

MATH 56A SPRING 2008
STOCHASTIC PROCESSES

KIYOSHI IGUSA

CONTENTS

5. Martingales	111
5.1. Intuitive description of martingale	111
5.2. probability theory and information	113
5.3. definition of conditional expectation	119
5.4. Optimal Sampling Theorem (OST)	124
5.5. integrability conditions	126
Homework 5	
Martingales	130

5. MARTINGALES

On the first day I gave the intuitive definition of “information,” and martingales. On the second day I gave you the mathematical definition of “information” and the proof of the “law of iterated expectation.” Later we need to discuss 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. A martingale is a random variable which you expect to have the same value tomorrow as it has today. What this amounts to is: You want to “hedge” an asset so that its value will be constant. I used a really simple example to illustrate this.

5.1.1. *intuitive definition of 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, \dots, X_n . Suppose that we have a function

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

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

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

5.1.2. *example: Bernoulli.* Suppose that X_1, X_2, \dots , 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:

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

The question is: What is the expected value of S_n given \mathcal{F}_m ?

Example Suppose that after 5 tosses of this coin we have:

$$\begin{array}{cccc|cc} & & & \mathcal{F}_3 & & \\ n = & 1 & 2 & 3 & 4 & 5 \\ X_n = & 1 & 1 & -1 & 1 & 1 \\ S_n = & 1 & 2 & 1 & 2 & 3 \end{array}$$

Then \mathcal{F}_3 consists of the information on the left of the vertical line. The conditional expected value

$$\mathbb{E}(S_3 | \mathcal{F}_3) = S_3 = 1$$

This is because, at time $n = 3$, we know the values of X_1, X_2, X_3 and their sum S_3 . It is not random. We know it is 1. Similarly,

$$\mathbb{E}(S_2 | \mathcal{F}_3) = S_2 = 2.$$

On the other hand,

$$\mathbb{E}(S_4 | \mathcal{F}_3) = S_3 + \mathbb{E}(X_4) = 1 + 2p - 1 = 2p.$$

This is because, first of all

$$\mathbb{E}(X_i) = p(1) + (1-p)(-1) = 2p - 1$$

and, at time 3,

$$S_4 = \underbrace{X_1 + X_2 + X_3}_{\text{known}} + X_4 = 1 + X_4.$$

In general,

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

5.1.3. *the martingale.* In the case when $p = 1/2$ we have $(n - m)(2p - 1) = 0$ in the above equation. So, we get:

$$\mathbb{E}(S_n | \mathcal{F}_m) = \begin{cases} S_m & \text{if } n > m \\ S_n & \text{if } n \leq m \end{cases}$$

This means that S_n is a martingale according to the following definition.

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

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

for $n \geq m$.

Continuing with the same example, let

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

This is the random number S_n minus its expected value. When $n \geq m$ we have:

$$\begin{aligned} \mathbb{E}(M_n | \mathcal{F}_m) &= \mathbb{E}(S_n | \mathcal{F}_m) - n(2p - 1) \\ &= S_m + (n - m)(2p - 1) - n(2p - 1) \\ &= S_m - m(2p - 1) = M_m \end{aligned}$$

Therefore, M_n is a martingale wrt \mathcal{F}_n .

The definition of a martingale is sometimes written as:

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

This says: With the information we have today, we expect the value of M to be the same tomorrow as it is today.

I proved that this definition is equivalent to the previous definition using induction and the “law of iterated expectation” which says

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

Proof. (\Rightarrow) Certainly the first definition implies the second since the first definition says that the value of M is expected to be the same any number of days in the future.

(\Leftarrow) We are given that

$$\mathbb{E}(M_1 | \mathcal{F}_0) = M_0,$$

$$\mathbb{E}(M_2 | \mathcal{F}_1) = M_1.$$

To go to two days we use the law of iterated expectation:

$$\begin{aligned} \mathbb{E}(M_2 | \mathcal{F}_0) &= \mathbb{E}(\underbrace{\mathbb{E}(M_2 | \mathcal{F}_1)}_{M_1} | \mathcal{F}_0) \\ &= \mathbb{E}(M_1 | \mathcal{F}_0) = M_0. \end{aligned}$$

Continue by induction to increase the 2 to any positive integer. \square

5.2. probability theory and information. I reviewed basic probability theory so that I could give the rigorous definition of “information.”

