# *Measure Theoretic Conditional Expectation in an Elementary Setting*

**Author:** *R. Scott McIntire*

**Date:** Sep 6, 2024

---

## Overview

The measure theoretic approach to conditional expectation can be confusing when compared to the traditional approach – especially in a discrete setting. In what follows we go through a conditional expectation problem within a discrete and familiar setting in an attempt to reduce this confusion. In the example shown, we show explicitly that the conditional expectation function is non-measurable in the origin measure space.

## Elementary Probability Example Using Measure Theory

Let $X = \{D_1, D_2, D_3, D_4, D_5, D_6\}$ and define a function $P$ by $P(D_i) = \frac{1}{6}$, for $i \in \{1, 2, 3, 4, 5, 6\}$. The intent is that $P$ will become a probability measure for the space we construct. Let $\mathcal{E} = 2^X$ be the sigma algebra consisting of the power set of $X$. We extend $P$ for every element in the sigma algebra. Since the sigma algebra consists of all sets we need an assignment for an arbitrary set, $A$. The assignment is $P(A) = \frac{|A|}{6}$; that is the cardinality of the set divided by 6. We now have a measure space; in fact, a probability space: $(P, X, \mathcal{E})$. Note that for a probability space we need an event space, $X$, a sigma algebra of sets (in the discrete case just an algebra), and a function $P$ which takes elements of the sigma algebra to [0,1] with the property that

$$P\big(\bigcup_{i=1}^{N} A_i\big) = \sum_{i=1}^{N} P(A_i) \quad \text{when } A_k \cap A_j = \emptyset \quad k \neq j$$

We now consider a random variable from which we will get a sub-sigma algebra. Let

$$g(D_i) = \begin{cases} 0 & \text{if } i \text{ is even;} \\ 1 & \text{if } i \text{ is odd} \end{cases}$$

In a discrete space the sigma algebra generated from $g$ is the algebra of sets generated from the sets: $g^{-1}(0), g^{-1}(1), g^{-1}(a)$, for $a \neq 0, 1$. It is not too hard to see that $\mathcal{F} = \{\emptyset, \{1, 3\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$.

For a discrete space a function, $f$, is measurable with respect to a sigma algebra if $f^{-1}(a)$ is an element in the sigma algebra for all $a \in (-\infty, \infty)$. This has implications for the sigma algebra $\mathcal{F}$.

The claim is that any function which is measurable over $\mathcal{F}$ has the property that $f$ is constant on the sets $\{1, 3\}$ and $\{2, 4, 6\}$. More generally, we claim that any measurable function $f$ in $\mathcal{F}$ must be constant on the *minimal* elements of the algebra – elements which have no non-trivial subsets.†

---

† By non-trivial subset we mean a subset that is not the empty set.

To see this suppose that $f(1)$ differs from $f(3)$. Then $\{1,3\} \cap f^{-1}(f(1))$ is a non trivial subset of $\{1,3\}$ as is $\{1,3\} \cap f^{-1}(f(3))$. These two sets must differ since $f(1)$ and $f(3)$ differ. This means that they are respectively $\{1\}$ and $\{3\}$. Since $\mathcal{F}$ is a sigma algebra and $\{1,3\}$, $F^{-1}(f(1))$, and $f^{-1}(f(3))$ are sets in $\mathcal{F}$, any intersection of these sets is also in $\mathcal{F}$. But we have that $\{1\}$ and $\{3\}$ are sets that result from such intersections and consequently must be in $\mathcal{F}$. However, we know that they also are *not* in $\mathcal{F}$ – contradiction. Therefore, our premise that $f(1)$ and $f(3)$ could take differing values is incorrect. Using the same argument one can show that $f$ is constant on the other minimal set $\{2,4,6\}$.

Notice that while any function over the measure space $(P, X, \mathcal{E})$ is measurable, we can write down a specific function that is non-measurable with respect to $\mathcal{F}$. We know that all we have to do is come up with a function that differs on either of the sets $\{1,3\}$ or $\{2,4,6\}$. For instance, the function: $f(D_i) = i$, for $i \in \{1,2,3,4,5,6\}$, is a non-measurable function in $\mathcal{F}$.

# Conditional Expectation

Given a probability space $(P, X, \mathcal{E})$, the conditional expectation of a measurable function $f$ with respect to a sub-sigma algebra $\mathcal{F}$ is the unique $\mathcal{F}$ measurable function (random variable) labeled, $\mathbb{E}[f/\mathcal{F}]$, such that

$$\int_{\Lambda} \mathbb{E}[f/\mathcal{F}]\, dP = \int_{\Lambda} f\, dP \quad \forall \Lambda \in \mathcal{F} \tag{1}$$

That is, $\mathbb{E}[f/\mathcal{F}]$ is a measurable function in the probability space $(P, X, \mathcal{F})$. Although it seems that $f$ itself satisfies this equation you have to be careful. The function we are looking for must be measurable with respect to $\mathcal{F}$, and since $\mathcal{F}$ is a sub-sigma algebra of $\mathcal{E}$, it is quite possible that $f$ is not $\mathcal{F}$ measurable. However, if $f$ is measurable with respect to $\mathcal{F}$ then it is its own conditional expectation. We have the following basic facts about conditional expectation:

- $\mathbb{E}[f/\mathcal{F}] = f$ when $f$ is measurable with respect to $\mathcal{F}$;
- $\mathbb{E}[\mathbb{E}[f/\mathcal{F}]/\mathcal{F}] = \mathbb{E}[f/\mathcal{F}]$;
- $\mathbb{E}[(f - \mathbb{E}[f/\mathcal{F}])/\mathcal{F}] = 0$.

