FIN 508- MARTINGALE THROUGH MEASURE THEORY
LECTURE 9

MARTINGALES IN DISCRETE-TIME

In the 1950's, Doob wrote up a systemic account on the theory of martingales in his book "Stochastic Processes". Doob's book, although about 60 years old, remains very useful to researchers and still in print. The fundamental results in the martingale theory (in the restricted sense) include the optional stopping theorem, martingale inequalities and the martingale convergence theorem.

This chapter is devoted to the theory of martingales in discrete-time. We will only present the basic aspects of this subject with the emphasis on the use of filtrations (information flows), stopping times (random times) and sample paths of stochastic sequences.

In probability theory, we study probabilistic properties of random variables: properties determined by the distributions of random variables. It can be a very subtle problem to give a good description of laws of random variables taking values in infinite dimensional spaces. The classical probability deals with sequences of random variables, such as the law of large numbers, central limit theorems etc., typically starts with the assumption of independence among elements in the sequence. When we consider stochastic processes, that is, parametrized families of random variables, we will be interested in relationships between elements in the family and in particular properties determined by their (finite dimensional) joint distributions.

The basic concepts in the theory of martingales become natural and apparent as we will see, if we are allowed ourselves to use *a family of different σ algebras* on the same sample space instead one fixed collection of events, the technical used to prove deep limiting theorems, which were mastered only by few experts in the past, become systemic tools as long as we accept the notion of random times. It took some years for the probability society to digest these two fundamental ideas, and it took a generation to rewrite our textbooks on probability theory which introduce the basic theory of martingales from the very beginning.

Let us begin with the concept of *filtrations* (which model flows of information).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathbb{Z}_+ = \{0, 1, \cdots\}$ denote the ordered set of non-negative integers, and $\overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{\infty\}$.

Definition 9.1 *A family* $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$ *of sub* σ *algebras of* $\mathcal{F}$ *is called* a filtration, *if* $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ *for every* $n \in \mathbb{Z}_+$.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *together with a filtration* $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$ *is called a* filtered probability space, *denoted by* $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$.

It is useful to consider $\mathcal{F}_n$ as the information available to us up to time n.

Given a sequence of random variables $X = (X_n)_{n\in\mathbb{Z}_+}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for every n, let $\mathcal{F}_n^X$ be the smallest σ-algebra with respect to which $X_0, \cdots, X_n$ are measurable, i.e. $\mathcal{F}_n^X = \sigma\{X_m : m \le n\}$. $(\mathcal{F}_n^X)$ is called the *filtration* generated by X. A sequence of random variables $X = (X_n)_{n\in\mathbb{Z}_+}$ can be considered as the state of some random process evolving in discrete time $n = 0, 1, 2, \cdots$. For example the value of the share price of a particular company at the end of each trading day. $\mathcal{F}_n^X$ is the information about this random evolution up to time n – that is, the history of the price process. In particular, each X_n is measurable with respect to $\mathcal{F}_n^X$, i.e. $X_n \in \mathcal{F}_n^X$, so that $X = (X_n)_{n\ge 0}$ is *adapted* to the filtration $(\mathcal{F}_n^X)$, which means that as long as we reach time n, then we know the value taken by the random variable X_n at that time. Here we abuse the system of notations: which doesn't mean X_n is an element of $\mathcal{F}_n^X$, but $\{X_n \in B\} \in \mathcal{F}_n^X$ for every Borel set B, as a convention, here $\{X_n \in B\}$ is the abbreviation of $\{\omega \in \Omega : X_n(\omega) \in B\}$, and the same convention applies to similar situations.

In stochastic analysis, a stochastic process is any parameterized family of random variables valued in an arbitrary (measurable) state space. In this book, however, by a *stochastic process* we will mean a sequence of random variables (X_n), on a filtered probability space. The name "stochastic process" (stochastic derives from the Greek for random) is used to underline the fact we are more concerned with the behavior of a random sequence evolving with time n, and we are not so interested in the properties of the individual random variables, although naturally the distribution of each random variable X_n will contribute to the global and limiting behavior of the whole sequence (X_n).

Definition 9.2 *1) A sequence* $(X_n)_{n\in\mathbb{Z}_+}$ *of random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is adapted to a filtration* $(\mathcal{F}_n)$, *if for every* $n \in \mathbb{Z}_+$, X_n *is* $\mathcal{F}_n$*-measurable. In this case we say* $(X_n)_{n\in\mathbb{Z}_+}$ *is an* adapted sequence, *or* adapted process *(with respect to* $(\mathcal{F}_n)$*).*

2) If $X_0 \in \mathcal{F}_0$ and if X_n is $\mathcal{F}_{n-1}$-measurable for any $n \in \mathbb{N}$, then we say (X_n) is predictable *or* previsible.

