MATH 56A: STOCHASTIC PROCESSES
CHAPTER 5

5. Martingales

On the first day I gave the intuitive definition of "information," conditional expectation and martingale using the fair value of your place in a game. On the second day I gave you the mathematical definition of "information." On the third day I explained the mathematical definition of conditional expected value. We also discussed the definition of "integrability" and "uniform integrability" and the two theorems: Optimal Sampling Theorem and the Martingale Convergence Theorem.

5.1. **Intuitive description of martingale.** In the previous chapter we talked about optimal stopping time in a game in which the worst thing that could happen is you don't get anything. This time we are talking about a martingale: You have the opportunity to buy a "share" in a random game that someone else is playing. The game may or may not be fair. The question is: How much should you pay? This question becomes easier if you assume that you can sell your share after one round of play. So the formula or strategy should tell you how much your share of the game will be tomorrow. If we don't have any transaction fees or discount rate then the fair price you should pay today should be exactly equal to the price that you expect to sell it for tomorrow given the information that you have today.

5.1.1. *information.* We have a stochastic process X_n in discrete time n. X_n is not necessarily Markovian.

$\mathcal{F}_n$ represents all the information that you have about X_n for time $\leq n$. This is basically just $X_0, X_1, \cdots, X_n$. Suppose that we have a function

$$Y_n = f(X_0, X_1, \cdots, X_n).$$

Then, given $\mathcal{F}_n$, Y_n is known. Given $\mathcal{F}_0$, Y_n is random but $\mathbb{E}(Y_n \mid \mathcal{F}_0)$ is known. As time progresses (gets closer to n), you usually have a better idea of what Y_n might be until finally,

$$\mathbb{E}(Y_n \mid \mathcal{F}_n) = Y_n$$

Date: November 12, 2006.

5.1.2. *example: Bernoulli.* Suppose that $X_1, X_2, \cdots$, are independent identically distributed (i.i.d.) with distribution

$$X_n = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } 1-p \end{cases}$$

Let $Y_n = S_n$ be the sum:

$$Y_n = S_n = X_1 + X_2 + \cdots + X_n$$

The information at time 0 (before we flip the first coin) is $\mathcal{F}_0 : (S_0 = 0)$.

Suppose first that $p = 1/2$ Then S_n is simple random walk on $\mathbb{Z}$. The expected value of S_n changes with time. At the beginning we expect it to be zero: $\mathbb{E}(S_n \,|\, \mathcal{F}_0) = 0$. But later our estimate changes. For example,

$$\mathbb{E}(S_n \,|\, \mathcal{F}_{n-1}) = S_{n-1}$$

Why is that? Given $\mathcal{F}_{n-1}$ we know S_{n-1} but X_n is still random:

$$S_n = \underbrace{S_{n-1}}_{\text{known}} + \underbrace{X_n}_{\pm 1}$$

When $p = 1/2$ the expected value of X_n is zero: $\mathbb{E}(X_n) = 0$.

5.1.3. *expectation.* Before doing the case of general p I reviewed the definition of *expectation*:

$$\mathbb{E}(Y) := \sum_y y\mathbb{P}(Y = y) \quad \text{for discrete } Y$$

$$\mathbb{E}(Y) := \int_{-\infty}^{\infty} y f(y)\, dy \quad \text{for continuous } Y$$

So,

$$\begin{aligned} \mathbb{E}(X_n) &= 1 \cdot \mathbb{P}(X_n = 1) + (-1) \cdot \mathbb{P}(X_n = -1) \\ &= 1 \cdot p + (-1)(1-p) \\ &= p - 1 + p = 2p - 1 \end{aligned}$$

The expected value is what we use when we don't know X_n:

$$\mathbb{E}(X_n \,|\, \mathcal{F}_m) = \begin{cases} 2p-1 & \text{if } n > m \\ X_n & \text{if } n \le m \end{cases}$$

Recall that $\mathbb{E}$ is a *linear function.* So,

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = n(2p-1)$$

$$\mathbb{E}(S_n \,|\, \mathcal{F}_{n-1}) = \underbrace{\mathbb{E}(X_1 \,|\, \mathcal{F}_{n-1}) + \cdots + \mathbb{E}(X_{n-1} \,|\, \mathcal{F}_{n-1})}_{\text{not random}} + \underbrace{\mathbb{E}(X_n \,|\, \mathcal{F}_{n-1})}_{\text{random}}$$

$$= X_1 + X_2 + \cdots + X_{n-1} + 2p - 1$$

So,

$$\mathbb{E}(S_n \mid \mathcal{F}_{n-1}) = S_{n-1} + 2p - 1$$

In general,

$$\mathbb{E}(S_n \mid \mathcal{F}_m) = \begin{cases} S_m + (n-m)(2p-1) & \text{if } n > m \\ S_n & \text{if } n \le m \end{cases}$$

If $p \neq 1/2$ the value of S_n is expected to change in the future. If S_n is the payoff function you want to play this game if $p > 1/2$ and you don't want to play if $p < 1/2$.

