

1 March 3rd, 2022

For this chapter, we will start by rigorously introducing conditional expectation, and then move on to martingale theory.

1.1 Conditional Expectation

In modern probability, we first give the conditional expectation in a sigma field, and then note that other cases are special cases of this. Here we will go the opposite direction. The most general definition is not constructive, meaning we usually can't write out an explicit formula, but it still maintains special properties that are useful.

Recalling the basic level of conditional expectation, considering a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with random variable $X = (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$, the conditional expectation of X given an event $A \in \mathcal{F}$, with $\mathbb{P}(A) > 0$ is:

$$\mathbf{E}(X|A) = \frac{\mathbf{E}X \mathbb{1}_A}{\mathbb{P}(A)} = \frac{\int_A X d\mathbb{P}}{\mathbb{P}(A)}.$$

Remark 1.1 — We can write:

$$\mathbf{E}X = \frac{\mathbf{E}X \mathbb{1}_\Omega}{\mathbb{P}(\Omega)}.$$

Example 1.2

The conditional probability of B given A , is:

$$\mathbb{P}(B|A) = \mathbf{E}(\mathbb{1}_B|A) = \frac{\mathbf{E}\mathbb{1}_B \cdot \mathbb{1}_A}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Example 1.3

The conditional expectation of X given $Y = y$ (i.e. $Y^{-1}(\{y\}) \in \mathcal{F}$), is:

$$\mathbf{E}[X|Y = y] = \psi(y)$$

since it is a function of y . This leads to a second level of conditional expectation.

The conditional expectation of X given Y is:

$$\mathbf{E}[X|Y] = \psi(Y)$$

which is a random variable.

Remark 1.4 — The conditional expectation of a variable given an event is a number, the conditional expectation of a variable given a variable is a random variable.

To motivate this idea, consider the following example:

Example 1.5

Consider the probability space:

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad \mathcal{F} = 2^\Omega, \quad \mathbb{P}(\{i\}) = \frac{1}{6}.$$

Let $X(\omega) = \omega$ and $Y(\omega) = \begin{cases} 1, & \omega \in \{2, 4, 6\} \\ 0, & \text{otherwise} \end{cases}$.

- The expectation $\mathbf{E}X$ is the "best guess" of X without any information. In this case, for example we might care about the mean square error

$$\operatorname{argmin}_e \mathbf{E}(X - e)^2 = \mathbf{E}X = 3.5.$$

- The conditional expectation given event $Y = y$ ($\mathbf{E}[X|Y = y]$) is the "best guess" given info $Y = y$. For example:

$$\mathbf{E}[X|Y = 1] = \frac{\mathbf{E}X \mathbb{1}_{Y=1}}{\mathbb{P}(Y = 1)} = 2 \cdot \frac{1}{6}(2 + 4 + 6) = 4.$$

On the other hand:

$$\mathbf{E}[X|Y = 0] = \frac{\mathbf{E}X \mathbb{1}_{Y=0}}{\mathbb{P}(Y = 0)} = 2 \cdot \frac{1}{6}(1 + 3 + 5) = 3.$$

one way of thinking is we are allowed to change our guess given new information.

- The conditional expectation given Y ($\psi(Y)$) is given by:

$$\psi(Y) = 4\mathbb{1}(Y = 1) + 3\mathbb{1}(Y = 0).$$

In this case, we will still be given the information of Y , but we are not allowed to change our guess. Instead, we are allowed to have a "strategy" of future information you'll provide. In other words, this is the "best guess" based on the future information of Y .

Example 1.6

We have $\mathbf{E}[X|X] = X$, since the best guess based on getting the value of X is itself. On the other hand, $\mathbf{E}[X|Y] = \mathbf{E}X$ if $X \perp Y$, since the future information is useless.

Going beyond these two levels, we will define the conditional expectation given a σ -field $\mathcal{H} \subseteq \mathcal{F}$. If we go back to how we explained the first two examples, the "information" generated by Y is $\sigma(Y)$. In the previous example, this would be $\{\emptyset, \Omega, \{2, 4, 6\}, \{1, 3, 5\}\}$. If $\mathcal{H} = \sigma(Y)$, then we shall have:

$$\mathbf{E}[X|\sigma(Y)] = \mathbf{E}[X|Y].$$

We will define this by looking at the properties of the above definition.

- The first property we see is that $\mathbf{E}[X|Y]$ is $\sigma(Y)$ -measurable, since the information contained cannot be more than that of $\sigma(Y)$.
- $\mathbf{E}[X|Y]$ is a constant on any atoms $A \in \sigma(Y)$, since we cannot differentiate items within the atoms.

Definition 1.7 (atom of a *sigma*-field). There does not exist any nontrivial event D s.t. $D \subset A$

This also means:

$$\mathbf{E}[\mathbf{E}[X|Y]]\mathbb{1}_A = \mathbf{E}[X|Y = Y(\omega)]\mathbf{E}\mathbb{1}_A = \mathbf{E}X\mathbb{1}_A,$$

Since this is true for the atoms, by linearity of expectation, it is true for any event:

$$\mathbf{E}[\mathbf{E}[X|Y]]\mathbb{1}_E = \mathbf{E}X\mathbb{1}_E, \quad \forall E \in \sigma(Y)$$

Replacing $\sigma(Y)$ by a general \mathcal{H} , we have the following definition:

Definition 1.8. The conditional expectation of X given $\mathcal{H} \subseteq \mathcal{F}$ is defined as any random variable Z satisfying:

1. Z is \mathcal{H} -measurable
2. $\mathbf{E}Z\mathbb{1}_A = \mathbf{E}X\mathbb{1}_A$ for all $A \in \mathcal{H}$.

Example 1.9

Suppose X is \mathcal{H} -measurable. We have: $\mathbf{E}[X|\mathcal{H}] = X$.

