STOCHASTIC ANALYSIS: AN INTRODUCTION

ESKO VALKEILA

Abstract. The purpose of these lectures is to introduce stochastic integrals with respect to standard Brownian motion, or more generally with respect to continuous square integrable martingales. Before this we discuss discrete time parameter martingales to be familiar with some of the techniques needeed later. We also prove some results for discrete time parameter martingales, like moment inequalities and the important martingale convergence theorem. After this, and after defining stochastic integrals, we give some classical applications of the fundamental Itô formula: Lévy theorem to characterize Brownian motion and Girsanov theorem. Another application are the iterated integrals with respect to Brownian motion; these are useful for example in Malliavin calculus. We try to cover also the basic facts about stochastic differential equations. If the time permits, we discuss some other applications.

Contents

1. Basic set-up, background

1.1. Practicalities. Lectures are on Thursdays and on Tuesdays, first lecture on Thursday, March 13, and last on Tuesday, April 29. The course consists of three parts: lectures, exercises, individual homework and final exam.

1.2. Background. There are several excellent books on this area. We will compose our lectures mainly from the following sources:

- K.L. Chung, R.J. Williams: Introduction to Stochastic Integration, Birkhuser
- J. Jacod, P. Protter: Probability Essentials, Springer
- Karatzas, S.E. Shreve: Brownian Motion and Stochastic Calculus, Springer
- D. Revuz, M. Yor: Continuous Martingales and Brownian Motion, Springer
- A.D. Ventsel: A Course in the Theory of Random Processes (in Russian), Nauka[1]

1.3. Probability theory. If you feel uncomfortable with the following material, please consult ASAP the book by Jacod and Protter mentioned above [or consult your favorite book on measure theoretical probability].

1.3.1. *Probability spaces and random variables.* Recall that a probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set containing all possible elementary events, and $\mathbb{P}$ is a probability measure defined on the measurable space $(\Omega, \mathcal{F})$. Random variables X are *measurable* mappings from $(\Omega, \mathcal{F}, \mathbb{P})$ to some other measurable space $(S, \mathcal{S})$. If X is such a random variable, then P_X is a probability measure on $(S, \mathcal{S})$ defined by $P_X(B) = \mathbb{P}\{\omega : X(\omega) \in B\}$ for $B \in \mathcal{S}$. The probability measure P_X is called the *law* of X. Typically, $(S, \mathcal{S})$ is the real line with its Borel sets or the set of nonnegative integers $\mathbb{N}$ with all subsets in $\mathbb{N}$. In the latter case we say that X takes integer values, and the law of X can be given in terms of the discrete probabilities α_k:

$$\alpha_k = \mathbb{P}(X = k) = P_X(\{k\}).$$

Of course, we have $\sum_k \alpha_k = 1$ and $\alpha_k \geq 0$. On the other hand, if X takes its values on the real line, it often haves a density f_X, which completely describes the law P_X of X:

$$\int_a^b f_X(y)dy = \mathbb{P}(X \in (a,b)) = P_X(\{(a,b)\}).$$

1.3.2. *Expectations [or integrals with respect to probability measure].* If X is a simple random variable on real line, which means that we can write X as

$$X = \sum_{k=1}^{n} y_k 1_{B_k},$$

[1] At least a German translation of this book exists, based on an earlier edition of this interesting book, which is suitable for self study.

then the expectation $\mathbb{E}X$ is written as

$$\mathbb{E}X = \sum_{k=1}^{n} y_k \mathbb{P}(X \in B_k) = \sum_{k=1}^{n} y_k P_X(B_k) = E_{P_X} X. \tag{1.1}$$

Having this we can show that every nonnegative random variable X has an expectation $\mathbb{E}X$ which is defined as

$$\mathbb{E}X = \sup\{\mathbb{E}Y : 0 \le Y \le X; Y \text{ is simple}\};$$

At this point we can allow $\mathbb{E}X = \infty$. Recall that a nonnegative random variable X is *integrable*, if $\mathbb{E}X < \infty$. For general real valued random variables X we define $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$, and then we have $X = X^+ - X^-$; the random variable X is integrable, if X^+ and X^- are integrable, and then we put

$$\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-.$$

Example 1.1.