5.2.1. *basic probability.*

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

- Ω = 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 $\mathbb{P}(A) \leq 1$ for every $A \in \mathcal{F}$. I.e., \mathbb{P} is a function:

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$$

The only condition aside from the definition of “measure” and “ σ -algebra” is: $\mathbb{P}(\Omega) = 1$.

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 $A_1, A_2, \dots \in \mathcal{F}$ then

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

(2) \mathcal{F} is closed under taking complements. ($A \in \mathcal{F} \Rightarrow \Omega - A \in \mathcal{F}$.)

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

Note that, by DeMorgan's law, (1) and (2) imply:

(4) \mathcal{F} is closed under countable intersection.

The word “closed” means the operation gives another element of the same set. For example if a club is “closed under friendship” it means that: If you are a member of the club, so are all of your fiends. This set is probably either empty or includes everybody.

A *measure* $\mathbb{P} : \mathcal{F} \rightarrow [0, \infty)$ 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:

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

Definition 5.4. If \mathcal{B} is any set of subsets of Ω then the σ -algebra generated by \mathcal{B} is defined to be the intersection of all σ -algebras containing \mathcal{B} .

For example, the σ -algebra generated by the collection of all half-open intervals $(a, b]$ in the real line \mathbb{R} is the set of *Borel measurable* subsets of \mathbb{R} . In probability theory we use the Borel measurable sets instead of the Lebesgue measurable sets so that we only need to consider intervals.

Definition 5.5. A function $X : \Omega \rightarrow \mathbb{R}$ 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} . (Because we are using the Borel σ -algebra on \mathbb{R} , this is the same as saying that the inverse images of half open intervals $(a, b]$ are measurable or, equivalently, that

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

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. Albert said in class that a function is measurable if the inverse image of every open set is measurable. This is true since every open interval is a countable union of half open intervals.)

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, \dots, 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, \dots, 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 $\mathcal{F}_0 = \{\emptyset, \Omega\}$

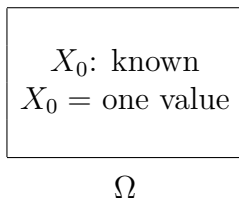
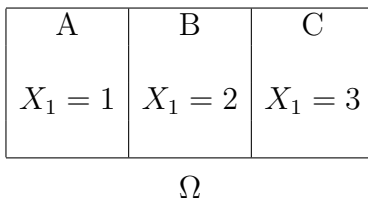
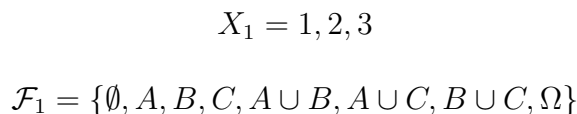


TABLE 2. The σ -subalgebra \mathcal{F}_1 has $2^3 = 8$ elements given by the values of X_0, X_1



Using Tables 2 and 3 I explained the concept of “measurability with respect to \mathcal{F}_n .” The random variable X_2 is \mathcal{F}_2 -measurable because the two sets

$$X_2^{-1}(1) := \text{the subset of } \Omega \text{ on which } X_2 = 1$$

$$X_2^{-1}(2) := \text{the subset of } \Omega \text{ on which } X_2 = 2$$

are elements of the σ -algebra \mathcal{F}_2 . ($X_2^{-1}(1)$ is the union of the top 3 boxes and $X_2^{-1}(2)$ is the union of the bottom three boxes in Table 2.) However, X_2 is not \mathcal{F}_1 -measurable since $X_2^{-1}(1)$ cuts across the three boxes in Table 1 and is therefore not one of the 8 sets listed in Table 1.

TABLE 3. The σ -subalgebra \mathcal{F}_2 has $2^6 = 64$ elements given by the values of X_0, X_1, X_2 . These are the subsets given by the 6 blocks in the table and unions of these.

$$X_2 = 1, 2$$

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

$$\Omega$$

Intuitively, “ Y is \mathcal{F}_n -measurable” means that, at time n , we will have enough information to calculate the value of Y . We need the precise mathematical definition to prove theorems about information and martingales.

In the case where the filtration \mathcal{F}_n is given by the values of the stochastic random variable X_n , Y is \mathcal{F}_n -measurable if and only if

$$Y = f(X_0, X_1, \dots, X_n)$$

where f is a measurable function. (Unlike other concepts such as continuity or differentiability, any function that you can write down is measurable.)