We may think that the sample point $\omega \in \Omega$ is chosen by the fates and over time the choice is revealed to us through the values taken by the process X_n. Thus at time n the σ-algebra $\mathcal{F}_n$ contains all those sets which can be resolved, i.e. we know if ω is in them or not. That is the meaning of adaptness

For a *predictable sequence* (X_n), you know X_n before the present time n, so it is *previsible* and you can certainly predict it!

Another important concept, *stopping times* [which are random times], allows us to articulate the idea of making a decision about when to stop a process based on the observations of its past behavior. However stopping times have far-reaching applications than its superficial definition. The concept of stopping times really synthesize many important technical like random partitions, localizations etc.

Definition 9.3 *Let $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. A measurable function $T : \Omega \to \{0, 1, 2, \cdots, \infty\}$ [thus it may take value ∞] is called a* stopping time *(with respect to $(\mathcal{F}_n)$; if one wishes to emphasize the underlying filtration in question), if $\{T = n\} \in \mathcal{F}_n$ for every n.*

A stopping time T is a random variable and$\{T = \infty\} \in \mathcal{F}$. Both finite constant time $T \equiv n$ and the infinity time $T \equiv \infty$ are stopping times.

Let $\mathcal{F}_\infty = \sigma\{\mathcal{F}_n : n \in \mathbb{Z}_+\} \subset \mathcal{F}$. If T is a stopping time, then

$$\{T = \infty\} = \Omega \setminus \bigcup_{n=0}^{\infty}\{T = n\} = \bigcap_{n=0}^{\infty}\{T > n\}$$

belongs to $\mathcal{F}_\infty$, and for every n

$$\{T \le n\} = \bigcap_{k=0}^{n}\{T = k\} \in \mathcal{F}_n$$

and

$$\{T > n\} = \{T \le n\}^c \in \mathcal{F}_n$$

for every $n \in \mathbb{Z}_+$.

If S and T are two stopping time, then $S+T$, $S\vee T = \max\{S, T\}$ and $S\wedge T = \min\{S, T\}$ are stopping times too. In fact

$$\{S + T = n\} = \bigcup_{j=0}^{n}\{S = j\} \cap \{T = n - j\},$$

$$\{S \vee T = n\} = (\{S = n\} \cap \{T \le n\}) \bigcup (\{T = n\} \cap \{S \le n\})$$

and

$$\{S \wedge T \le n\} = \{S \le n\} \cap \{T \le n\}$$

belong to $\mathcal{F}_n$ for every n.

In the literature prior to the French School establishing the general theory of stochastic processes, stopping times had been called Markov times (for example, see K. Ito and H. P. J. McKean: *Diffusion Processes and Their sample Paths.* Berlin, Springer-Verlag 1965).

Example 9.4 *Let $(X_n)_{n\in\mathbb{Z}_+}$ be an adapted process on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$, and $B \in \mathcal{B}(\mathbb{R})$. Then the first time T at which the process $(X_n)_{n\in\mathbb{Z}_+}$ hits B:*

$$T = \inf\{n \geq 0 : X_n \in B\}$$

is a stopping time with respect to $(\mathcal{F}_n)$. More precisely, T is a random variable defined by

$$T(\omega) = \inf\{n \geq 0 : X_n(\omega) \in B\} \quad \forall \omega \in \Omega$$

together with the convention that $\inf \emptyset = \infty$. Hence

$$\{T = n\} = \bigcap_{k=0}^{n-1} \{X_k \in B^c\} \cap \{X_n \in B\}.$$

Since (X_n) is adapted, therefore $\{X_k \in B^c\} \in \mathcal{F}_k$ and $\{X_n \in B\} \in \mathcal{F}_n$, so that $\{T = n\} \in \mathcal{F}_n$. T is a stopping time, called a hitting time.

Hitting times are essentially the only stopping times we are interested in.