- *For integer valued X we have*

$$\mathbb{E}X = \sum_k k\alpha_k.$$

- *For integrable and real valued X with density f_X we have*

$$\mathbb{E}X = \int_{-\infty}^{\infty} y f_X(y) dy.$$

Notation: $\mathbb{E}X = \int X d\mathbb{P} = \int_\Omega X(\omega) d\mathbb{P}(\omega)$.

1.3.3. *Convergence modes.* Now X_n, $n \ge 1$ and X are random variables. Recall the following types of convergence:

- $X_n \xrightarrow{\mathbb{P}} X$ if for any ϵ

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0$$

as $n \to \infty$. This is *stochastic convergence.*

- $X_n \xrightarrow{\text{a.s.}} X$ if

$$\mathbb{P}\left(\omega : \lim_n X_n(\omega) = X(\omega)\right) = 1.$$

This is *almost sure convergence.*

- Take $p \ge 1$: $X_n \xrightarrow{L^p} X$ if

$$\lim_n \mathbb{E}|X_n - X|^p = 0.$$

This is *convergence in L^p*. Typically we will use convergence in L^2.

If X_n co verges to X almost surely or in L^p, then X_n converges to X stochastically. If X_n converges to X stochastically, then X_n can fail to converge to X almost surely or in L^p. If X_n converges almost surely to X, then it can fail to converge to X in L^p, and finally, if X_n converges to X in L^p, it can fail to converge to X almost surely.
Recall also that if $X_n \to X$ stochastically, then there exists a subsequence n_k such that $X_{n_k} \to X$ almost surely.

1.3.4. *Expectation and limits.* In this section the sequence X_n converges stochastically or almost surely, and we write simply $X_n \to X$. Recall the following theorems

- Monotone converge theorem (MCT): If $0 \leq X_1 \leq X_2 \leq \cdots$ and $X_n \to X$, then
$$\lim_n \mathbb{E} X_n = \mathbb{E}\left(\lim_n X_n\right) = \mathbb{E} X.$$
- Fatou's lemma: If $X_n \geq Y$ and Y is an integrable random variable, then
$$\mathbb{E} \liminf_n X_n \leq \liminf_n \mathbb{E} X_n.$$
Note that we do assume any convergence here.
- Dominated convergence theorem (DCT): If $X_n \to X$ and Y is an integrable random variable such that $|X_n| \leq Y$, then
$$\lim_n \mathbb{E} X_n = \mathbb{E}\left(\lim_n X_n\right) = \mathbb{E} X.$$

1.3.5. *The space* $L^2(\Omega, \mathcal{F}, \mathbb{P}$. We close the necessary background by recalling some properties of the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$.
The space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a *Hilbert space.* The inner product is given by
$$(X, Y)_{L^2(\mathbb{P})} = \mathbb{E}\,(X, Y)\,.$$
The norm of the vector X is $||X||_{L^2(\mathbb{P})} = \sqrt{\mathbb{E} X^2}$. *Orthogonality* of two vectors $X, Y \in L^2(\mathbb{P})$ is defined in the usual way: $X \perp Y \Leftrightarrow \mathbb{E}(XY) = 0$.
X_n is a Cauchy-sequence: for any $\epsilon > 0$ there exists an n_ϵ such that $n, m \geq n_\epsilon$ and then $||X_n - X_m||_{L^2(\mathbb{P})} < \epsilon$. The space $L^2(\mathbb{P})$ is *complete*: If X_n is a c-sequence, then there exists a vector $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \stackrel{L^2}{\to} X$.
We end by recalling the *projection* theorem. Recall that $H \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a subspace, if $X, Y \in H, \alpha, \beta \in \mathbb{R} \Rightarrow \alpha X + \beta Y \in H$ and H is closed: if X_n converges in $L^2(\mathbb{P})$ to X, and $X_n \in H$, then also $X \in H$.

Proposition 1.1 (Projection). *Let $H \subset L^2(\Omega, \mathbb{F}, \mathbb{P})$ be a subspace. Then for every $X \in L^2(\mathbb{P})$ there exists a unique $Y = \Pi_H X \in H$, the* projection *of X on H with the properties*

- $||X - Y||_{L^2(\mathbb{P})} = \inf\left\{||X - Z||_{L^2(\mathbb{P})} : Z \in L^2(\mathbb{P})\right\}$.
- $X - \Pi_H X \perp H$.