5.1.4. *the martingale.* Continuing with the same example, let

$$M_n = X_1 + \cdots + X_n - n(2p-1) = S_n - n(2p-1)$$

This is the random number S_n minus its expected value. Then

$$\begin{aligned} \mathbb{E}(M_n \mid \mathcal{F}_m) &= \mathbb{E}(S_n \mid \mathcal{F}_m) - n(2p-1) \\ &= \begin{cases} S_m - m(2p-1) = M_m & \text{if } n > m \\ S_n - n(2p-1) = M_n & \text{if } n \le m \end{cases} \end{aligned}$$

Definition 5.1. *A sequence of random variables $M_0, M_1, \cdots$ with $\mathbb{E}(|M_i|) < \infty$ is a* martingale *with respect to $\{\mathcal{F}_n\}$ if*

$$\mathbb{E}(M_n \mid \mathcal{F}_m) = M_m$$

It follows by induction on n that this definition is equivalent to the condition:

$$\mathbb{E}(M_n \mid \mathcal{F}_{n-1}) = M_{n-1}$$

For example,

$$\mathbb{E}(M_2 \mid \mathcal{F}_0) = \mathbb{E}(\mathbb{E}(M_2 \mid \mathcal{F}_1) \mid \mathcal{F}_0) = \mathbb{E}(M_1 \mid \mathcal{F}_0) = M_0$$

(using the rule of iterated expectation)

5.2. **theory: conditional expectation with respect to information.** On the second and third days I tried to explain the mathematical definition of information as a σ-subalgebra of the σ-algebra of all events. I started with a review of basic probability.

5.2.1. *basic probability.*

Definition 5.2. *A* probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *consists of*

- $\Omega =$ *the* sample space*,*
- $\mathcal{F} =$ *the σ-algebra of all measurable subsets of Ω, (elements of $\mathcal{F}$ are called* events*) and*
- $\mathbb{P} =$ *the* probability measure *which assigns a measure $\mathbb{P}(A)$ for every $A \in \mathcal{F}$.*

The only condition is: $\mathbb{P}(\Omega) = 1$. *Note that*

$$A \in \mathcal{F} \iff \mathbb{P}(A) \text{ is defined}$$

This definition assumes the definition of "measure." "measurable" and "σ-algebra."

Definition 5.3. *A* σ-algebra *on a set* Ω *is a collection* $\mathcal{F}$ *of subsets* A *(called* measurable subsets *of* Ω*) satisfying the following axioms:*

(1) $\mathcal{F}$ *is closed under countable union. I.e., if* $A_1, A_2, \cdots$ *are measurable (elements of* $\mathcal{F}$*) then*

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

(2) *If* A *is measurable then so is its complement* $\Omega - A$. *(This implies that* $\mathcal{F}$ *is closed under countable intersection.)*

(3) $\emptyset, \Omega \in \mathcal{F}$.

A measure $\mathbb{P} : \mathcal{F} \to [0, \infty)$ *is a function which assigns to each* $A \in \mathcal{F}$ *a nonnegative real number s.t.* $\mathbb{P}$ *takes countable disjoint union to sum:*

$$\mathbb{P}\left(\coprod A_i\right) = \sum \mathbb{P}(A_i).$$

(Compare with the definition: A *topology* on Ω is a collection of subsets called *open subsets* which is closed under finite intersection and arbitrary union. The complement of an open set may not be open.)

Definition 5.4. *A function* $X : \Omega \to \mathbb{R}$ *is called* measurable *with respect to* $\mathcal{F}$ *if the inverse image of every measurable subset of* $\mathbb{R}$ *is measurable, i.e., an element of* $\mathcal{F}$. *(This is the same as saying that the inverse images of open, closed and half open intervals* (a, b), $[a, b]$, $(a, b], [a, b)$ *are measurable or, equivalently, the subset of* Ω *on which* $a < X \leq b$ *is measurable and therefore the measure*

$$\mathbb{P}(a < X \leq b)$$

is defined.) Measurable functions on Ω *are called* random variables.

