Measure Theoretic Conditional Expectation in an Elementary Setting

Author: *R. Scott McIntire*

Date: Sep 25, 2024

Overview

The measure theoretic approach to conditional expectation can be confusing when compared to the traditional approach. 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

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 σ-algebra consisting of the power set of X. We extend P for every element in the σ-algebra. Since the σ-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 σ-algebra of sets (in the discrete case just an algebra), and a function P which takes elements of the σ-algebra to [0,1] with the following properties:

- $P(X) = 1$
- $P(\emptyset) = 0.$
- $P\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{i=1}^{N} P(A_i) \quad \text{when } A_k \cap A_j = \emptyset \quad k \neq j;$

We now consider a random variable from which we will get a sub σ-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 σ-algebra generated from g is the algebra of sets generated from the sets: $g^{-1}(a)$, $a \in (-\infty, \infty)$. Clearly, the interesting sets come from the values 0 and 1, all other values lead to the empty set. Consequently,

$$\mathcal{F} = \{\emptyset, \{D_1, D_3, D_5\}, \{D_2, D_4, D_6\}, \{D_1, D_2, D_3, D_4, D_5, D_6\}\}$$

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

claim: Any function, f, which is measurable over $\mathcal{F}$ has the property that f is constant on the sets $\{D_1, D_3, D_5\}$ and $\{D_2, D_4, D_6\}$. More generally, we claim that any measurable function, f, in a given σ-

algebra must be constant on the *minimal* elements of the algebra – elements which have no non-trivial subsets.†

To see this suppose that $f(1)$ differs from $f(3)$. Then $\{D_1, D_3, D_5\} \cap f^{-1}(f(D_1))$ is a non trivial subset of $\{D_1, D_3, D_5\}$ as is $\{D_1, D_3, D_5\} \cap f^{-1}(f(D_3))$. These two sets must differ since $f(D_1)$ and $f(D_3)$ differ. Since $\mathcal{F}$ is a σ-algebra and $\{D_1, D_3, D_5\}$, $f^{-1}(f(D_1))$, and $f^{-1}(f(D_3))$ are sets in $\mathcal{F}$, any intersection of these sets is also in $\mathcal{F}$. But since the two sets differ and are non-empty subsets of $\{D_1, D_3, D_5\}$, one of them must be a strict subset of $\{D_1, D_3, D_5\}$. And, since they are each in $\mathcal{F}$ with $\mathcal{F}$ an algebra of sets, this strict subset must be in $\mathcal{F}$. However, we know that they also are *not* in $\mathcal{F}$ – contradiction. Therefore, our premise that $f(D_1)$ and $f(D_3)$ could take differing values is incorrect. The same argument shows that $f(D_1)$ and $f(D_5)$ do not differ. One can repeat the above argument to show that f is constant on the other minimal set $\{D_2, D_4, D_6\}$.

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 $\{D_1, D_3, D_5\}$ or $\{D_2, D_4, D_6\}$. 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 σ-algebra $\mathcal{F}$ is the *unique* $\mathcal{F}$ *measurable* function labeled, $\mathbb{E}[f/\mathcal{F}]$, such that‡

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

Let us write this again in, perhaps, an unusual way.

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

In this second equation we use the fact that the conditional expectation is a *measurable* function with respect to $\mathcal{F}$. We do this by replacing the measure P on the left hand side with $P_{\mathcal{F}}$ to indicate that we are using the same measure, but one that is restricted to σ-algebra $\mathcal{F}$.

Although it seems that f – itself – satisfies the first equation, we have to be careful. The second equation is a way to emphasize that the conditional expectation function must be *measurable* with respect to the sigma algebra we are using for integration. The second equation makes clear that while, on the right, we are dealing with a $\mathcal{E}$ measurable function on $\mathcal{E}$, on the left we are not – we are dealing with a $\mathcal{F}$ measurable function. Consequently, while f seems like a nature candidate for the conditional expectation it is not necessarily $\mathcal{F}$ measurable. However, if f is measurable with respect to $\mathcal{F}$ then, by the definition above, it is its own conditional expectation.

The modern theory of probability – including conditional expectation – relies on measure theory which has its roots in Lebesgue measure theory – used to provide an alternative to the theory of Riemann Integration. From Lebesgue Measure Theory one is introduced to measurable and non-measurable sets/functions. The problem, in the context of Lebesgue, is that non-measurable sets/functions exist but specific examples are

† Specifically, in a discrete setting, a set Z in a σ-algebra, $\mathcal{H}$, is *minimal* in $\mathcal{H}$ if there is no non-empty, strict subset, Y, of Z with $Y \in \mathcal{H}$.

‡ That such a unique function exists is a consequence of the Radon-Nikodym theorem.

hard to come by. However, non-measurable functions appear *implicitly* in the definition of conditional expectation in that if they didn't the definition wouldn't be of any interest.