Example 1.10

Suppose X is independent of \mathcal{H} , i.e. $X^{-1}(B) \perp A$ for all $B \in \mathcal{B}(\mathbb{R})$ and $A \in \mathcal{H}$. We have $\mathbf{E}[X|\mathcal{H}] = \mathbf{E}X$.

Thus, we arrive at the definition of conditional expectation:

Definition 1.11. $\mathbf{E}[X|Y] = \mathbf{E}[X|\sigma(Y)]$. For an event, given an event $A \in \mathcal{F}$, let $\mathcal{H} = \{\emptyset, \Omega, A, A^C\}$, we have:

$$\mathbf{E}[X|\mathcal{H}] = \frac{\mathbf{E}X\mathbb{1}_A}{\mathbb{P}(A)} \cdot \mathbb{1}_A + \frac{\mathbf{E}X\mathbb{1}_{A^C}}{\mathbb{P}(A^C)} \cdot \mathbb{1}_{A^C}$$

If we observe event A , then this reduces.

Now, we will consider the existence and uniqueness. First, for uniqueness, we will show that any two variables that satisfy these two properties are the same almost surely. Suppose we have Z and \tilde{Z} which both satisfy 1 and 2. For any $\epsilon > 0$, let:

$$A_\epsilon = \{\omega : Z(\omega) - \tilde{Z}(\omega) > \epsilon\} \in \mathcal{H}$$

by property 1. By property 2, we have:

$$0 = \int_{A_\epsilon} (Z - \tilde{Z}) d\mathbb{P} = \int_{A_\epsilon} (Z - \tilde{Z}) d\mathbb{P} \geq \epsilon \mathbb{P}(A_\epsilon) \implies \mathbb{P}(A_\epsilon) = 0 \implies Z = \tilde{Z} \text{ a.s.}$$

Now we will consider the existence.

Definition 1.12. A measure ν is absolutely continuous w.r.t μ , written as $\nu \ll \mu$ if $\mu(A) = 0 \implies \nu(A) = 0$.

Example 1.13

Consider μ be the Lebesgue measure and ν be a distribution of a continuous r.v. Then:

$$\nu(A) = \int_A f d\mu$$

meaning that $\nu \ll \mu$.

Theorem 1.14 (Radon-Nikodym Theorem)

Let μ and ν be σ -finite measures on (Ω, \mathcal{F}) . If $\nu \ll \mu$, then there is a function f which is \mathcal{F} -measurable, and for all $A \in \mathcal{F}$:

$$\nu(A) = \int_A f d\mu,$$

with f being the **Radon-Nikodum derivative** often denoted by $\frac{d\nu}{d\mu}$.

Now we will prove the existence of $\mathbf{E}[X|\mathcal{H}]$:

Proof. Let the original probability space be $(\Omega, \mathcal{F}, \mathbb{P})$. Consider $X \geq 0$. Let $\mu = \mathbb{P}$ and set:

$$\nu(A) = \int_A X d\mathbb{P} = \mathbf{E}X \mathbb{1}_A, \quad \forall A \in \mathcal{H}$$

apparently $\nu \ll \mu$. By R-N theorem, we know that $\frac{d\nu}{d\mu}$ is \mathcal{H} -measurable and:

$$\nu(A) = \int_A X d\mu = \mathbf{E}X \mathbb{1}_A$$

but also:

$$\nu(A) = \int_A \frac{d\nu}{d\mu} d\mu = \mathbf{E}[\mathbf{E}[X|\mathcal{H}]] \mathbb{1}_A$$

□

1.2 Properties of Conditional Expectation

Some of the main properties of conditional expectation can be categorized into those related to measurability, and those related to independence.

1.2.1 Pulling out independent factors

- If $X \perp \mathcal{H}$, $\mathbf{E}[X|\mathcal{H}] = \mathbf{E}X$.
- If X is independent of $\sigma(Y, \mathcal{H})$, $\mathbf{E}[XY|\mathcal{H}] = \mathbf{E}X\mathbf{E}[Y|\mathcal{H}]$. (note needs both independent of Y and \mathcal{H}).
- If $X \perp Y$, $\mathcal{G} \perp \mathcal{H}$, $X\mathcal{H}$, $X\mathcal{G}$:

$$\mathbf{E}[\mathbf{E}[XY|\mathcal{G}]|\mathcal{H}] = \mathbf{E}X\mathbf{E}Y = \mathbf{E}[\mathbf{E}[XY|\mathcal{H}]|\mathcal{G}].$$

Proof. Note that $\mathbf{E}[XY|\mathcal{G}]$ is \mathcal{G} -measurable, so $\mathbf{E}[XY|\mathcal{G}] \perp \mathcal{H}$. The proof then follow using the law of total expectation, $\mathbf{E}[\mathbf{E}[XY|\mathcal{G}]] = \mathbf{E}[XY]$. \square

1.2.2 Pulling out known factors

- If X is \mathcal{H} -measurable, $\mathbf{E}[X|\mathcal{H}] = X$.
- $\mathbf{E}[f(X)|X] = f(X)$.
- If X is \mathcal{H} -measurable, then $\mathbf{E}[XY|\mathcal{H}] = X\mathbf{E}[Y|\mathcal{H}]$.
- $\mathbf{E}[f(X)Y|X] = f(X)\mathbf{E}[Y|X]$.