MATH 526: Discrete State Stochastic Processes
Lecture 19

Chapter 5: Martingales

Prakash Chakraborty

University of Michigan

Conditioning on events

Recall the notion of conditional probability of event A, given event B:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

where $\mathbb{P}(B) > 0$.

Example

1. Think of a probability of survival of a species in an uncertain future environment. We may not know which random environment will be realized in the future (e.g. we are not sure about the future average annual temperature), but we may still be able to estimate the conditional probability of survival of a given species (e.g. humans) given the type of environment.

2. Alternatively, a hedge fund manager may be interested in performance of her investment portfolio relative to the overall market. Then, instead of looking at the distribution of the value of the fund, she looks at its conditional distribution, given each future state of the market.

The above definition, immediately allows us to make the following observations:

- If events A and B are independent, then

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

  i.e., conditioning on an independent event does not add any new information.

- For a fixed B, the conditional probability $\mathbb{P}(\cdot \mid B)$ defines a new probability measure:

$$\tilde{\mathbb{P}}(A) = \mathbb{P}(A \mid B)$$

  for any event A. Then, we can define the expectation of a r.v. X with respect to this measure, i.e. the conditional expectation of X, given the event B:

$$\mathbb{E}(X \mid B) = \int x\tilde{\mathbb{P}}(X \in dx) = \int x\mathbb{P}(X \in dx \mid B).$$

  Assume, for simplicity, that X takes values in a discrete space. Then:

$$\mathbb{E}[X \mid B] = \sum_x x\mathbb{P}(X = x|B) = \sum_x x\frac{\mathbb{P}(X = x, B)}{\mathbb{P}(B)} = \frac{\sum_x x\mathbb{E}[\mathbf{1}_{\{X=x\}}\mathbf{1}_B]}{\mathbb{P}(B)}$$

$$= \frac{\mathbb{E}[\left(\sum_x x\mathbf{1}_{\{X=x\}}\right)\mathbf{1}_B]}{\mathbb{P}(B)} = \frac{\mathbb{E}\left[X\mathbf{1}_B\right]}{\mathbb{P}(B)}. \tag{1}$$

  The identity $\mathbb{E}[X|B] = \mathbb{E}\left[X\mathbf{1}_B\right]/\mathbb{P}(B)$ holds for non-discrete X as well.

$\mathbb{E} = \int x \, d\mathbb{P}$

- The conditional expectation of X, given B, satisfies:

$$\mathbb{E}\left[\mathbb{E}[X \mid B]\mathbf{1}_B\right] = \mathbb{E}\left[\frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}\mathbf{1}_B\right] = \frac{\mathbb{E}\left(X\mathbf{1}_B\right)}{\mathbb{P}(B)}\mathbb{P}(B) = \mathbb{E}\left(X\mathbf{1}_B\right).$$

- In addition, computing the conditional expectation of an indicator we recover the conditional probability of the corresponding event:

$$\mathbb{E}(\mathbf{1}_A \mid B) = \frac{\mathbb{E}\left(\mathbf{1}_A\mathbf{1}_B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(A \cap B\right)}{\mathbb{P}(B)} = \mathbb{P}(A \mid B)$$

  So, the conditional probability can be expressed through conditional expectation. Thus, we can forget about conditional probabilities and focus on conditional expectations.

- Recall that a r.v. X is independent of event B if and only if the random variables X and $\mathbf{1}_B$ are independent. Then, if a r.v. X is independent of event B, we have

$$\mathbb{E}(X \mid B) = \frac{\mathbb{E}\left(X\mathbf{1}_B\right)}{\mathbb{P}(B)} \overset{\text{Why?}}{=} \frac{\mathbb{E}X\mathbb{E}\mathbf{1}_B}{\mathbb{P}(B)} = \mathbb{E}X$$

Conditioning on random variables

Notation: $A^c = \Omega \setminus A$ is the compliment of A.