The increasing sequence of σ -algebras

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

is an example of a *filtration*.

Definition 5.6. A *filtration* is an increasing sequence of σ -subalgebras of \mathcal{F} .

The intuitive idea is that, as time progresses, you have more and more information. If \mathcal{F}_n were not contained in \mathcal{F}_{n+1} it would mean you will have forgotten some of the information between days n and $n + 1$.

5.2.4. *conditional expectation.* I still need to explain the definition of “conditional expectation.” I used the integral formula for conditional expectation to prove:

Theorem 5.7 (rule of iterated expectation). *If \mathcal{F}_n is a filtration and $n > m$ then*

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

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

Proof. I gave the proof of this equation in the following form:

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

Here X represents the information given by \mathcal{F}_n . I am assuming that X is one random variable. The RHS is given by

$$\mathbb{E}(Y) = \int y \mathbb{P}(y < Y \leq y + dy)$$

But

$$\mathbb{P}(y < Y \leq y + dy) = f_Y(y) dy$$

So,

$$\mathbb{E}(Y) = \int y \mathbb{P}(y < Y \leq y + dy) = \int_{-\infty}^{\infty} y f_Y(y) dy$$

On the LHS we have

$$\mathbb{E}(Y | X = x) = \int y \mathbb{P}(\underbrace{y < Y \leq y + dy}_A | \underbrace{x < X \leq x + dx}_B)$$

But

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} = \frac{f(x, y) dx dy}{f_X(x) dx}$$

where $f(x, y)$ is the joint density function. So,

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

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

$$\begin{aligned} \mathbb{E}(\mathbb{E}(Y | X)) &= \int_{-\infty}^{\infty} \mathbb{E}(Y | X = x) f_X(x) dx \\ &= \int_{-\infty}^{\infty} \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f_X(x)} f_X(x) dx \\ &= \iint_{\mathbb{R}^2} y f(x, y) dy dx \\ &= \int y \left[\int f(x, y) dx \right] dy = \int y f_Y(y) dy = \mathbb{E}(Y) \end{aligned}$$

where we switched the order of integration (Fubini's theorem) and used the identity

$$\int f(x, y) dx = f_Y(y).$$

□

5.3. definition of conditional expectation. The definition of a martingale is: If $n \geq m$ then

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

But what is the definition of this \uparrow conditional expectation?

5.3.1. *conditional expectation wrt information.*

Definition 5.8. Given: Y a measurable function with respect to $\mathcal{F}_\infty = \cup \mathcal{F}_n$. (This means we will “eventually” know the value of Y .) Then

$$Y' = \mathbb{E}(Y | \mathcal{F}_m)$$

is defined to be “the \mathcal{F}_m -measurable function which best approximates Y .”

To explain what this says I used two examples.

Example 5.9. \mathcal{F}_n is given by X_0, X_1, X_2, \dots . Then \mathcal{F}_2 is given by the table:

$\mathcal{F}_2 :$				$X_2 = 1$
				$X_2 = 2$
	$X_1 = 1$	$X_1 = 2$	$X_1 = 3$	

To say that $Y' = \mathbb{E}(Y | \mathcal{F}_2)$ is \mathcal{F}_2 -measurable means that Y' takes only 6 values, one in each of the little rectangles in the diagram. (Each little box is a subset of Ω .)

$Y' =$	y_{11}	y_{21}	y_{31}
	y_{12}	y_{22}	y_{32}

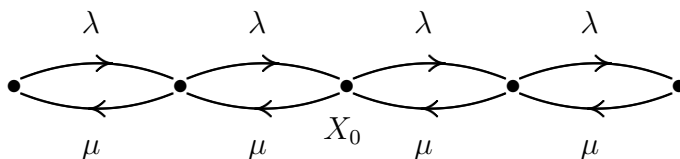
The numbers are the best guess as to the value of Y given the information at time $n = 2$:

$$y_{ij} = \mathbb{E}(Y | X_1 = i, X_2 = j)$$

The law of iterated expectation says:

$$\begin{aligned} \mathbb{E}(Y | \mathcal{F}_0) &= \mathbb{E}(\mathbb{E}(Y | \mathcal{F}_2) | \mathcal{F}_0) = \mathbb{E}(Y' | \mathcal{F}_0) \\ &= \sum_{i=1}^3 \sum_{j=1}^2 y_{ij} \mathbb{P}(X_1 = i, X_2 = j) \end{aligned}$$

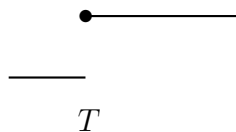
The conditional expectation with respect to information is the function which takes 6 values in the previous example. Each of these numbers is a conditional expectation with respect to an event. The next example explains this and gives the correct formula for conditional jump time (something that we needed in Chapter 3).