If one started out learning measure theory from the Lebesgue theory, the definition of conditional expectation might make readers somewhat uneasy as we are confronted with non-measurable sets from the very start. From this context, it is harder to get an intuitive idea of conditional expectation.

In the next section we examine the measure theoretic version of conditional expectation in a discrete probabilistic setting where such non-measurable functions are simple and explicit. In the process, we will provide an intuitive idea of the measure-theoretic definition of conditional expectation which coincides with the traditional approach – except in the "singular" case which we discuss in the last section.

Example Calculation of Conditional Expectation

Consider the function from the second on an elementary dice example, $f(D_i) = i$, which is measurable in the space $(P, X, \mathcal{E})$. Let $\mathcal{F}$ be the sub σ-algebra of that section. We now compute the conditional expectation of f with respect to $\mathcal{F}$. Using (1) we choose the *minimal* sets: $\Lambda_1 = \{D_1, D_3, D_5\}$ and $\Lambda_2 = \{D_2, D_4, D_6\}$. Since the conditional expectation function is constant on each of these sets, equation (1) will provide a way to find each value on each the two sets. Since the union of Λ_1 and Λ_2 constitute the entire set, X, we will know the value of the conditional expectation function on all of X; hence we will know the conditional expectation function.

Proceeding, we have

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

and

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

We know $\mathbb{E}\left[f/\mathcal{F}\right]$ is *constant* on Λ_1. We label this value as $\mathbb{E}\left[f/\mathcal{F}\right](\Lambda_1)$. From (2) we have

$$\begin{aligned}
\int_{\Lambda_1} \mathbb{E}\left[f/\mathcal{F}\right]\, dP_{\mathcal{F}} &= \int_{\Lambda_1} f\, dP \\
\mathbb{E}\left[f/\mathcal{F}\right](\Lambda_1) \int_{\Lambda_1} dP_{\mathcal{F}} &= \int_{\Lambda_1} f\, dP \\
\mathbb{E}\left[f/_{\mathcal{F}}\right](\Lambda_1) * P_{\mathcal{F}}(\Lambda_1) &= \int_{\Lambda_1} f\, dP = f(D_1) * P(D_1) + f(D_3) * P(D_3) + f(D_5) * P(D_5) && (**) \\
\mathbb{E}\left[f/\mathcal{F}\right](\Lambda_1) &= \int_{\Lambda_1} f\, dP = f(D_1) * \frac{P(D_1)}{P(\Lambda_1)} + f(D_3) * \frac{P(D_3)}{P(\Lambda_1)} + f(D_5) * \frac{P(D_5)}{P(\Lambda_1)} \\
\mathbb{E}\left[f/\mathcal{F}\right](\Lambda_1) &= \int_{\Lambda_1} f\, dP = 1 * \frac{\frac{1}{6}}{\frac{1}{2}} + 3 * \frac{\frac{1}{6}}{\frac{1}{2}} + 5 * \frac{\frac{1}{6}}{\frac{1}{2}} \\
\mathbb{E}\left[f/\mathcal{F}\right](\Lambda_1) &= \int_{\Lambda_1} f\, dP = 1 * \frac{1}{3} + 3 * \frac{1}{3} + 5 * \frac{1}{3} \\
\mathbb{E}\left[f/\mathcal{F}\right](\Lambda_1) &= 3
\end{aligned}$$

Likewise, $\mathbb{E}\left[f/\mathcal{F}\right]$ is constant on the set Λ_2. As we did above, we find the value of $\mathbb{E}\left[f/\mathcal{F}\right]$ on the set Λ_2. As with Λ_1, label $\mathbb{E}\left[f/\mathcal{F}\right](\Lambda_2)$ as the constant value of $\mathbb{E}\left[f/\mathcal{F}\right]$ on Λ_2. We have