We know how to condition on events. It turns out that, in many cases, it is also useful to condition on random variables.

Imagine that, at the beginning, we don't know whether the event B will occur or not (it is a random event). But, even when we find it out, we, still, will not know the value of the r.v. X. Roughly speaking, X "has more randomness than B". Then, we want to find the conditional expectation of a r.v. X in both cases:

$$\mathbb{E}(X \mid B) \quad \text{and} \quad \mathbb{E}(X \mid B^c)$$

5.1: Conditional probabilities and expectations

Example. Assume that we are interested in the stock price of a company one year from now, and the quarterly report of the firm is due next month. For simplicity, think of the report as an r.v. that can take two values: good or bad. This report will, clearly, affect the future performance of the stock of this company. However, the stock value one year from now will not be known at the time when the report comes out. Hence, we need to find the conditional distribution of the stock value given all possible values of the report.

To make the above idea more formal, we introduce the conditional expectation of X, given $\mathbf{1}_B$: earlier event

r.v.

$$\mathbb{E}(X \mid \mathbf{1}_B) = f(\mathbf{1}_B) = \begin{cases} \mathbb{E}(X \mid B), & \text{if } \mathbf{1}_B = 1, \\ \mathbb{E}(X \mid B^c), & \text{if } \mathbf{1}_B = 0. \end{cases}$$

We see that the above conditional expectation is no longer a number, but rather an r.v. – a function of $\mathbf{1}_B$.

The average of this r.v. over the event B (i.e. the expectation of the product of this r.v. and the indicator of B) is the same as the average of X over B:

$$\mathbb{E}\left(\mathbb{E}\left(X \middle| \mathbf{1}_B\right)\right)\mathbf{1}_B) = \mathbb{E}\left(\mathbb{E}(X \mid B)\mathbf{1}_B\right) \underset{\text{by (1)}}{=} \mathbb{E}\left(X\mathbf{1}_B\right).$$

B occurs $\Rightarrow \mathbf{1}_B = 1$.

The same holds for $\mathbf{1}_{B^c}$ instead of $\mathbf{1}_B$. In fact, we have for any function g,

$$\mathbb{E}\left(\mathbb{E}\left(X \middle| \mathbf{1}_B\right) g\left(\mathbf{1}_B\right)\right) = \mathbb{E}\left(X g\left(\mathbf{1}_B\right)\right).$$

The above property becomes the definition of a conditional expectation with respect to arbitrary random variables.

Remark

It is not trivial to construct conditional expectations with respect to an arbitrary random variable. We have provided a natural construction of conditional expectations with respect to the indicator of an event. But what about conditioning on a general random variable?

It seems natural to define the conditional expectation of an r.v. X, given an r.v. Y, as function $f(Y)$:

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

defined for all possible values y of Y, as follows:

$$f(y) = \mathbb{E}(X \mid Y = y) = \frac{\mathbb{E}[X \mathbf{1}_{\{Y=y\}}]}{\mathbb{P}(Y = y)}.$$

This construction is, indeed, possible when Y takes values in a discrete space (like in the case considered above, where $Y = \mathbf{1}_B$). However, in general, it does not hold when Y has continuous distribution. The easiest way to see this is to notice that both numerator and denominator of the above fraction vanish if Y has continuous distribution: $\mathbb{P}(Y = y) = 0$.

We thus need a more general definition of conditional expectation.

Definition (Conditional Expectation with respect to Random Variables)

Consider a r.v. X and a finite collection of r.v.'s $(Y_1, \ldots, Y_n)$, all of them taking values in $\mathbb{R}$. We define the conditional expectation of X, given $(Y_1, \ldots, Y_n)$ as an r.v. Z, denoted

$$Z = \mathbb{E}(X \mid Y_1, \ldots, Y_n),$$

such that:

1. $Z = F(Y_1, \ldots, Y_n)$, for some (measurable) function F,
2. and the following holds

$$\mathbb{E}\left[Zg(Y_1, \ldots, Y_n)\right] = \mathbb{E}\left[Xg(Y_1, \ldots, Y_n)\right],$$

   for all (measurable) functions g, for which the expectation in the right hand side is well defined.