Example 5.10. Consider the continuous Markov chain (on the set of integers \mathbb{Z}): The rate of movement to the right is $a(n, n + 1) = \lambda$ and to the left is $a(n, n - 1) = \mu$. We want to calculate the conditional expected value of the jump time:

$$\begin{aligned} T &= \text{time of 1st jump} \\ &= \inf\{t \mid X_t \neq X_0\} \end{aligned}$$

This is a stopping time because X_t is *right continuous*: At the mo-



ment that you jump you will know:

$$X_T = \text{the state that you jump to}$$

X_T will be either to the right R or left L :

$$R = \text{event } "X_T = X_0 + 1" \text{ (jump right)}$$

$$L = \text{event } "X_T = X_0 - 1" \text{ (jump left)}$$

These two events give a σ -algebra

$$\mathcal{F}_1 = \sigma\text{-algebra generated by } R, L:$$

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

The conditional expectation of the jump time T with respect to this information is:

$$T' = \mathbb{E}(T \mid \mathcal{F}_1) = \mathcal{F}_1\text{-measurable function approximating } T$$

T' being \mathcal{F}_1 -measurable means that it takes two values:

$$T' = \begin{array}{|c|c|} \hline t_L & t_R \\ \hline \end{array}$$

The two values are:

$$t_L = \mathbb{E}(T \mid L)$$

$$t_R = \mathbb{E}(T | R)$$

We know from before (and I will compute it again below) that

$$\mathbb{E}(T | \mathcal{F}_0) = \mathbb{E}(T) = \frac{1}{\lambda + \mu}$$

I gave an intuitive proof of this by drawing a picture:

$$0 \text{ --- } R \text{ --- } L \text{ --- } L \text{ --- } R \text{ --- } R \text{ --- } 1$$

Here $\lambda = 3, \mu = 2$. Then on an average unit interval of time we will see 3 jumps to the right and 2 jumps to the left for a total of $\lambda + \mu = 5$ jumps per unit time. The average time between jumps will be

$$\frac{1}{\lambda + \mu} = \frac{1}{5}$$

The law of iterated expectation implies that this is

$$\begin{aligned} &= \mathbb{E}(\mathbb{E}(T | \mathcal{F}_1) | \mathcal{F}_0) = \mathbb{E}(T' | \mathcal{F}_0) \\ &= \mathbb{E}(T | L)\mathbb{P}(L) + \mathbb{E}(T | R)\mathbb{P}(R) \\ &= t_L \cdot \frac{\mu}{\lambda + \mu} + t_R \cdot \frac{\lambda}{\lambda + \mu} \end{aligned}$$

This means that the intuitive idea that

$$t_L = \frac{1}{\mu}, \quad t_R = \frac{1}{\lambda} \quad \text{is WRONG}$$

because it would give

$$\frac{1}{\lambda + \mu} = \frac{2}{\lambda + \mu} \quad !!$$

So, the question is: What is $\mathbb{E}(T | R)$?

5.3.2. conditional expectation wrt an event.

Definition 5.11. If A is any event then the conditional expectation of T given A is defined to be:

$$\mathbb{E}(T | A) := \frac{\mathbb{E}(T \cdot I_A)}{\mathbb{P}(A)}$$

where I_A is the *indicator function* of A . This is the function which is 1 on A and 0 outside of A . (So, $\mathbb{E}(I_A) = \mathbb{P}(A)$. This is an equation we saw before.)

