $\sigma(Y)$ contains all the sets of the form:
$A_{a,b}\{Y \in (a, b]\} = \{\omega : a < Y(\omega) \leq b\}, -\infty < a < b < \infty$.

Conditional Expectations

i From an elementary probability theory we know the conditional probability of A given B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

ii Consider the case of a discrete random variable Y on Ω that assumes the distinct values y_i and on the sets A_i

$$A_i = \{\omega : Y(\omega) = y_i\}, A_i \cap A_j = \emptyset, i \neq j, \cup_{i=1}^{N} A_i = \Omega$$

For a r.v. $X \in \Omega$ with $\mathbb{E}(X) < \infty$ we define the conditional expectation of X given Y as discrete r. v.

$$\mathbb{E}(X|Y)(\omega) = \mathbb{E}(X|A_i) = \mathbb{E}(X|Y = y_i), \omega \in A_i$$

Properties of Conditional Expectations

a. The conditional expectations are linear:

$$\mathbb{E}\left([c_1X_1 + c_2X_2]|Y)\right) = c_1\mathbb{E}(X_1|Y) + c_2\mathbb{E}(X_2|Y)$$

b. The expectations of X and $\mathbb{E}(X|Y)$ are the same:

$$\mathbb{E}\left(\mathbb{E}(X|Y)\right) = \sum_{i=1}^{N} \mathbb{E}(X|A_i)\mathcal{P}(A_i) = \sum_{i=1}^{N} \mathbb{E}(XI_{A_i}) =$$

$$\mathbb{E}(X\sum_{i=1}^{N} I_{A_i}) = \mathbb{E}X$$

We used the fact that

$$A_i \cap A_j = \emptyset, i \neq j, \bigcup_{i=1}^{\infty} A_i = \Omega$$

Summary on conditional expectations

So far, we have:

- The conditional expectation $\mathbb{E}(X|Y)$ of X given a discrete random variable Y is a discrete random variable
- It coincides with the classical conditional expectation $\mathbb{E}(X|Y = y_i)$ on the sets $A_i = \{\omega : Y(\omega) = y_i\}$
- The fewer values Y has, the coarser the random variable $\mathbb{E}(X|Y)$. In particular, if $Y = const$, then $\mathbb{E}(X|Y) = \mathbb{E}(X)$
- The conditional expectation $\mathbb{E}(X|Y)$ is not a function of X, but merely a function of Y : $\mathbb{E}(X|Y) = g(Y)$, where g is:

$$g(Y) = \sum_{i=1}^{N} \mathbb{E}(X|Y = Y_i) I_{\{y_i\}}(y)$$

Thus, the conditional expectation can be understood as a random variable constructed from a collection $\sigma(Y)$, so

$$\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma(Y))$$

Simple Example

Consider a probability space Ω with four elements $\Omega = \{a, b, c, d\}$. We define a probability measure $\mathcal{P}$ by

$$\mathcal{P}(a) = \frac{3}{8}, \mathcal{P}(b) = \frac{1}{8}, \mathcal{P}(c) = \frac{1}{6}, \mathcal{P}(d) = \frac{1}{3}$$

and the probability of every other set is the sum of probabilities of the elements in the set. For example, $\mathcal{P}\{a, b\} = \mathcal{P}(a) + \mathcal{P}(b)$
We define two random variables X and Y by the formula:

$$X(a) = 1, X(b) = 1, X(c) = -1, X(d) = -1$$

$$Y(a) = -1, Y(b) = 1, Y(c) = 1, Y(d) = -1$$

1. List all sets in σ-algebra $\mathcal{F}$
2. List all sets in $\sigma(X)$
3. Determine $\mathbb{E}[Y|X]$

Information carried a by r.v

a. If X is a r.v. $\Omega \to \mathbb{R}$. We will say that a set $A \subset \Omega$ is *determined* by the r.v. X if, knowing only the value $X(\omega)$ of the r.v. we can decide if whether or not $\omega \in A$.

b. Another way of saying this is that for every $y \in \mathbb{R}$, either $X^{-1}(y) \subset A$ or $X^{-1} \cap A = \emptyset$.

c. The collection of subsets of Ω determined by X is σ-algebra generated by X and denote by $\sigma(X)$.

d. If the r.v X takes finitely many values, then $\sigma(X)$ is generated by the collection of sets

$$\{X^{-1}(X(\omega)) | \omega \in \Omega\}$$

which are called the atoms

e. In general, if X is a r.v. $\Omega \to \mathbb{R}$, then $\sigma(X)$ is given by