Consider the function of the last section, $f(D_i) = i$, which is measurable in the space $(P, X, \mathcal{E})$. Let $\mathcal{F}$ be the sub-sigma algebra of the last section. We now compute the conditional expectation of $f$ with respect to $\mathcal{F}$. Using (1) we choose two $\Lambda$'s: $\Lambda_1 = \{1,3\}$ and $\Lambda_2 = \{2,4,6\}$. We have

$$\int_{\Lambda_1} \mathbb{E}[f/\mathcal{F}]\, dP = \int_{\Lambda_1} f\, dP \tag{2}$$

and

$$\int_{\Lambda_2} \mathbb{E}[f/\mathcal{F}]\, dP = \int_{\Lambda_2} f\, dP \tag{3}$$

We know $\mathbb{E}[f/\mathcal{F}]$ is constant on $\Lambda_1$. We label this value as $C_{\Lambda_1}$. From (2) we have

$$\begin{aligned}
\int_{\Lambda_1} \mathbb{E}[f/\mathcal{F}]\, dP &= \int_{\Lambda_1} f\, dP \\
C_{\Lambda_1} \int_{\Lambda_1} dP &= \int_{\Lambda_1} f\, dP \\
C_{\Lambda_1} * P(\Lambda_1) &= \int_{\Lambda_1} f\, dP = f(D_1) * P(D_1) + f(D_3) * P(D_3) \\
C_{\Lambda_1} * P(\Lambda_1) &= 1 * \frac{1}{6} + 3 * \frac{1}{6} = \frac{2}{3}
\end{aligned}$$

Since $P(\Lambda_1) = \frac{2}{6} = \frac{1}{3}$, we have that

$$C_{\Lambda_1} = 2$$

Consequently,

$$\mathbb{E}\left[f/\mathcal{F}\right](D_1) = \mathbb{E}\left[f/\mathcal{F}\right](D_3) = C_{\Lambda_1} = 2$$

Likewise, $\mathbb{E}\left[f/\mathcal{F}\right]$ is constant on the set $\Lambda_2$. As we did above, we find the value of $\mathbb{E}\left[f/\mathcal{F}\right]$ on the set $\Lambda_2$. Let $C_{\Lambda_2}$ be the constant value of $\mathbb{E}\left[f/\mathcal{F}\right]$ on $\Lambda_2$. We have

$$\begin{aligned}
\int_{\Lambda_2} \mathbb{E}\left[f/\mathcal{F}\right]\, dP &= \int_{\Lambda_2} f\, dP \\
C_{\Lambda_2} \int_{\Lambda_2} dP &= \int_{\Lambda_2} f\, dP \qquad (\dagger) \\
C_{\Lambda_2} * P(\Lambda_2) &= \int_{\Lambda_2} f\, dP = f(D_2) * P(D_2) + f(D_4) * P(D_4) + f(D_6) * P(6) \\
C_{\Lambda_2} * P(\Lambda_2) &= 2 * \frac{1}{6} + 4 * \frac{1}{6} + 6\frac{1}{6} = 2
\end{aligned}$$

Since $P(\Lambda_2) = \frac{3}{6} = \frac{1}{2}$, we have that

$$C_{\Lambda_2} = 4$$

Consequently,

$$\mathbb{E}\left[f/\mathcal{F}\right](D_2) = \mathbb{E}\left[f/\mathcal{F}\right](D_4) = \mathbb{E}\left[f/\mathcal{F}\right](D_6) = C_{\Lambda_2} = 4$$

Since a function is determined once we know what its values are on every $x \in X$, we have found the conditional expectation function, $\mathbb{E}\left[f/\mathcal{F}\right]$, as we know its values on every $x \in X$.

From equation (†) we see, that the value of the function $(f/\mathcal{F})$ evaluated on any 'x' value in a minimal set, $\Lambda$, is

$$\mathbb{E}\left[f/\mathcal{F}\right](x) = \frac{\int_{\Lambda} f\, dP}{P(\Lambda)}$$

More generally, discrete probability or otherwise, we may think of the sigma algebra, $\mathcal{F}$, as a coarser "mesh" than the sigma algebra, $\mathcal{E}$. And we can think of the value of $\mathbb{E}\left[f/\mathcal{F}\right]$ on any "minimal" element of the mesh as the average of the function, $f$, over this minimal element with respect to the finer "mesh" – the sigma algebra, $\mathcal{E}$.

You can think of the conditional expectation as giving the "best" representation of a function given a cruder mesh. Just as the "best" representation of an image when using a larger pixels would be an average of the more refined image over the cruder mesh