Theorem 2.2.14. *Linearity of Conditional Expectations*
For random variables X and Y, if g and h are functions of X, then $\forall x \in \text{Supp}[X]$,

$$\text{E}\big[g(X)Y + h(X)\,\big|\,X = x\big] = g(x)\text{E}[Y|X = x] + h(x).$$

Proof: Let X and Y be either discrete random variables with joint PMF f or jointly continuous random variables with joint PDF f, and let g and h be functions of X. If X and Y are discrete, then by Theorem 2.2.11, $\forall x \in \text{Supp}[X]$,

$$\begin{aligned}
\text{E}\big[g(X)Y + h(X)\,\big|\,X = x\big] &= \sum_y \big(g(x)y + h(x)\big) f_{Y|X}(y|x) \\
&= g(x)\sum_y y f_{Y|X}(y|x) + h(x)\sum_y f_{Y|X}(y|x) \\
&= g(x)\text{E}[Y|X = x] + h(x)\cdot 1 \\
&= g(x)\text{E}[Y|X = x] + h(x).
\end{aligned}$$

Likewise, if X and Y are jointly continuous, then by Theorem 2.2.11, $\forall x \in \text{Supp}[X]$,

$$\begin{aligned}
\text{E}\big[g(X)Y + h(X)\,\big|\,X = x\big] &= \int_{-\infty}^{\infty} \big(g(x)y + h(x)\big) f_{Y|X}(y|x)dy \\
&= g(x)\int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy + h(x)\int_{-\infty}^{\infty} f_{Y|X}(y|x)dy \\
&= g(x)\text{E}[Y|X = x] + h(x)\cdot 1 \\
&= g(x)\text{E}[Y|X = x] + h(x). \quad \square
\end{aligned}$$

We now turn to a central definition of this book: the *conditional expectation function (CEF)*. The CEF is a function that takes as an input x and returns the conditional expectation of Y given $X = x$. The CEF is extremely useful, since it is a single function that characterizes all possible values of $\text{E}[Y|X = x]$. If we are interested in characterizing the way in which the conditional distribution of Y depends on the value of X, the CEF is a natural summary feature of the joint distribution to target. Furthermore, we will see that the CEF can be closely linked to many topics that this book considers, including regression (Chapter 4), missing data (Chapter 6), and causal inference (Chapter 7).

Definition 2.2.15. *Conditional Expectation Function (CEF)*
For random variables X and Y with joint PMF/PDF f, the *conditional expectation function* of Y given $X = x$ is

$$G_Y(x) = \text{E}[Y|X = x], \forall x \in \text{Supp}[X].$$

A few remarks on notation are in order here. We will generally write $\mathrm{E}[Y|X=x]$ to denote the CEF rather than $G_Y(x)$. The above definition is merely meant to emphasize that, when we use the term CEF, we are referring to the *function* that maps x to $\mathrm{E}[Y|X=x]$, rather than the value of $\mathrm{E}[Y|X=x]$ at some specific x. It is also intended to clarify that the CEF, $\mathrm{E}[Y|X=x]$, is a univariate function of x. It is *not* a function of the random variable Y. So, for example, if X is a Bernoulli random variable, then the CEF of Y given X is

$$G_Y(x) = \mathrm{E}[Y|X=x] = \begin{cases} \mathrm{E}[Y|X=0] & : \quad x=0 \\ \mathrm{E}[Y|X=1] & : \quad x=1. \end{cases}$$

$G_Y(X)$ is a function of the random variable X and is therefore itself a random variable[14] whose value depends on the value of X. That is, when X takes on the value x, the random variable $G_Y(X)$ takes on the value $G_Y(x) = \mathrm{E}[Y|X=x]$.

We write $\mathrm{E}[Y|X]$ to denote $G_Y(X)$ (since $\mathrm{E}[Y|X=X]$ would be confusing). So, for example, if X is a Bernoulli random variable with $p=\frac{1}{2}$, then $\mathrm{E}[Y|X]$ is a random variable that takes on the value of $\mathrm{E}[Y|X=0]$ with probability $\frac{1}{2}$ and the value of $\mathrm{E}[Y|X=1]$ with probability $\frac{1}{2}$. Also note that we can analogously define the *conditional variance function*, $H_Y(x) = \mathrm{V}[Y|X=x]$, which we will generally write as $\mathrm{V}[Y|X=x]$, and write $\mathrm{V}[Y|X]$ to denote the random variable $H_Y(X)$.

Theorem 1.3.2 (*Equality of Functions of a Random Variable*) implies that we can write statements about conditional expectation functions in the more compact $\mathrm{E}[Y|X]$ form. For example, Theorem 2.2.14 can equivalently be stated as follows: for random variables X and Y, if g and h are functions of X, then

$$\mathrm{E}\big[g(X)Y + h(X)\,\big|\,X\big] = g(X)\mathrm{E}[Y|X] + h(X).$$

From this point on, we will generally state theorems involving conditional expectation functions using this more concise notation.

We now illustrate these concepts with our familiar example of flipping a coin and rolling a four- or six-sided die.

Example 2.2.16. *Flipping a Coin and Rolling a Die*
Consider, again, the generative process from Example 1.1.11. Recall from Example 1.3.5 that the conditional PMF of Y given $X=x$ is

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{4} & : \quad x=0, y \in \{1,2,3,4\} \\ \frac{1}{6} & : \quad x=1, y \in \{1,2,3,4,5,6\} \\ 0 & : \quad \text{otherwise}. \end{cases}$$

14 Ignoring some unimportant measure-theoretic regularity conditions, namely $X(\Omega) = \mathrm{Supp}[X]$.