(Compare with the definition: A function is *continuous* if the inverse image of every open set is open.)

5.2.2. *information.* is defined to be a σ-subalgebra of the σ-algebra $\mathcal{F}$ of all events $A \subseteq \Omega$. When the book says that $\mathcal{F}_n$ is the information given by $X_0, \cdots, X_n$ it means that $\mathcal{F}_n$ is the collection of all subsets of Ω which are given by specifying the values of $X_0, X_1, \cdots, X_n$.

A random variable Y' is $\mathcal{F}_n$*-measurable* if it can be written as a function of $X_0, X_1, \cdots, X_n$:

$$Y' = f(X_0, X_1, \cdots, X_n)$$

5.2.3. *filtration.* $\{\mathcal{F}_n\}$ is called a *filtration*. I drew the following diagrams to illustrate what that means in the case when X_1 takes 3 values and X_2 takes two values:

TABLE 1. The σ-subalgebra $\mathcal{F}_0$ has only the two required elements $\mathcal{F}_0 = \{\emptyset, \Omega\}$

X_0: known $X_0 =$ one value

Ω

TABLE 2. The σ-subalgebra $\mathcal{F}_1$ has $2^3 = 8$ elements given by the values of X_0, X_1

$$X_1 = 1, 2, 3$$

$$\mathcal{F}_1 = \{\emptyset, A, B, C, A \cup B, A \cup C, B \cup C, \Omega\}$$

A	B	C
$X_1 = 1$	$X_1 = 2$	$X_1 = 3$

Ω

TABLE 3. The σ-subalgebra $\mathcal{F}_2$ has $2^6 = 64$ elements given by the values of X_0, X_1, X_2

$$X_2 = 1, 2$$

$X_1 = 1$ $X_2 = 2$	$X_1 = 2$ $X_2 = 2$	$X_1 = 3$ $X_2 = 2$
$X_1 = 1$ $X_2 = 1$	$X_1 = 2$ $X_2 = 1$	$X_1 = 3$ $X_2 = 1$

Ω

The increasing sequence of σ-algebras

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$$

is an example of a *filtration.*

Definition 5.5. *A* filtration *is an increasing sequence of σ-subalgebras of $\mathcal{F}$.*

$$Y' \text{ is } \mathcal{F}_0 - \text{measurable} \iff Y' \text{ is constant}$$

$$Y' \text{ is } \mathcal{F}_n - \text{measurable} \iff Y' = f(X_0, X_1, \cdots, X_n)$$

5.2.4. *conditional expectation.* The definition of martingale uses conditional expectation with respect to information. This is defined mathematically by:

$\mathbb{E}(Y \mid \mathcal{F}_n) :=\ Y'$: the $\mathcal{F}_n$-measurable function which best approximates Y

In the example above, $Y' = \mathbb{E}(Y \mid \mathcal{F}_2)$ is a random variable which takes only 6 values, one for each of the 6 blocks in the third figure above. For example, in the lower left corner we have

$$Y' = \mathbb{E}(Y \mid X_1 = 1, X_2 = 2)$$

Theorem 5.6 (rule of iterated expectation)**.** *If $\mathcal{F}_n$ is a filtration and $n > m$ then*

$$\mathbb{E}(\mathbb{E}(Y \mid \mathcal{F}_n) \mid \mathcal{F}_m) = \mathbb{E}(Y \mid \mathcal{F}_m)$$

Assuming that $\mathbb{E}(|Y| \mid \mathcal{F}_m) < \infty$.

Proof. I gave the proof in the case when $n = m + 1$ assuming that X_m is given and $X = X_n$ is random. Then we have to show

$$\mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(Y)$$

The RHS is given by

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

Substituting the formula for $f_Y(y)$ in terms of the joint distribution $f(x, y)$:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

gives

$$\mathbb{E}(Y) = \iint_{\mathbb{R}^2} y f(x, y)\, dx dy$$

On the LHS we have

$$\mathbb{E}(Y \mid X = x) = \frac{\int_{-\infty}^{\infty} y f(x,y)\,dy}{f_X(x)}$$