Question

Does the conditional expectation defined above exist and is it is unique?

Theorem

For any r.v. X, with $\mathbb{E}|X| < \infty$, and any collection of random variables $(Y_1, \ldots, Y_n)$, $\mathbb{E}[X \mid Y_1, \ldots, Y_n]$ exists uniquely.

We now show that the definition coincides with the discrete case.

Example (Conditional expectation for discrete r.v.'s) Consider a r.v. Y taking values in a finite state space $\mathbb{S} = \{1, \ldots, m\} \subset \mathbb{R}$; and an r.v. X, taking values in $\mathbb{R}$. Let us construct $\mathbb{E}[X|Y]$.

Remark

By considering a single r.v. Y we do not lose the generality. If we have several r.v.'s $(Y_1, \ldots, Y_n)$, taking values in a finite state space $\tilde{\mathbb{S}}$, then, the vector $Y = (Y_1, \ldots, Y_n)$, again, has only finite number of possible values (each value is a vector), and we define $\mathbb{S}$ to be a space of all possible vectors of length n, with entries from $\tilde{\mathbb{S}}$. The resulting state space is still finite.

Recall that the conditional expectation has to be a function of Y. Hence, it is of the form:

$$\mathbb{E}(X \,|\, Y) = f(Y) = \sum_{i=1}^{m} f(i)\mathbf{1}_{\{Y=i\}},$$

because any r.v. $Y : \Omega \to \mathbb{S}$ can be written as: $Y(\omega) = \sum_{i=1}^{m} i\mathbf{1}_{\{Y(\omega)=i\}}$.

Introducing the events $A_i = \{\omega \in \Omega \,:\, Y(\omega) = i\}$, we can write the conditional expectation in terms of these events:

$$\mathbb{E}(X \,|\, Y) = f(Y) = \sum_{i=1}^{m} f(i)\mathbf{1}_{A_i}.$$

We only need to determine the constants $f(i)$, for $i = 1, \ldots, m$. Recall the definition of conditional expectation:

$$\mathbb{E}\left[f(Y)g(Y)\right] = \mathbb{E}\left[\mathbb{E}[X \mid Y]g(Y)\right] = \mathbb{E}\left(Xg(Y)\right),$$

for all functions g.

Let us fix an arbitrary $i \in \mathcal{S}$ and check that this property holds for $g(y) = \mathbf{1}_{\{y=i\}}$. Notice that,

$$\mathbb{E}\left[f(Y)g(Y)\right] = \mathbb{E}\left(f(Y)\mathbf{1}_{A_i}\right) = \mathbb{E}\Big[\sum_{j \in \mathbb{S}} f(j)\mathbf{1}_{\{Y=j\}}\mathbf{1}_{\{Y=i\}}\Big] = f(i)\mathbb{P}(A_i).$$

Hence, $f(i)$ needs to satisfy:

$$f(i)\mathbb{P}(A_i) = \mathbb{E}\left(X\mathbf{1}_{A_i}\right), \qquad \text{for all } i = 1, \ldots, m.$$

Thus, we need to set

$$f(i) = \frac{\mathbb{E}\left(X\mathbf{1}_{A_i}\right)}{\mathbb{P}(A_i)} = \mathbb{E}(X \mid Y = i), \quad \text{for all } i = 1, \ldots, m.$$

Example. Consider two i.i.d. random variables X_1 and X_2, with $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$. Let us compute the conditional expectation of X_1, given $X_1 + X_2$:

$$\mathbb{E}\left(X_1 \,|\, X_1 + X_2\right) = \sum_{i=0}^{2} \mathbb{E}(X_1 \,|\, X_1 + X_2 = i)\mathbf{1}_{\{X_1+X_2=i\}}.$$

