MATH 56A SPRING 2008
STOCHASTIC PROCESSES

KIYOSHI IGUSA

Contents

Date: March 12, 2008.

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 $\le n$. 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 $\mathbb{E}(Y_n \,|\, \mathcal{F}_0)$ 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 $X_1, X_2, \cdots$, 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 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:

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

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

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

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,

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

On the other hand,

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

This is because, first of all

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

and, at time 3,

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

In general,

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

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:

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

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

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

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

for $n \ge m$.

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. When $n \ge m$ we have:

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

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

The definition of a martingale is sometimes written as:

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

This say: 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

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

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

($\Leftarrow$) We are given that

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

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

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

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

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

Continue by induction to increase the 2 to any positive integer. □

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

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

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 $A_1, A_2, \cdots \in \mathcal{F}$ 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* $\mathbb{P} : \mathcal{F} \to [0, \infty)$ 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:

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

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 $X : \Omega \to \mathbb{R}$ 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 $\mathcal{F}$. (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 $X_0, \cdots, X_n$ it means that $\mathcal{F}_n$ is the collection of all subsets of Ω which are given by specifying the values of $X_0, X_1, \cdots, X_n$.

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

Using Tables 2 and 3 I explained the concept of "measurablity 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

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.

$X_2 = 1, 2$

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

Ω

boxes in Table 1 and is therefore not one of the 8 sets listed in Table 1.

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

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

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

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

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

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

But

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

So,

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

On the LHS we have

$$\mathbb{E}(Y \,|\, X = x) = \int y\mathbb{P}(\underbrace{y < Y \le y + dy}_{A} \,|\, \underbrace{x < X \le x + dx}_{B})$$

But

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

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:

$$\mathbb{E}(\mathbb{E}(Y \,|\, X)) = \int_{-\infty}^{\infty} \mathbb{E}(Y \,|\, X = x) f_X(x)\, dx$$

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

$$= \iint_{\mathbb{R}^2} y f(x, y)\, dydx$$

$$= \int y \left[\int f(x, y)\, dx \right] dy = \int y f_Y(y)\, dy = \mathbb{E}(Y)$$

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

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

$\square$