$\mathbb{E}(\mathbb{E}(Y \mid X))$ is the expected value of this function:

$$\begin{aligned}\mathbb{E}(\mathbb{E}(Y \mid X)) &= \int_{-\infty}^{\infty} \mathbb{E}(Y \mid X = x) f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} \frac{\int_{-\infty}^{\infty} y f(x,y)\,dy}{f_X(x)} f_X(x)\,dydx \\ &= \iint_{\mathbb{R}^2} y f(x,y)\,dydx = \mathbb{E}(Y)\end{aligned}$$

assuming that $|y|f(x,y)$ has a finite integral. (This is Foubini's Theorem. You can reverse the order of integration only when the absolute value has a finite integral.) □

5.3. **Optimal sampling theorem.** The optimal sampling theorem says that, under certain conditions,

$$\mathbb{E}(M_T \mid \mathcal{F}_0) = M_0$$

where M_n is a martingale and T is a stopping time. We know that this is not always true because of the Monte Carlo gambling strategy:

5.3.1. *Monte Carlo stopping time.* This is the strategy where you stop when you win and double your bet if you lose. You can express it as a stopping time for a martingale as follows.

Suppose that $X_1, X_2, \cdots$ are independent Bernoulli variables where X_n takes values $\pm 2^{n-1}$ with equal probability. Then

$$M_n = S_n = X_1 + \cdots + X_n$$

is a martingale with $M_0 = 0$. It represents the game where you keep doubling your bet no matter what happens.

Now, let T be the first time that you win:

$$T = \inf\{n \geq 1 \mid X_n > 0\}$$

Since the simple random walk on $\mathbb{Z}$ is (null) recurrent, your probability is 1 that $T < \infty$. And when you stop, you will have $M_T = 1$. So,

$$\mathbb{E}(M_T \mid \mathcal{F}_0) = 1 \neq M_0 = 0$$

The optimal sampling theorem does not hold for Monte Carlo. So, we had better make sure that the statement excludes this case and all "similar" cases.

One way to avoid this counterexample is to put an upper bound (a time limit) on T.

Theorem 5.7 (1st OST)**.** *Suppose that M_n is a martingale. Then $\mathbb{E}(M_T) = M_0$ if T is a bounded stopping time (i.e., $T < C$).*

5.3.2. *integrability.* Now we have a bunch of theorems that assume integrability. A random variable Y is called *integrable* (or, more precisely, L^1) if $\mathbb{E}(|Y|) < \infty$. I don't remember if I got to this in class but it is in my notes:

Theorem 5.8. *Suppose that $\mathcal{F}_n$ is a filtration and Y_n is $\mathcal{F}_n$ measurable. Suppose*

1. *T is a stopping time and*
2. $\mathbb{P}(T < \infty) = 1$

Then $M_n := \mathbb{E}(Y_T \,|\, \mathcal{F}_n)$ is a martingale assuming that each M_n is integrable.

Proof. By definition we have:

$$\mathbb{E}(M_{n+1} \,|\, \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(Y_T \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n)$$

By Theorem 5.6 this is

$$= \mathbb{E}(Y_T \,|\, \mathcal{F}_n) = M_n$$

□

5.3.3. *2nd OST and uniform integrability.* The second optimal sampling theorem requires "uniform integrability."

Theorem 5.9 (2nd OST)**.** *Suppose that $M_0, M_1, \cdots$ is a martingale with respect to the filtration $\mathcal{F}_n$ and T is a stopping time. Then*

$$\mathbb{E}(M_T \,|\, \mathcal{F}_0) = M_0$$

provided that

1. $\mathbb{P}(T < \infty) = 1$
2. $\mathbb{E}(|M_T|) < \infty$ *(M_T is integrable).*
3. *$M_0, M_1, \cdots$ are uniformly integrable*

When you say that Y is integrable, you mean that the improper integral

$$\int_{-\infty}^{\infty} y f(y)\, dy = \lim_{K\to\infty} \int_{-K}^{K} y f(y)\, dy$$

converges.

Definition 5.10. *The functions $M_0, M_1, \cdots$ are* uniformly integrable *if the corresponding improper integrals for M_n converge uniformly. In other words, for every $\epsilon > 0$ there is a $K > 0$ so that*

$$\int_{-\infty}^{-K} |y| f_n(y)\, dy + \int_{K}^{\infty} |y| f_n(y)\, dy < \epsilon$$

for all $n \geq 0$ where f_n is the density function for M_n: The tails are getting smaller at the same rate.

In the book the sum of the two tails is written as a single integral:

$$\int_{-\infty}^{-K} |y| f_n(y)\, dy + \int_{K}^{\infty} |y| f_n(y)\, dy = \int_{-\infty}^{\infty} I_{|y|\geq K} |y| f(y)\, dy$$

where $I_{|y|\geq K}$ is the *indicator function* equal to 1 when the condition $(|y| \geq K)$ is true and 0 when it is false.