Finally, we compute the conditional expectations, given events $X_1 + X_2 = i$, using the formula from the beginning of this lecture:

$$\mathbb{E}(X_1 \,|\, X_1 + X_2 = i) = \frac{\mathbb{E}\left(X_1 \mathbf{1}_{\{X_1+X_2=i\}}\right)}{\mathbb{P}(X_1 + X_2 = i)} = \begin{cases} \frac{0}{1/4} = 0, & i = 0, \\ \frac{1/4}{1/2} = \frac{1}{2}, & i = 1, \\ \frac{1/4}{1/4} = 1, & i = 2. \end{cases}$$

Thus,

$$\begin{aligned} \mathbb{E}\left(X_1 \,|\, X_1 + X_2\right) &= \sum_{i=0}^{2} \mathbb{E}(X_1 \,|\, X_1 + X_2 = i)\mathbf{1}_{\{X_1+X_2=i\}} \\ &= \frac{1}{2}\mathbf{1}_{\{X_1+X_2=1\}} + \mathbf{1}_{\{X_1+X_2=2\}} \\ &= \frac{X_1 + X_2}{2}. \end{aligned}$$

Notice that, for a discrete r.v. Y, conditioning on Y is the same as conditioning on all the possible values of Y: $\{Y = i\}$. Hence, in the discrete case, conditional expectation with respect to an r.v. reduces to conditional expectations with respect to events.

When the r.v. Y, on which we condition, has continuous distribution (e.g. exponential or normal), the above conditioning on events will not work! In particular, the expression:

$$\mathbb{E}\left(X \mid Y = y\right) = \frac{\mathbb{E}\left(X\mathbf{1}_{\{Y=y\}}\right)}{\mathbb{P}(Y = y)},$$

is not well defined, as both the numerator and the denominator are zero.

This is why, in general, we may not be able to construct a conditional expectation explicitly. However, we may be able compute conditional expectations using their properties. In fact, we will rarely use the definition of conditional expectation itself, and will more often use its properties, outlined in the following theorem.

Theorem (Basic Properties of Conditional Expectation)

Assume we are given random variables X and Z, such that $\mathbb{E}|X| < \infty$ and $\mathbb{E}|Z| < \infty$, and a collection of random variables $\mathcal{F} = \{Y_1, \ldots, Y_n\}$.

1. *If $X = c$ is constant with probability one, then*

$$\mathbb{E}(X \mid \mathcal{F}) = c.$$

2. *If each Y_i is constant with probability one, then*

$$\mathbb{E}(X \mid \mathcal{F}) = \mathbb{E}X.$$

3. *Linearity:*

$$\mathbb{E}(\lambda X + \mu Z \mid \mathcal{F}) = \lambda \mathbb{E}(X \mid \mathcal{F}) + \mu \mathbb{E}(Z \mid \mathcal{F}).$$

   for any $\lambda, \mu \in \mathbb{R}$.

4. *Monotonicity: if $X \leq Z$, with probability one, then*

$$\mathbb{E}(X \mid \mathcal{F}) \leq \mathbb{E}(Z \mid \mathcal{F}), \quad \text{with probability one.}$$

Theorem (Contd...)

5. *Tower property*: consider $\mathcal{G} = (\tilde{Y}_1, \ldots, \tilde{Y}_m)$, where each $\tilde{Y}_i$ is a deterministic function of $(Y_1, \ldots, Y_n)$, for $i = 1, \ldots, m$. In other words, $\mathcal{G}$ contains less information than $\mathcal{F}$. Then, we have:

在G的信息比F少的情况下

$\tilde{Y}_1 = f_1(Y_1, \ldots, Y_n)$

$= \sum \mathbb{P}(G=i) \cdot \mathbb{E}(X \mid \mathcal{F} \in G^{-1}(i))$

$$\mathbb{E}\left(\mathbb{E}(X \mid \mathcal{F}) \mid \mathcal{G}\right) = \mathbb{E}(X \mid \mathcal{G}).$$

