

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

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

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:

$$\mathcal{F}_2 : \begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline X_1 = 1 & X_1 = 2 & X_1 = 3 \\ \hline \end{array} \begin{array}{l} X_2 = 1 \\ X_2 = 2 \end{array}$$

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  $\Omega$ .)

$$Y' = \begin{array}{|c|c|c|} \hline y_{11} & y_{21} & y_{31} \\ \hline y_{12} & y_{22} & y_{32} \\ \hline \end{array}$$

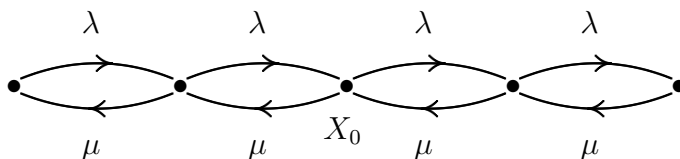
The numbers are the best guess as to the value of  $Y$  given the information at time  $n = 2$ :

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

The law of iterated expectation says:

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

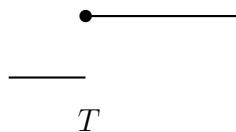
The conditional expectation with respect to information is the function which takes 6 values in the previous example. Each of these numbers is a conditional expectation with respect to an event. The next example explains this and gives the correct formula for conditional jump time (something that we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

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

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  $\sigma$ -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  $\pm 1$  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  $B$ .*