5.3.4. *example: random walk.* The OST can be used in reverse. If $\mathbb{E}(M_T \mid \mathcal{F}_0) \neq M_0$ then it must be because one of the conditions does not hold. I gave an example using simple random walk on $\mathbb{Z}$. You take $X_0 = 0$ and let T be the first visit to 1. Then M_n is a martingale, but

$$M_T = 1 \neq M_0 = 0$$

So, the (conclusion of) OST does not hold. Let's check the conditions:

(1) $\mathbb{P}(T < \infty) = 1$. This holds because the Markov chain is recurrent.
(2) $M_T = 1$ is constant and therefore integrable. $\mathbb{E}(|M_T|) = 1 < \infty$.

The conclusion is that the third condition must fail: $M_0, M_1, \cdots$ are *not* uniformly integrable. "The tails remain fat."

5.3.5. *example: optimal stopping time.* Suppose that $X_n \in \{1, 2, 3, 4, 5, 6\}$ and T is the 1st visit to the set $\{1, 3, 4, 5, 6\}$, i.e., this is the optimal stopping time in the game that we analyzed in the last chapter when the payoff is equal to X_n when it is > 1 and zero if you every reach 1.

Let

$$M_n = v(X_n) = \mathbb{E}(f(X_T) \mid \mathcal{F}_n)$$

Then,

(1) M_n is a martingale and
(2) Optimal sampling holds. I.e., $\mathbb{E}(M_T \mid \mathcal{F}_0) = M_0$.

In your homework you computed the value function v using the iteration algorithm which assumes that $v(X_n)$ is a Martingale.

5.4. **Martingale convergence theorem.** The other question we dealt with is: When does a martingale converge?

Theorem 5.11 (Martingale convergence theorem)**.** *Suppose that $\{M_n\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$.*

(1) *If there exists $C < \infty$ so that $\mathbb{E}(|M_n|) < C$ for all n then*

$$M_n \to M_\infty$$

where M_∞ is some integrable random variable.

(2) *If M_n are uniformly integrable and $M_n \to M_\infty$ then*

$$\mathbb{E}(M_n) \to \mathbb{E}(M_\infty)$$

5.4.1. *example: log normal distribution.* Suppose that $X_1, X_2, \cdots$ are i.i.d. where each X_i can take only two values $3/2$ and $1/2$ with equal probability:

$$\mathbb{P}(X_i = 3/2) = \frac{1}{2} = \mathbb{P}(X_i = 1/2)$$

The expected value of each X_i is

$$\mathbb{E}(X_i) = \frac{1}{2}(3/2 + 1/2) = 1$$

Let $M_0 = 1$ and $M_n = X_1X_2 \cdots X_n$ (the product). Since these are independent we have

$$\mathbb{E}(M_n) = \mathbb{E}(X_1)\mathbb{E}(X_2) \cdots \mathbb{E}(X_n) = 1$$

$$\mathbb{E}(M_{n+1} \mid \mathcal{F}_n) = \underbrace{X_1 \cdots X_n}_{M_n} \underbrace{\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)}_{1} = M_n$$

So, M_n is a martingale. Also, since $M_n > 0$ it is equal to its absolute value and

$$\mathbb{E}(|M_n|) = \mathbb{E}(M_n) = 1 \text{ is bounded}$$

Therefore, the first part of the martingale convergence theorem tells us that M_n converges to some function M_∞. But, the following calculation shows that $\mathbb{E}(M_\infty) = 0$. Therefore, the second part of the theorem tells us that M_n are not uniformly integrable.

Here is the calculation. Take the natural log of M_n:

$$\ln M_n = \sum_{i=1}^{n} \ln X_i$$

For each i,

$$\mathbb{E}(\ln X_i) = \frac{1}{2}(\ln 3/2 + \ln 1/2) \approx -.1438$$

By the strong law of large numbers we have that

$$\frac{1}{n} \ln M_n \to \mathbb{E}(\ln X_i) \approx -.1438$$

with probability one. Therefore, $\ln M_n \to -\infty$ and $M_n \to 0$ with probability one.

By the central limit theorem, $\frac{1}{n} \ln M_n$ becomes normal for large n. Then M_n becomes "log normal" which means that its logarithm is normal. For example, the size of grains of sand is distributed approximately log normally since each time it breaks the size is multiplied by a random factor.