$$
\begin{aligned}
\int_{\Lambda_2} \mathbb{E}\,[f/\mathcal{F}]\; dP_{\mathcal{F}} &= \int_{\Lambda_2} f\, dP \\
\mathbb{E}\,[f/\mathcal{F}](\Lambda_2) \int_{\Lambda_2} dP_{\mathcal{F}} &= \int_{\Lambda_2} f\, dP \\
\mathbb{E}\,[f/_{\mathcal{F}}](\Lambda_2) * P_{\mathcal{F}}(\Lambda_2) &= \int_{\Lambda_2} f\, dP = f(D_2) * P(D_2) + f(D_4) * P(D_4) + f(D_6) * P(D_6) \\
\mathbb{E}\,[f/\mathcal{F}](\Lambda_2) &= \int_{\Lambda_2} f\, dP = f(D_2) * \frac{P(D_2)}{P(\Lambda_2)} + f(D_4) * \frac{P(D_4)}{P(\Lambda_2)} + f(D_6) * \frac{P(D_6)}{P(\Lambda_2)} \\
\mathbb{E}\,[f/\mathcal{F}](\Lambda_2) &= \int_{\Lambda_2} f\, dP = 2 * \frac{\frac{1}{6}}{\frac{1}{2}} + 4 * \frac{\frac{1}{6}}{\frac{1}{2}} + 6 * \frac{\frac{1}{6}}{\frac{1}{2}} \\
\mathbb{E}\,[f/\mathcal{F}](\Lambda_2) &= \int_{\Lambda_2} f\, dP = 2 * \frac{1}{3} + 4 * \frac{1}{3} + 6 * \frac{1}{3} \\
\mathbb{E}\,[f/\mathcal{F}](\Lambda_2) &= 4
\end{aligned}
$$

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

From equation $(**)$ we see, that the value of the function $\mathbb{E}\,[f/\mathcal{F}]$ evaluated on any ‘x’ value in a minimal set, Λ, is‡

$$
\begin{aligned}
\mathbb{E}\,[f/\mathcal{F}](x) &= \frac{\int_\Lambda f\, dP}{P(\Lambda)} \\
&= \int_\Lambda f\, dP_\Lambda
\end{aligned}
\tag{4}
$$

That is, the value of the conditional expectation on any minimal set is the *weighted average* of f over the minimal set. The weights are determined by normalizing the probability measure P over the minimal set.

More generally, discrete probability or otherwise, we may think of the σ-algebra, $\mathcal{F}$, as a coarser “mesh” than the σ-algebra, $\mathcal{E}$. And we can think of the value of $\mathbb{E}\,[f/\mathcal{F}]$ 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 σ-algebra, $\mathcal{E}$.

From this perspective, you can think of the conditional expectation as giving the “best” representation of a function given a cruder mesh, $\mathcal{F}$, than the refined mesh, $\mathcal{E}$. Just as the “best” representation of an image at a larger pixel scale (crude mesh) would be an average over smaller scale pixels (refined mesh) of the larger pixel.

Example Redux

Let’s adjust the dice example probabilities and redo the calculation of the conditional expectation. Set the probability measure, P, as:

$$
P(A) = \frac{|A \cap \{D_1, D_3, D_5\}^c|}{3} \qquad A \in \mathcal{E}
$$

† In the “Redux” section we examine more closely what “knowing the value on every x” means.

‡ The notation, P_Λ means that we have normalized the measure, P, so that $P_\Lambda(\Lambda) = 1$. To do this we are *assuming* that $P(\Lambda) \neq 0$.

That is, the probability of a set, A, is the number of die in A that are not in the set $\{D_1, D_3, D_5\}$ divided by 3. With this definition, $P(\Lambda_1) = 0$ and $P(\Lambda_2) = 1$.

Given P, we can find the value of $\mathbb{E}[f/\mathcal{F}]$ on the set Λ_2 as before using equation (4). However, we can't use it to find the value $\mathbb{E}[f/\mathcal{F}]$ on the set Λ_1 as $P(\Lambda_1) = 0$.

It turns out, we don't have to determine the values with any specificity on Λ_1. We can *set* the values of $\mathbb{E}[f/\mathcal{F}]$ on Λ_1 to 0; or, to any other *single* value for that matter†. The reason for this is that the definition of conditional expectation determines a unique element in the function space on $\mathcal{F}$. However, each element in such a space is actually an equivalence class of measurable functions over $\mathcal{F}$ who differ by a set of measure 0.

We can't directly compute what the value of a representative of $\mathbb{E}[f/\mathcal{F}]$ is on the set Λ_1, but we don't have to, since it's a set of measure 0, any value on this set will give us a function that is in the equivalence class of the conditional expectation element. Therefore, we may take $\mathbb{E}[f/\mathcal{F}]$ to be 0 on Λ_1. This gives us a representative for the conditional expectation as we have values for all $x \in X$. Through this representative we can produce the associated equivalence class of functions. Consequently, we know $\mathbb{E}[f/\mathcal{F}]$ in the function space $L_2(P, X, \mathcal{F})$.

In the case of singular sets (set of probability 0) we may revise equation (4) and write that on minimal sets, $\mathbb{E}[f/\mathcal{F}]$ is constant and satisfies:

$$\mathbb{E}[f/\mathcal{F}](x) * P(\Lambda) = \int_{\Lambda} f \, dP \tag{5}$$

Here, with a measure theoretic methodology, we have an elegant way to talk about conditional expectation in a singular context that doesn't exist in the traditional approach.

† We know that Λ_1 is minimal; so, for the conditional expectation function to be measurable it must have the same value on all elements of Λ_1.