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\}$$

Independence

1. As usual we act on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$
   Let $\mathcal{H}, \mathcal{G}$ are sub-algebras of $\mathcal{F}$. We will say that these sigma fields are independent, if

   $$\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B), \forall A \in \mathcal{H}, B \in \mathcal{G}$$

2. Let X and Y be two r.v., which generate two σ-fields $\sigma(X)$ and $\sigma(Y)$. Then we say X and Y are independent, if their σ-algebras are independent.
3. If X and Y are independent r.v., then

   $$\mathbb{E}(X|Y) = \mathbb{E}(X) \tag{1}$$

   The proof is straightforward and is left as exercise.

σ-algebra, generated by a random variable

Let X be a r. v. defined on a nonempty sample space Ω.
Definition of $\sigma(X)$: The σ-algebra generated by X, denoted by $\sigma(X)$, is the collection of all subsets of Ω of the form $\{X \in \mathcal{B}\}$, where $\mathcal{B}$ is any Borel subset of $\mathbb{R}$.

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\}$$

Definition of measurability : Let $\mathcal{G}$ be a σ-algebra of subsets of Ω. If every set in $\sigma(X)$ is also in $\mathcal{G}$, we say that X is $\mathcal{G}$-measurable. A random variable X is $\mathcal{G}$-measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X.

Existence and uniqueness

1. **Proposition A** Consider the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and $\mathcal{G}$ be a σ field included in $\mathcal{F}$. If X is $\mathcal{G}$-measurable r.v. such that
$$\int_A XdP = 0, \forall A \in \mathcal{G}$$
then $X = 0$ a.s. (almost surely).
2. **Theorem A** Let $(\Omega, \mathcal{F}, \mathcal{P})$ and $\mathcal{G}$ be a σ field included in $\mathcal{F}$. Then for any r.v. X there is a $\mathcal{G}$-measurable r.v Y such that
$$\int_A XdP = \int_A YdP, \forall A \in \mathcal{G}$$
Moreover, this r.v Y is unique due to the proposition A. The r.v. Y plays the role if the expectation of X given the partial information $\mathcal{G}$.

General Conditional Expectations

Let X be a r.c. on probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let $\mathcal{G}$ be a σ algebra, $\mathcal{G} \in \mathcal{F}$. It is natural to ask what is the expectation of X given the information $\mathcal{G}$. This is a r.v. denoted as $\mathbb{E}[X|\mathcal{G}]$, satisfying the following properties:

1. $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$ -measurable.
2. $$\int_A \mathbb{E}[X|\mathcal{G}] dP = \int_A X dP, \forall A \in \mathcal{G}$$

Example 1 If $\mathcal{G} = \{\emptyset, \Omega\}$ then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}(X)$.
Example 2 The conditional expectation of X given the total information $\mathcal{F}$ is the r.v. itself:

$$\mathbb{E}[X|\mathcal{F}] = X$$

The existence of r.v. $\mathbb{E}[X|\mathcal{G}]$ is guaranteed by theorem A.

Properties of general expectations

We have all listed previously properties a, b, c (linearity, expectation of expectation, expectation under independence). Moreover, we have:

d "Taking out what is known" if X, Y are two r.v. , X is $\mathcal{G}$-measurable, then

$$\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}] \tag{2}$$

e. "Tower Property". If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then

$$\mathbb{E}[\mathbb{E}[X\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] \tag{3}$$

f. An independent condition drops out:

$$\mathbb{E}[\mathbb{E}[X\mathcal{G}] = \mathbb{E}[X]$$

if X is independent of $\mathcal{G}$.

Jensen inequality

Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function and X is an integrable r.v. on probability space $(\Omega, \mathcal{F}, \mathcal{P})$. If $\phi(X)$ is integrable, then

$$\phi(\mathbb{E}[X)] \leq \mathbb{E}[\phi(X)] \tag{4}$$

almost surely (so inequality might fail on a set of probability zero).
Proof: We assume that $\phi(x)$ is twice differentiable with $\phi^{''}$ continuous. Let $\mu = \phi(X)$. We expand $\phi(x)$ in a Taylor series around μ:

$$\phi(x) = \phi(\mu) + \phi^{'}(\mu)(x - \mu) + \frac{1}{2}\phi^{''}(y)(y - \mu)^2$$

where y is between μ and x. Since ϕ is convex, $\phi^{''}(y) > 0$ and

$$\phi(x) \geq \phi(\mu) + \phi^{'}(\mu)(x - \mu)$$

Replacing x by r.v. X and taking expectations leads to the inequality:

$$\mathbb{E}[\phi(X)] \geq \mathbb{E}[\phi(\mu) + \phi^{'}(\mu)(X - \mu)] = \phi(\mu) = \phi(\mathbb{E}(X)]$$