Example 1.2. *The following is a fundamental observation for us. Let $\mathcal{G} \subset \mathcal{F}$ and put $H = L^2(\Omega, \mathcal{G}, \mathbb{P})$. Then H is a subspace: H is clearly linear, and it is closed, because the almost sure limit of $\mathcal{G}$ measurable random variables is $\mathcal{G}$ measurable. Then by the Projection theorem there exists unique $Y = \Pi_H X$, which is $\mathcal{G}$ measurable and $X - Y \perp H$.*

After there preparations we can slow down and start to discuss conditioning.

1.4. **Conditioning.**

1.4.1. *From concrete to abstract.* One of the key features of probability theory is the notion of conditional probability and expectation. If $A \in \mathcal{F}$ and $0 < \mathbb{P}(A) < 1$, we define the conditional probability of an arbitrary event B with the formula

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

Note that in this case we also can define the conditional probability with respect to the event $A^c := \Omega \backslash A$. We can also define the conditional expectation with respect to the event A of a simple random variable $X = \sum_{k=1}^m x_k 1_{B_k}$ with the formula

$$\mathbb{E}(X|A) = \sum_{k=1}^m x_k \mathbb{P}(B_k|A).$$

All the above is based on the fact that we know $\mathbb{P}(A)$ or the event A^c. A very useful idea is to think that we know only that A will happen or A^c will happen. Then the conditional expectation of X will be a random variable, depending weather A has happened of A^c has happened; we can write this as follows

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X|A)1_A + \mathbb{E}(X|A^c)1_{A^c},$$

where $\mathcal{G}$ is the σ- algebra $\mathcal{G} = \{\emptyset, A, A^c, \Omega\}$. An exercise is to check that if we take Y to be the random variable

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

then Y satisfies the two conditions:

- The random variable Y is measurable with the elementary σ- algebra $(\emptyset, A, A^c, \Omega)$.
- The random variable Y satisfies the *integral test*: for every $G \in \{\emptyset, A, A^c, \Omega\}$ we have
$$\int_G Y d\mathbb{P} = \int_G X d\mathbb{P}.$$

To generalize the above is the topic of the next section.

1.5. **Conditional expectations.**

The following definition is one of the contibutions of Kolmogorov to probability theory.

Definition 1.1 (Kolmogorov). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X \in L^1(\mathbb{P})$ and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. The conditional expectation of random variable X with respect to $\mathcal{G}$ $\mathbb{E}[X|\mathcal{G}]$ is a random variable $Y \in \mathcal{G}$, which satisfies that*

$$\int_G X d\mathbb{P} = \int_G Y d\mathbb{P} \ \forall G \in \mathcal{G}. \tag{1.2}$$

For us random variables are in fact equivalence classes of random variables, which are almost surely the same. With this interpretation the conditional expectation $\mathbb{E}(X|\mathcal{G})$ is unique.

If the random variable $X \in L^2(\mathbb{P})$, then the conditional expectation of X with respect to $\mathcal{G}$ is the projection of X to the subspace $H = L^2(\Omega, \mathcal{G}, \mathbb{P})$. Indeed, $Y = \Pi_H X$ is $\mathcal{G}$ measurable, and by definition of the projection

$X - Y \perp 1_G$ for arbitrary $G \in \mathcal{G}$. But this is exactly the integral test (1.2) in the definition 1.1:

$$(X - Y, 1_G)_{L^2(\mathbb{P})} = 0 \Leftrightarrow \int_G X d\mathbb{P} = \int_G Y d\mathbb{P}.$$

If X is an integrable random variable with $X \geq 0$, then we can define the conditonal expectation of X by approximating the random variable X by random varibales $X^{(n)} = X \wedge n$, and defining

$$\mathbb{E}(X|\mathcal{G}) = \lim_n \mathbb{E}(X^{(n)}|\mathcal{G}).$$

For an integrable random variable X we define the conditional expectation by the formula

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X^+|\mathcal{G}) - \mathbb{E}(X^-|\mathcal{G}).$$