The expectation of T and of $T \cdot I_A$ are given by integration:

$$\begin{aligned} \mathbb{E}(T) &= \int_0^\infty t f_T(t) dt \\ \mathbb{E}(T \cdot I_A) &= \int_0^\infty t \cdot I_A f_T(t) dt \end{aligned}$$

$$\mathbb{E}(T \cdot I_R) = \int_0^\infty t \cdot I_R f_T(t) dt = \int_0^\infty t \lambda e^{-(\lambda+\mu)t} dt = \frac{\lambda}{\lambda + \mu} \mathbb{E}(T) = \frac{\lambda}{(\lambda + \mu)^2}$$

Since the probability of jumping to the right is $\mathbb{P}(R) = \frac{\lambda}{\lambda + \mu}$ we get:

$$\mathbb{E}(T | R) = \frac{\mathbb{E}(T \cdot I_A)}{\mathbb{P}(R)} = \frac{\lambda/(\lambda + \mu)^2}{\lambda/(\lambda + \mu)} = \frac{1}{\lambda + \mu}$$

Similarly,

$$\mathbb{E}(T | L) = \frac{1}{\lambda + \mu}$$

This means that the information of which way you jump tells us nothing about T ! (The time of the jump and the direction of the jump are independent.)

We then moved on to the Optimal Sampling Theorem.

5.4. **Optimal Sampling Theorem (OST).** First I stated it a little vaguely:

Theorem 5.12. *Suppose that*

- (1) T is a stopping time
- (2) M_n is a martingale wrt the filtration \mathcal{F}_n
- (3) certain other conditions are satisfied.

Then:

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

The first thing I explained is that this statement is NOT TRUE for Monte Carlo. This is the gambling strategy in which you double your bet every time you lose. Suppose that you want to win \$100. Then you go to a casino and you bet \$100. If you lose you bet \$200. If you lose again, you bet \$400 and so on. At the end you get \$100. The probability is zero that you lose every single time. In practice this does not work since you need an unlimited supply of money. But in mathematics we don't have that problem.

To make this a martingale you do the following. Let

$$X_1, X_2, X_3, \dots$$

be i.i.d. Bernoulli random variables which are equal to ± 1 with equal probability:

$$X_i = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

In other words, we are assuming each game is fair. Then

$$\mathbb{E}(X_i) = 0.$$

Let

$$M_n = X_1 + 2X_2 + 4X_3 + \dots + 2^{n-1}X_n$$

This is the amount of money you will have at the end of n rounds of play if you bet 1 on the first game, 2 on the second, 4 on the third, etc. and keep playing regardless of whether you win or lose. To see that this is a martingale we calculate:

$$M_{n+1} = X_1 + 2X_2 + \dots + 2^{n-1}X_n + 2^n X_{n+1} = M_n + 2^n X_{n+1}$$

At time n we know the first n numbers but we don't know the last number. So,

$$\begin{aligned} \mathbb{E}(M_{n+1} | \mathcal{F}_n) &= M_n + \mathbb{E}(2^n X_{n+1}) \\ &= M_n + 2^n \mathbb{E}(X_{n+1}) = M_n + 0 = M_n \end{aligned}$$

I.e., the expected future value is the same as the known value on each day. So, this is a martingale.

T = the first time you win. Then

$$\mathbb{P}(T < \infty) = 1.$$

The argument about random walk being null recurrent actually does not apply here. I will explain on Monday what that was about. In the Monte Carlo case it is obvious that $T < \infty$ since

$$\mathbb{P}(T > n) = \frac{1}{2^n} \rightarrow 0.$$

In any case,

$$M_T = 1$$

since, at the moment you win, your net gain will be exactly 1. So,

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

In other words, the Optimal Sampling Theorem does not hold. We need to add a condition that *excludes Monte Carlo*. We also know that we cannot prove a theorem which is false. So, we need some other condition in order to prove OST. The simplest condition is boundedness:

Theorem 5.13 (OST1). *The OST holds if T is bounded, i.e., if $T \leq B$ for some constant B .*

5.5. **integrability conditions.** The OST says that

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

under “certain conditions.” These are integrability conditions which I want to explain (but just the definition).

Definition 5.14. Suppose that Y is a random variable. Then

(1) Y is *integrable* (L^1) if

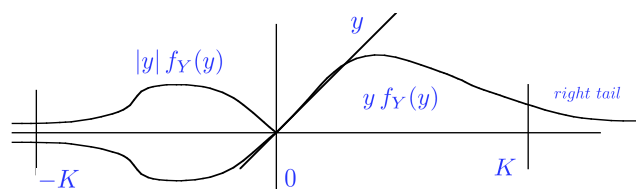
$$\mathbb{E}(|Y|) < \infty$$

I.e, if the integral

$$\int_{-\infty}^{\infty} |y| f_Y(y) dy \quad \text{converges.}$$

(2) Y is *square integrable* (L^2 , p. 217 in book) if

$$\mathbb{E}(Y^2) < \infty$$



If Y is *not* integrable then one of the tails must be fat: I.e., either the right tail

$$\int_K^{\infty} y f_Y(y) dy = \infty$$

or the left tail:

$$\int_{-\infty}^{-K} |y| f_Y(y) dy = \infty$$

If we cut off any finite piece, you still have infinity left. So, the same will be true for any value of K .