Illustration of Jensen inequality

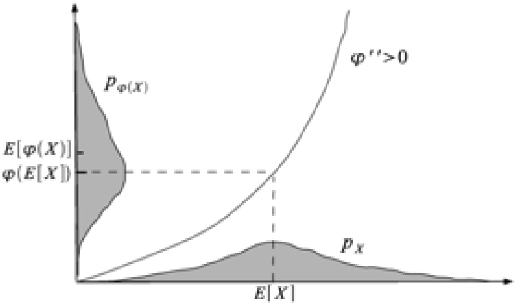

Figure: Jensen Inequality

Applications of the Jensen inequality

A r.v. $X : \Omega \to \mathbb{R}$ is called *square integrable*, if

$$\mathbb{E}[X^2] = \int_\Omega (X(\omega))^2 dP(\omega) = \int_{\mathbb{R}} x^2 p(x) dx < \infty$$

Let X be a nonnegative r.v. Define a *moment generating function*:

$$\psi_X(t) = \mathbb{E}e^{tX}$$

1. Application 1: If X is square integrable r.v., then it is integrable. It follows then that the variance of r.v. X (if exists), is non-negative:

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0$$

2. Application 2. If $\psi_X(t)$ is a moment generating function of r.v. X with mean μ , then

$$\psi_X(t) \geq e^{t\mu}$$

Other inequalities

1. *Markov inequality*: For $\lambda, p > 0$, we have

$$\mathcal{P}(\omega : |X(\omega)| \geq \lambda) \leq \frac{1}{\lambda^p}\mathbb{E}[|X|^p] \tag{5}$$

   To prove, define a set $A = \{\omega : |X(\omega)| \geq \lambda\}$ and calculate the expectation $\mathbb{E}[|X|^p]$.

2. *Tchebychev inequality*: If X is a r.v. with mean μ and variance σ^2, then

$$\mathcal{P}\left(\omega : |X(\omega) - \mu| \geq \lambda\right) \leq \frac{\sigma^2}{\lambda^2} \tag{6}$$

   To prove define a set $A = \{\omega : |X(\omega) - \mu| \geq \lambda\}$ and calculate the variance of X.

Limits of sequences of Random Variables

Consider a sequence $(X_n)_{n\geq 1}$ of r.v. defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$. There are several ways of making sense of the limit expressions $X = \lim_{n\to\infty} X_n$.

1. Almost certain limit : The sequence X_n converges *almost certainly or strongly* to X , if for states of the world ω except a set of probability zero we have

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

An example includes a sequence of i.i.d r.v X_n with the same mean μ:

$$a.c \lim_{n\to\infty} \frac{[X_1 + \ldots + X_n]}{n} = \mu$$

2. Mean Square Limit : We say X_n converges to X in the mean square sense, if

$$\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0$$

The mean square convergence will be useful to define the Itô integral.

Example of limits of r.v.

1. Almost certain convergence
   A fair coin is tossed once. Thus, the sample space is $\Omega = \{H, T\}$. We repeat it n times, for each trial number n define a r.v. X_n as follows:

   $$X_n = \begin{cases} \frac{1}{n+1} & \text{if } \omega = \text{H} \\ 1 & \text{otherwise} \end{cases}$$

   When $n \to \infty$, $X_n(\omega)$ converges to a r.v. $X(\omega)$, where $X(H) = 0$ and $X(T) = 1$ (classical Bernoullli r.v).
2. Mean-square convergence
   Consider a sequence of r.v. X_n so that $\mathbb{E}[X_n] \to k$ and $Var[X_n] \to 0$, as $n \to \infty$. Then X_n converges to a r.v. $X = k$ (constant) in the mean-square sense.

Symmetric random walks

We begin with a symmetric random walk: 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 $q = 1 - p = \frac{1}{2}$. The successive outcomes of the tosses are denoted by $\omega = \omega_1 \omega_2 \omega_3 ...$ 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:

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

Increments of the Symmetric Random Walk

A random walk has *independent* increments. 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

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

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.

In addition,

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

Variance of the increments

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

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