1.5.1. *Properties of conditional expectations.* Let $X, Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} \subset \mathcal{F}$. Put $Y = \mathbb{E}[X|\mathcal{G}]$.
If $X \in \mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = X$. This follows from the defintion 1.1.
From the integral test (1.2) we obtain that

$$\mathbb{E}X = \int_\Omega X d\mathbb{P} = \int_\Omega Y d\mathbb{P} = \mathbb{E}Y = \mathbb{E}(\mathbb{E}[X|\mathcal{G}]).$$

Moreover, the conditional expectation satisfies

- $\mathbb{E}[aX + bZ|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Z|\mathcal{G}]$, when $a, b \in \mathbb{R}$; hence the conditional expectation is linear.
- Let $\mathcal{H} \subset \mathcal{G}$. Then
$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}].$$
This is quite obvious from the projection theorem for square integrable X.
- Let $Z \in \mathcal{G}$ and assume that $XZ \in \mathcal{L}^1$. Then
$$\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}].$$
To check this property consider first the case $Z = 1_F$, where $F \in \mathcal{G}$, and then the claim follows directly from the definition of conditional expectation. By lineriaty the claim is true for simple Z, and after this positive Z, X, and finally for arbitrary Z and X.
- If $g : \mathbb{R} \to \mathbb{R}$ is convex and we have that $g(X) \in \mathcal{L}^1$. Then the following *Jensen's inequality* holds:
$$g(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[g(X)|\mathcal{G}].$$
The proof of this claim follows from the linearity of the conditional expectation and from the fact that for any convex g there are two sequences a_n, b_n of real numbers such that
$$g(x) = \sup_n (a_n x + b_n).$$

The MCT is valid for conditional expectations:

$$0 \leq X^{(n)} \uparrow Y \in \mathcal{L}^1 \Rightarrow \mathbb{E}[X^{(n)}|\mathcal{G}] \uparrow \mathbb{E}[Y|\mathcal{G}].$$

Having this we can prove the DCT for conditional expectations: if $|X^{(n)}| \leq Z \in \mathcal{G}$, $X^{(n)} \xrightarrow{\text{a.s.}} X$, then $\mathbb{E}[X^{(n)}|\mathcal{G}] \xrightarrow{\text{a.s.}} \mathbb{E}[X|\mathcal{G}]$.

The Fatou's lemma is also true: if $X^{(n)} \geq 0$, then

$$\mathbb{E}[\liminf_n X^{(n)}|\mathcal{G}] \leq \liminf_n \mathbb{E}[X^{(n)}|\mathcal{G}].$$

The last property of conditional expectation is

- Let $X \perp\!\!\!\perp \mathcal{G}$. Then
$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X.$$
This is easy to see if $X = 1_A$, and then the claim is true for simple X, then for nonnegative X, and finally for integrable X.

2. Discrete time martingales

2.1. **Histories and martingales.** Let $\mathbb{F} = (\mathcal{F}_n)_{n\geq 1}$ a family of σ-algebras, $\mathcal{F}_n \subset \mathcal{F}$, when $n \geq 1$. The family $\mathbb{F}$ is a *history*, if is increasing : $k \leq n \Rightarrow \mathcal{F}_k \subset \mathcal{F}_n$. Sometimes an history if called a *filtration*.
Let $X = (X_n)_{n\geq 1}$ be a stochastic process: this means only that X_n is a random varibale for $n \geq 1$. Let $\mathbb{F}$ be a history. If $X_n \in \mathcal{F}_n$ for all $n \geq 1$, then we say that X is $\mathbb{F}$- *adapted*; a short notation is $X \in \mathbb{F}$.

Example 2.1. *Let $X = (X_n)_{n\geq 1}$ be a stochastic process and*

$$\mathcal{F}_n^X \doteq \sigma\{X_1, \ldots, X_n\},$$

or in plain English: $\mathcal{F}_n^X$ is the smallest σ- algebra $\mathcal{H}$ with the property that the random variables X_1, $\ldots X_n$ are measurable with respect to $\mathcal{H}$. We say that $\mathcal{F}_n^X$ is the σ- algebra generated by the random variables $X_1, \ldots, X_n$. Clearly $X_n \in \mathcal{F}_n^X$. We say that $\mathbb{F}^X$ is the history *of the process X.*

Definition 2.1. *Let $\mathbb{F}$ be a history, $X = (X_n)_{n\geq 1}$ is a stochastic process. Assume that $X \in \mathbb{F}$ and $X_n \in \mathcal{L}^1$, when $n \geq 1$.*