5.5.1. *a beautiful theorem.* Here is a wonderful theorem which takes longer to state than to prove and which is related to what we just learned.

Theorem 5.15. *Suppose that*

- (1) Y_n is \mathcal{F}_n -measurable,
- (2) T is a stopping time and
- (3) $\mathbb{P}(T < \infty) = 1$.

Then

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

is a martingale wrt \mathcal{F}_n .

Proof.

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

So, M_n is a martingale. □

Example 5.16. Let $Y_n = f(X_n)$ be the payoff function.

X_n = state at time n .

T = optimal stopping time. Then

$Y_T = f(X_T)$ = optimal payoff.

$v(x)$ = value function.

Then

$$v(X_n) = \mathbb{E}(\underbrace{f(X_T)}_{Y_T} | \mathcal{F}_n)$$

As an example of the theorem we just proved, we have:

Corollary 5.17. $M_n = v(X_n)$ is a martingale!

Question: Does $v(X_n)$ satisfy OST? In other words:

$$i\mathbb{E}(v(X_T) | \mathcal{F}_0) = v(X_0)?$$

Answer: Yes, because $v(X_T) = f(X_T)$. (When you reach the state X_T you are supposed to stop and take the payoff.)

5.5.2. uniform integrability.

Theorem 5.18 (2nd Optimal Sampling Theorem). *Suppose that M_0, M_1, M_2, \dots is a martingale wrt the filtration \mathcal{F}_n . Suppose*

- (1) T = stopping time
- (2) $\mathbb{P}(T < \infty) = 1$.
- (3) M_T is integrable $\mathbb{E}(|M_T|) < \infty$
- (4) M_0, M_1, \dots are uniformly integrable (defined below).

Then OST holds, i.e., $\mathbb{E}(M_T | \mathcal{F}_0) = M_0$.

Note: The contrapositive is also true. I.e., if OST fails then one of the conditions must fail.

For example, in Monte Carlo, $X_i = \pm 1$ with probability 1/2,

$$M_n = X_1 + 2X_2 + 2^2X_3 + \dots + 2^{n-1}X_n$$

is a martingale

$$T = \text{smallest } n \text{ so that } X_n = 1.$$

This is a stopping time with $\mathbb{P}(T < \infty) = 1$ and $M_T = 1$ is integrable. But OST fails. So, it must be that this martingale is not uniformly integrable.

Definition 5.19. Y_n is integrable if for every $\epsilon > 0$ there is a $K_n > 0$ so that the K_n -tails have total area less than ϵ :

$$\int_{K_n}^{\infty} y f_{Y_n}(y) dy + \int_{-\infty}^{-K_n} |y| f_{Y_n}(y) dy < \epsilon$$

Y_n is *uniformly integrable* if the cutoff points are the same for all Y_n : $K_n = K$.

If a sequence Y_n is not uniformly integrable then, as time goes on, you are very likely to end up in the tail. (No matter where you cut it the tail has probability $\geq \epsilon > 0$. But you have an infinite sequence of random variable. If they are independent you are almost certain to end up in the tail.)

Finally, I asked: Why is Monte Carlo not uniformly integrable? It is not given by an integral. So, what does this mean?

5.5.3. *nonintegral meaning of uniform integrability.* We need a new definition of tail which applies to any random variable Y_n , not just the continuous ones. For any $\delta > 0$ define a δ -tail to be a set of values of Y_n with probability $\leq \delta$. Then uniform integrability implies that: $\forall \epsilon > 0 \exists \delta > 0$ so that

$$\int_{\delta\text{-tail}} |Y_n| < \epsilon$$

for all n . (In the discrete case the integral means you add up the probability times $|Y_n|$ for all points in the tail.) In the case of Monte Carlo, regardless of δ , we can take n so that $1/2^n < \delta$. Then the event that X_1, X_2, \dots, X_n are all 1 is in the δ -tail. It has probability $1/2^n$. But $M_n = 2^n - 1$ on this tail. So,

$$\int_{\delta\text{-tail}} |M_n| \geq \frac{2^n - 1}{2^n} \approx 1$$

which will not be $< \epsilon$. So, this sequence is not uniformly integrable.

