

## Conditional Expectation Function

Consider the random variable  $Y_i \in \mathbb{R}$  and the random vector  $X_i \in \mathbb{R}^k$ ,  $k \geq 1$ .<sup>1</sup>

### Definition

The Conditional Expectation Function (CEF) - denoted  $E[Y_i|X_i]$  - is a **random function**. It is a function that returns the expected value of  $Y_i$  for each realized value of  $X_i$ . Since  $X_i$  is a random vector the resulting function is random itself.

If we fix  $X_i = x$ , then the value at which we are evaluating the function is no longer random. The result is a constant: the expected value of  $Y_i$  at the given  $x$ .

$$E[Y_i|X_i = x] = \int y \cdot f_Y(y|X_i = x)dy = \int y dF_Y(y|X_i = x)$$

This follows the same logic that the expectation of a random variable is,  $E[Y_i]$ , is not random.

**Discrete case.** The book devotes a lot of time to the discussion of cases where  $X_i$  is a discrete random *variable*; using the notation  $W_i \in \{0, 1\}$  or  $D_i \in \{0, 1\}$ . In this unique case, we can write the CEF as,

$$E[Y_i|D_i] = E[Y_i|D_i = 0] + D_i \cdot (E[Y_i|D_i = 1] - E[Y_i|D_i = 0])$$

The above function returns  $E[Y_i|D_i = 0]$  when  $D_i = 0$  and  $E[Y_i|D_i = 1]$  when  $D_i = 1$ . This expression for the CEF will be useful in latter chapters of the book.

### Law of Iterated Expectations

The Law of Iterated Expectations says that given two random variables<sup>2</sup>  $[Y_i, X_i]$ , we can express the unconditional expected value of  $Y_i$  as the expected value of the conditional expectation of  $Y_i$  on  $X_i$ .

$$E[Y_i] = E[E[Y_i|X_i]]$$

---

<sup>1</sup>The subscript  $i$  is not necessary here. However, this notation is consistent with the rest of the book. In this book,  $Y_i$  denotes a random variable,  $\in \mathbb{R}$ , and  $Y$  a random vector,  $\in \mathbb{R}^n$ . Likewise,  $X_i$  is a random vector,  $\in \mathbb{R}^k$ , while  $X$  will represent a random matrix,  $\in \mathbb{R}^n \times \mathbb{R}^k$ .

<sup>2</sup>This can be extended to random vectors.

Where the outside expectation is with respect to  $X_i$ ,<sup>3</sup> since the CEF is a random function of  $X_i$ . We can expand this as follows,

$$E[Y_i] = \int t \cdot f_{Y_i}(t)dt = \int \int y \cdot f_{Y_i|X}(y|x)dyf_X(x)dx = E[E[Y_i|X_i]]$$

**Example 0.1.** Suppose  $Y_i$  and  $X_i$  are both discrete,  $Y_i \in \{1, 2\}$  and  $X_i \in \{3, 4\}$ , with the joint distribution:

Table 1:  $f_{Y,X}$

	$X_i = 3$	$X_i = 2$
$Y_i = 1$	<b>1/10</b>	<b>3/10</b>
$Y_i = 2$	<b>2/10</b>	<b>4/10</b>

We can then define the two marginal distributions,

Table 2:  $f_Y$

$Y_i = 1$	$Y_i = 2$
<b>4/10</b>	<b>6/10</b>

and,

Table 3:  $f_X$

$X_i = 3$	$X_i = 4$
<b>3/10</b>	<b>7/10</b>

Likewise, we know the conditional distribution  $f_{Y|X}$ ; which we get by dividing the joint distribution by the marginal distribution of  $X_i$ . Each column of the conditional distribution should add up to 1.

Table 4:  $f_{Y|X}$

	$X_i = 3$	$X_i = 4$
$Y_i = 1$	<b>1/3</b>	<b>3/7</b>
$Y_i = 2$	<b>2/3</b>	<b>4/7</b>

Now we can calculate the following objects:

<sup>3</sup>Some texts use the notation  $E_X[E[Y_i|X_i]]$  to demonstrate that the outside expectation is with respect to  $X_i$ .