$= \sum \mathbb{P}(G=i) \cdot \mathbb{E}(X \mid G=i)$

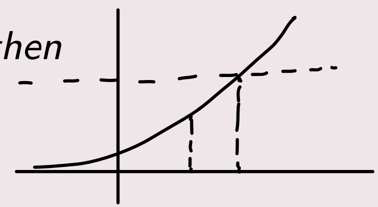
Choosing $m = 1$ and assuming that $\tilde{Y}_1$ is constant: → no information

$$\mathbb{E}\left(\mathbb{E}(X \mid \mathcal{F})\right) = \mathbb{E}X.$$

6. If Z is a function of $(Y_1, \ldots, Y_n)$, then: $\mathbb{E}(XZ \mid \mathcal{F}) = Z\mathbb{E}(X \mid \mathcal{F})$. In particular $\mathbb{E}(Z \mid \mathcal{F}) = Z$.

7. *Independence*: if X is independent of $(Y_1, \ldots, Y_n)$, then

$$\mathbb{E}(X \mid \mathcal{F}) = \mathbb{E}X.$$



8. *Jensen's inequality*: for any convex function ϕ, we have

$$\mathbb{E}\left(\phi(X) \mid \mathcal{F}\right) \geq \phi\left(\mathbb{E}\left(X \mid \mathcal{F}\right)\right) \qquad \text{(also true for non-conditional exp.)}$$

Example (MC with random initial condition) Consider a MC (X_n), with finite state space $\mathbb{S} = \{1, \ldots, m\}$ and with transition matrix p. Assume its initial condition X_0 is a random variable. Then,

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{\{X_n=j\}} \mid X_0) &= \sum_{i=1}^{m} \left[\mathbb{E}\left(\mathbf{1}_{\{X_n=j\}} \mid X_0 = i\right)\right] \mathbf{1}_{\{X_0=i\}} \\ &= \sum_{i=1}^{m} \mathbb{P}\left(X_n = j \mid X_0 = i\right) \mathbf{1}_{\{X_0=i\}} \\ &= \sum_{i=1}^{m} p^n(i,j) \mathbf{1}_{\{X_0=i\}}. \end{aligned}$$

While

$$\begin{aligned} \mathbb{P}(X_n = j) = \mathbb{E}\left(\mathbf{1}_{\{X_n=j\}}\right) &= \mathbb{E}\left[\mathbb{E}(\mathbf{1}_{\{X_n=j\}} \mid X_0)\right] \\ = \sum_{i=1}^{m} p^n(i,j) \mathbb{E}\mathbf{1}_{\{X_0=i\}} &= \sum_{i=1}^{m} p^n(i,j) q(i), \end{aligned}$$

where q is the distribution of X_0. Hence, the distribution of X_n is given by:

$$(\mathbb{P}(X_n = 1), \ldots, \mathbb{P}(X_n = m)) = qp^n.$$

Thus, using conditional expectations, we provide another derivation of the conditioning formula that we have been using all along.

Example (Random walk) Consider a discrete time stochastic process (S_n), given by $S_0 = 0$ and

$$S_n = \sum_{i=1}^{n} X_i,$$

where $\{X_i\}$ are independent identically distributed random variables, with mean μ.

Then, our best prediction for S_{n+1}, given the observations $(X_1, \ldots, X_n)$, is

$$\begin{aligned} \mathbb{E}\left(S_{n+1} \mid X_1, \ldots, X_n \right) &= \mathbb{E}\left(S_n + X_{n+1} \mid X_1, \ldots, X_n \right) \\ &= \mathbb{E}\left(S_n \mid X_1, \ldots, X_n \right) + \mathbb{E}\left(X_{n+1} \mid X_1, \ldots, X_n \right) \\ &= S_n + \mathbb{E}[X_{n+1}] \\ &= S_n + \mu. \end{aligned}$$