This δ -tail condition is not exactly the same uniform integrability. This will be explained at the end.

5.5.4. *Martingale convergence theorem.* I just stated this theorem without much explanation. It has two important integrality conditions.

Theorem 5.20 (Martingale convergence theorem). *Suppose that M_n is a martingale wrt the filtration \mathcal{F}_n . Then*

- (1) M_n converges to a random variable M_∞ if $\mathbb{E}(|M_n|) \leq C$ for some constant C .
- (2) $\mathbb{E}(M_n) \rightarrow \mathbb{E}(M_\infty)$ if M_n are uniformly integrable.

This ends what I said in class about martingales. What follows are some theoretical comments that I didn't have time to say. If we need them later I will go back and explain them.

It helps to know that the second conditions implies the first condition.

Lemma 5.21. *If Y_n are uniformly integrable then there is finite C so that $\mathbb{E}(|Y_n|) \leq C$ for all n .*

In fact there is the following theorem relating uniform integrability, this boundedness condition and the δ -tail interpretation.

Theorem 5.22. *A sequence of real valued random variables Y_n is uniformly integrable if and only if both of the following conditions hold.*

- (1) (uniform L^1 -boundedness) $\exists C < \infty$ s.t. $\mathbb{E}(|Y_n|) \leq C$ for all n
- (2) (δ -tail condition) $(\forall \epsilon > 0)(\exists \delta > 0)$

$$\int_{\delta\text{-tail}} |Y_n| < \epsilon$$

for all n .

5.5.5. *definition of uniform integrability.* The book gives the following definition of uniform integrability. This wording is intended to apply to all cases of real valued random variables.

Definition 5.23. A sequence or real valued random variables Y_n is *uniformly integrable* iff $(\forall \epsilon > 0)(\exists K > 0)$ so that

$$\mathbb{E}(|Y_n|I(|Y_n| > K)) \leq \epsilon$$

Where $I(|Y_n| > K)$ is the indicator function of the property $|Y_n| > K$, i.e., it is the function which is equal to 1 when $|Y_n| > K$ and 0 elsewhere. Expectation value are given by integrals for continuous random variables and sum for discrete random variables. So, this is always defined.

Proof of Lemma 5.21. Another one-line proof:

$$\mathbb{E}(|Y_n|) = \mathbb{E}(|Y_n|I(|Y_n| \leq K)) + \mathbb{E}(|Y_n|I(Y > K)) \leq 2K^2 + \epsilon.$$

□

HOMEWORK 5
MARTINGALES

These problems are due Wednesday, March 26. Answers will be posted the following week.

First problem: Suppose that X_1, X_2, \dots are i.i.d. random variables with expected value $\mu = \mathbb{E}(X_i)$.

a) Show that

$$M_n = \frac{X_1 X_2 \cdots X_n}{\mu^n}$$

is a martingale wrt X_1, X_2, \dots

b) Assuming that $X_i > 0$, is $\ln M_n$ (natural log of M_n) a martingale in general? [See next question.]

c) Modify $\ln M_n$ to make it into a martingale.

Second problem: (a special case of the first problem)

d) Suppose that (in the above problem)

$$X_i = \begin{cases} 1/2 & \text{with probability } 1/2 \\ 2 & \text{with probability } 1/2 \end{cases}$$

Then show that $T =$ first time that $M_n = 1/2$ is a stopping time. Does the Optimal Sampling Theorem hold?

e) Using the Central Limit Theorem (on $\ln M_n$), show that, with probability one, M_n converges to 0. Explain why this contradicts the Martingale convergence theorem. (What part of the statement holds? What part doesn't hold?)

Third problem Suppose that

$$X_i = \begin{cases} -1 & \text{with probability } 2/3 \\ 2 & \text{with probability } 1/3 \end{cases}$$

f) Show that

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

is a martingale wrt X_1, X_2, \dots ?

g) Show that the Optimal Sampling Theorem does not hold. [Let $T =$ first time that $S_n > 0$. Show that S_n is a recurrent Markov Chain and therefore $\mathbb{P}(T < \infty) = 1$]

h) Is S_n/\sqrt{n} a martingale wrt X_1, X_2, \dots ?