Given a stopping time T on $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$, the σ-algebra $\mathcal{F}_T$ representing the information available up to the random time T is the following σ-algebra

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \text{ s.t. } A \cap \{T \leq n\} \in \mathcal{F}_n \; \forall n = 0, 1, 2, \cdots\}.$$

Exercise 9.5 *If T is a stopping time on $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$, then $\mathcal{F}_T$ is a σ-algebra. If $T = n$ is a constant time, then $\mathcal{F}_T = \mathcal{F}_n$.*

Theorem 9.6 *Let $(X_n)_{n\in\mathbb{Z}_+}$ be an adapted random sequence on $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$, and T be a stopping time with respect to $(\mathcal{F}_n)$. Define*

$$X_T 1_{\{T<\infty\}}(\omega) = \begin{cases} X_{T(\omega)}(\omega), & \text{if } T(\omega) < \infty, \\ 0, & \text{if } T(\omega) = \infty. \end{cases}$$

Then $X_T 1_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable [In particular $X_T 1_{\{T<\infty\}}$ is a random variable.]

Proof. In fact

$$\begin{aligned} \left\{X_T 1_{\{T<\infty\}} \in G\right\} \cap \{T = n\} &= \{X_n \in G, T = n\} \\ &= \{X_n \in G\} \cap \{T = n\} \end{aligned}$$

which belongs to $\mathcal{F}_n$ for and $G \in \mathcal{B}(\mathbb{R})$ and for every $n = 0, 1, 2, \cdots$. Therefore $\left\{X_T 1_{\{T<\infty\}}\right\} \in \mathcal{F}_T$, which completes the proof. ■

Exercise 9.7 *Let $(X_n)_{n\in\mathbb{Z}_+}$ be a sequence of independent random variables with identical distribution:*

$$\mathbb{P}(X_n = 1) = p \;, \quad \mathbb{P}(X_n = 0) = 1 - p$$

where $0 < p < 1$. Let $(\mathcal{F}_n)$ be the filtration generated by (X_n), and

$$\begin{aligned} T_1 &= \inf\{n \geq 1 : X_n = 1\} \;, \\ T_{n+1} &= \inf\{T > T_n : X_n = 1\} \quad \text{if } n \geq 1 \;. \end{aligned}$$

T_n is the time that the n-th time 1 occurs in the sequence. Then each T_n is a stopping time, and the sequence

$$T_1, T_2 - T_1, \cdots, T_n - T_{n-1}, \cdots$$

is a sequence of independent, identically distributed (with a geometric distribution).

We now introduce the definition of a martingale. The word *martingale* originated in gambling, describing the double-or-quits strategy or part of a horse's harness. Mathematically it encapsulates the idea of a fair game. That is, whatever information from the past history of the game you use in order to determine your betting strategy, your expected return from playing the game is the same as your current fortune.

Definition 9.8 *Let $X = (X_n)_{n\in\mathbb{Z}_+}$ be an adapted process on a filtered probability space $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$. Suppose each X_n is integrable.*

1) X is a martingale, *if*

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] = X_n \quad a.s. \quad \forall n \in \mathbb{Z}_+.$$

2) X is a super-martingale *if*

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \leq X_n \quad a.s. \ \forall n \in \mathbb{Z}_+.$$

3) X is a sub-martingale *if*

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \geq X_n \quad a.s. \ \forall n \in \mathbb{Z}_+.$$

Exercise 9.9 *1) Prove that, an adapted, integrable random sequence (X_n) is a martingale if and only if*

$$\mathbb{E}\left[X_m|\mathcal{F}_n\right] = X_n \quad a.s. \ \forall m \geq n.$$

State a version of the statement for a super- or sub-martingale.

2) If (X_n) is a martingale, then $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for any n.

3) If (X_n) is a super-martingale, then $n \to \mathbb{E}[X_n]$ is decreasing, while $n \to \mathbb{E}[X_n]$ is increasing if (X_n) is a sub-martingale.

Example 9.10 *In these examples we are given a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$.*

1) Martingale by projection. Let $\xi \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ be an integrable random variable [i.e. $\mathbb{E}[|\xi|] < \infty$], and $X_n = \mathbb{E}[\xi|\mathcal{F}_n]$. Then (X_n) is a martingale.

2) Random walk. Let $(\xi_n)_{n\in\mathbb{Z}_+}$ be a sequence of adapted and integrable random variables. Suppose ξ_{n+1} and $\mathcal{F}_n$ are independent [i.e. $\sigma\{\xi_{n+1}\}$ and $\mathcal{F}_n$ are independent]. An example is that $\{\xi_n\}$ is a sequence of independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}_n = \sigma\{\xi_m : m \leq n\}$. Let $X_n = \sum_{k=0}^{n} \xi_k$ be the partial sum sequence. Then (X_n) is a martingale if $\mathbb{E}[\xi_n] = 0$ for any n, is a super-martingale if $\mathbb{E}[\xi_n] \leq 0$, and a sub-martingale if $\mathbb{E}[\xi_n] \geq 0$ for any n.