(a) *If in addition we have $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$, then X is a <u>martingale</u>.*
(b) *If in addition $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$, then X is <u>submartingale</u>.*
(c) *If in addition $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$, then X is <u>supermartingale</u>.*

Example 2.2 (Sums of independent random variables)**.** *Assume that $X^{(k)} \perp\!\!\!\perp$, $X^{(k)} \in \mathcal{L}^1$, when $k \geq 1$ and $\mu_k = \mathbb{E}X^{(k)}$. Put*

$$S_n = \sum_{k=1}^n X^{(k)}$$

and

$$\mathcal{F}_n^X = \sigma\{X^{(1)}, \ldots, X^{(n)}\} = \sigma\{S_1, \ldots, S_n\}.$$

Clearly $S_n \in \mathcal{F}_n$ and $S_n \in \mathcal{L}^1$. Since $X^{(n+1)} \perp\!\!\!\perp \mathcal{F}_n$, then we obtain

$$\begin{aligned} \mathbb{E}[S_{n+1}|\mathcal{F}_n] &= \mathbb{E}[S_n + X^{(n+1)}|\mathcal{F}_n] = S_n + \mathbb{E}[X^{(n+1)}|\mathcal{F}_n] \\ X^{(n+1)} \perp\!\!\!\perp \mathcal{F}_n &= S_n + \mu_{n+1}. \end{aligned}$$

We observe that S is a $\mathbb{F}$ martingale if and only if $\mu_k = 0$ for all $k \geq 1$, a submartingale if and only if $\mu_k \geq 0$ for all $k \geq 1$ and a supermartingale if and only if $\mu_k \leq 0$ for all $k \geq 1$.

Example 2.3 (products of independent random variables)**.** *Let* $X^{(k)} \perp\!\!\!\perp$, $k \geq 1$, $X^{(k)} > 0$ *and* $\mu_k = \mathbb{E}X^{(k)} > 0$, *when* $k \geq 1$. *Put* $Y_n = \prod_{k=1}^{n} X^{(k)}$ *and if*

$$\mathcal{F}_n = \sigma\{X_1, \dots, X_n\} = \sigma\{Y^{(1)}, \dots, Y^{(n)}\},$$

then $Y_n \in \mathcal{F}_n$. *Since* Y_n *is a product of independent random variables, then we have* $\mathbb{E}Y_n = \mu_1 \cdots \mu_n \in \mathcal{L}^1$. *let us compute the conditional expectation* $\mathbb{E}[Y_{n+1}|\mathcal{F}_n]$ *by using th einformation* $Y_{n+1} = Y_n X^{(n+1)}$ *and* $X^{(n+1)} \perp\!\!\!\perp \mathcal{F}_n$*:*

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] \overset{Y_n \in \mathcal{F}_n}{=} Y_n \mathbb{E}[X^{(n+1)}|\mathcal{F}_n] \overset{X^{(n+1)} \perp\!\!\!\perp \mathcal{F}_n}{=} Y_n \mu_{n+1}.$$

Hence Y *is a* $\mathbb{F}$*- martingale if and only if* $\mu_k = 1$, $k \geq 1$.

2.1.1. *Stopping times.*

Definition 2.2. *Let* $\mathbb{F} = (\mathcal{F}_k)_{k \geq 0}$ *be a history on probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. *Therandom variable* $\tau : \Omega \to \mathbb{N} \cup \{+\infty\}$ *is a* <u>*stopping time*</u>, *if*

$$\{\tau \leq k\} \in \mathcal{F}_k \quad \forall k \geq 0.$$

τ is a stopping time if and only if $\{\tau = k\} \in \mathcal{F}_k$ $\forall k \geq 0$.

Example 2.4. *Let* X *be a process and put*

$$\tau = \inf\{k : X_k \geq 3\}.$$

By convention $\tau = \infty$, *if* $X_k < 3$ *for all* $k \geq 1$. τ *is a* $\mathbb{F}^X$ *stopping time:*

$$\{\tau > n\} = \cap_{k=1}^{n} \{X_k < 3\} \in \mathcal{F}_n^X;$$

because $\mathcal{F}_n^X$ *is a* σ*-algebra, then also* $\{\tau \leq n\} \in \mathcal{F}_n^X$

13.3. 2008