1.  $E[Y_i]$

$$\begin{aligned} E[Y_i] &= 1 \cdot Pr(Y_i = 1) + 2 \cdot Pr(Y_i = 2) \\ &= 1 \cdot 4/10 + 2 \cdot 6/10 \\ &= 16/10 \end{aligned}$$

2.  $E[Y_i|X_i = 3]$

$$\begin{aligned} E[Y_i|X_i = 3] &= 1 \cdot Pr(Y_i = 1|X_i = 3) + 2 \cdot Pr(Y_i = 2|X_i = 3) \\ &= 1 \cdot 1/3 + 2 \cdot 2/3 \\ &= 5/3 \end{aligned}$$

3.  $E[Y_i|X_i = 4]$

$$\begin{aligned} E[Y_i|X_i = 4] &= 1 \cdot Pr(Y_i = 1|X_i = 4) + 2 \cdot Pr(Y_i = 2|X_i = 4) \\ &= 1 \cdot 3/7 + 2 \cdot 4/7 \\ &= 11/7 \end{aligned}$$

4.  $E[E[Y_i|X_i]]$

$$\begin{aligned} E[E[Y_i|X_i]] &= E[Y_i|X_i = 3] \cdot Pr(X_i = 3) + E[Y_i|X_i = 4] \cdot Pr(X_i = 4) \\ &= 5/3 \cdot 3/10 + 11/7 \cdot 7/10 \\ &= 16/10 \end{aligned}$$

We have therefore demonstrated the law of iterated expectations.

We can extend this principle to conditional expectations. Suppose you have three random variables/vectors  $\{Y_i, X_i, Z_i\}$ , we can express the conditional expected value of  $Y_i$  on  $X_i$  as the (conditional) expected value of the conditional expectation of  $Y_i$  on  $X_i$  and  $Z_i$ .

$$E[Y_i|X_i] = E[E[Y_i|X_i, Z_i]|X_i]$$

Here the outside expectation is with respect  $Z_i$  conditional on  $X_i$ . It utilizes the conditional distribution  $f_{Z|X}$  to form the outside expectation,

$$E[Y_i|X_i] = \int y \cdot f_{Y|X}(y|X_i) dt = \int \int y \cdot f_{Y|X,Z}(y|X_i, z) dy f_{Z|X}(z|X_i) dz = E[E[Y_i|X_i, Z_i]|X_i]$$

## Properties of the CEF

The following three theorems can be found in a range of Econometrics textbooks and Microeconomics texts, including MM & MHE

**Theorem 0.1.** *We can express the observed outcome  $Y_i$  as a sum of  $E[Y_i|X_i] + \varepsilon_i$  where  $E[\varepsilon_i|X_i] = 0$  (i.e., mean independent).*

*Proof.*

1.  $E[\varepsilon_i|X_i] = E[Y_i - E[Y_i|X_i]|X_i] = E[Y_i|X_i] - E[Y_i|X_i] = 0$
2.  $E[h(X_i)\varepsilon_i] = E[h(X_i)E[\varepsilon_i|X_i]] = E[h(X_i) \times 0] = 0$

□

**Theorem 0.2.**  *$E[Y_i|X_i]$  is the best predictor of  $Y_i$ .*

*Proof.*

$$\begin{aligned}(Y_i - m(X_i))^2 &= ((Y_i - E[Y_i|X_i]) + (E[Y_i|X_i] - m(X_i)))^2 \\ &= (Y_i - E[Y_i|X_i])^2 + (E[Y_i|X_i] - m(X_i))^2 \\ &\quad + 2(Y_i - E[Y_i|X_i]) \times (E[Y_i|X_i] - m(X_i))\end{aligned}$$

The last term (cross product) is mean zero. Thus, the function is minimized by setting  $m(X_i) = E[Y_i|X_i]$ . □

**Theorem 0.3.** *[ANOVA Theorem] The variance of  $Y_i$  can be decomposed as  $V(E[Y_i|X_i]) + E(V(Y_i|X_i))$*

*Proof.*

$$\begin{aligned}V(Y_i) &= V(E[Y_i|X_i] + \varepsilon_i) \\ &= V(E[Y_i|X_i]) + V(\varepsilon_i) \\ &= V(E[Y_i|X_i]) + E[\varepsilon_i^2]\end{aligned}$$

The second line follows from Theorem 1.1 (independence) and

$$E[\varepsilon_i^2] = E[E[\varepsilon_i^2|X_i]] = E[V(Y_i|X_i)]$$

□