3) Likelihood ratios. Let f, g be two probability density functions, with support on the whole of $\mathbb{R}$. Let (X_n) be a sequence of independent, identically distributed random variables from the distribution with probability density function f. The likelihood ratio is given by

$$R_n = \frac{g(X_1)g(X_2)\dots g(X_n)}{f(X_1)f(X_2)\dots f(X_n)}$$

with $R_0 = 1$. Then (R_n) is a martingale with respect to the filtration generated by X.

4) Polya's Urn. At time $t = 0$ an urn contains 1 red and 1 black ball. At each time a ball is chosen randomly from the urn and it is then replaced along with another ball of the same color. Thus at the time of the n-th draw there are $n+2$ balls in the urn and we let B_n be the number of black balls. Then $M_n = B_n/(n+2)$ is a martingale with respect to the filtration generated by B_n.

Example 9.11 *[*Martingale transform, discrete stochastic integral*] If* (H_n) *is a* predictable *process and* (X_n) *is a martingale, then*

$$(H.X)_n = \sum_{k=1}^{n} H_k(X_k - X_{k-1}),\ (H.X)_0 = 0$$

is a martingale.

Exercise 9.12 *1) If* (X_n) *and* (Y_n) *are two martingales (resp. super-martingale), so is* $(X_n + Y_n)$.

2) Show that $(X_n \wedge Y_n)$ *is a super-martingale, where* (X_n) *and* (Y_n) *are two martingales. In fact, since* $Z_n = \min\{X_n, Y_n\}$ *so that*

$$\mathbb{E}[Z_{n+1}|\mathcal{F}_n] \le \mathbb{E}[X_{n+1}|\mathcal{F}_n] \le X_n$$

and also

$$\mathbb{E}[Z_{n+1}|\mathcal{F}_n] \le \mathbb{E}[Y_{n+1}|\mathcal{F}_n] \le Y_n$$

hence $\mathbb{E}[Z_{n+1}|\mathcal{F}_n] \le Z_n$*, thus* Z *is also a super-martingale.*

Recall Jensen's inequality for conditional expectation: if $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, ξ, $\varphi(\xi) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G}$ is a sub σ-field of $\mathcal{F}$, then

$$\varphi(\mathbb{E}[\xi|\mathcal{G}]) \le \mathbb{E}[\varphi(\xi)|\mathcal{G}].$$

Functions $(t \ln t)\, 1_{(1,\infty)}(t)$, $t^+ = t1_{(0,\infty)}$ and $|t|^p$ (for $p \ge 1$) are examples of convex functions.

Theorem 9.13 *1) Let* (X_n) *be a* martingale, *and* $\varphi : \mathbb{R} \to \mathbb{R}$ *be a convex function. Suppose* $\varphi(X_n)$ *are integrable for every* n*. Then* $\{\varphi(X_n)\}$ *is a sub-martingale.*

2) Let (X_n) *be a sub-martingale, and* $\varphi : \mathbb{R} \to \mathbb{R}$ *be* increasing *and convex. Suppose* $\varphi(X_n)$ *are integrable for every* n*, then* $\{\varphi(X_n)\}$ *is a sub-martingale.*

Proof. 1) In fact, applying Jensen's inequality

$$\begin{aligned}\varphi(X_n) &= \varphi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) \quad \text{(martingale property)}\\ &\le \mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] \quad \text{(Jensen's inequality).}\end{aligned}$$

which proved 1). The proof of 2) is similar. ∎

$t^+ = \max\{t, 0\} = t1_{(0,\infty)}$ is increasing and convex, thus, if (X_n) is a sub-martingale, so is $X_n^+ = \max\{X_n, 0\}$. If $X = (X_n)$ is a super-martingale, then $-X_n$ is a sub-martingale, so that $X_n^- = \max\{-X_n, 0\}$ is a sub-martingale. That is, the positive part of a sub-martingale is again a sub-martingale, while the *negative part of a super-martingale* is however a *sub-martingale*. Therefore, if X_n is a martingale, then both its positive part and its negative part are sub-martingales, so is its absolute value $|X_n| = X_n^+ + X_n^-$.