

Illustration of Conditional Expectation

Tianchen Qian

Nov. 25, 2013

Abstract

In this note, we explained why the conditional expectation of a random variable Y given a σ -field \mathcal{G} can be seen as “smoothed version of Y over \mathcal{G} ” (in Example 2), and we briefly related the definition of conditional expectation to the elementary $E(Y | X = x)$ notation.

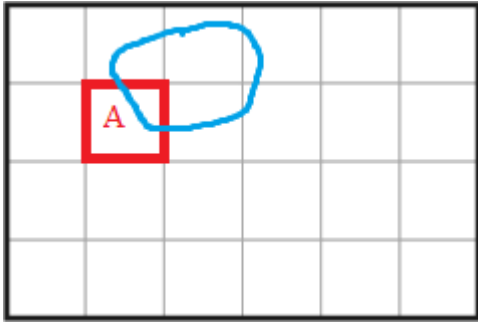
1 Preliminaries

Definition 1. Let (Ω, \mathcal{F}) and (Ω', \mathcal{F}') be two measurable spaces (that is to say, \mathcal{F} is a σ -field on Ω , and \mathcal{F}' on Ω'), and let X be a mapping from Ω to Ω' . We say X is \mathcal{F}/\mathcal{F}' -measurable, if $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \forall B \in \mathcal{F}'$.

Remark 1. Usually, when there is no confusion, we abbreviate X is \mathcal{F}/\mathcal{F}' -measurable as X is \mathcal{F} -measurable. Another convention is that, when people write $X : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$, sometimes they are implying X is \mathcal{F}/\mathcal{F}' -measurable (and we'll adopt this convention in this note).

Example 1. Say $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, i.e. assume X is $\mathcal{F}/\mathcal{B}(\mathbb{R})$ -measurable, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ -field on the real line. Consider the following figure (see top of next page), where the large rectangle represents Ω , and the small grids represent \mathcal{F} , so in this case the σ -field \mathcal{F} consists of all the small cells, and all combinations of them. In this case, say A is one of the “finest” element in \mathcal{F} , indicated in red (the “finest” here means that there is no nonempty subset of A that is also in \mathcal{F}), then X **must be constant on A** , i.e. there exists some $c \in \mathbb{R}$ such that $X(\omega) = c \forall \omega \in A$. This is because, if X could take two values on A , say $X(\omega_1) = c_1$ and $X(\omega_2) = c_2$ for some $\omega_1, \omega_2 \in A$ where $c_1 \neq c_2$, then $X^{-1}(\{c_1\})$ must be a part of A and maybe along with some other part of Ω outside A , illustrated in blue, and hence this pullback cannot be in \mathcal{F} (recall that \mathcal{F} consists of all the small cells, and all combinations of them), which violates the assumption that X is $\mathcal{F}/\mathcal{B}(\mathbb{R})$ -measurable. Note that this observation is also true when the target space is a general (Ω', \mathcal{F}') , i.e., when

$X : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$, because X is \mathcal{F} -measurable, X must be constant on each “finest” piece of \mathcal{F} .



In particular, take \mathcal{F} to be $\sigma(X)$ (recall that $\sigma(X)$ is defined as the smallest σ -field such that X is measurable). Since X is always $\sigma(X)$ measurable, we know that roughly speaking X is constant on each “finest” piece of $\sigma(X)$. To me, this is why people always say that $\sigma(X)$ “contains the information” of X : by knowing how “fine” $\sigma(X)$ is, one knows how “complicated” X is, i.e., how many different values X could take.

However, do note that this observation is valid when $\sigma(X)$ or \mathcal{F} is “discrete” in the sense that you can identify the “finest” grid, like in the above example. Otherwise, if say $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then we cannot find a “finest” piece of \mathcal{F} - you can say a single number in \mathbb{R} is a “finest” piece of $\mathcal{B}(\mathbb{R})$, but this won’t be useful, because X will always be constant on a single point in the sample space.

2 Illustration of conditional expectation

Definition 2. (Billinsley, Section 34, P445) Suppose Y is an integrable random variable on (Ω, \mathcal{F}, P) , and that \mathcal{G} is a sub σ -field of \mathcal{F} (i.e. $\mathcal{G} \subset \mathcal{F}$). Then there exists a function $Z : \Omega \rightarrow \Omega'$, called the *conditional expectation* of Y given \mathcal{G} , denoted as $E(Y | \mathcal{G})$, such that Z has the following two properties:

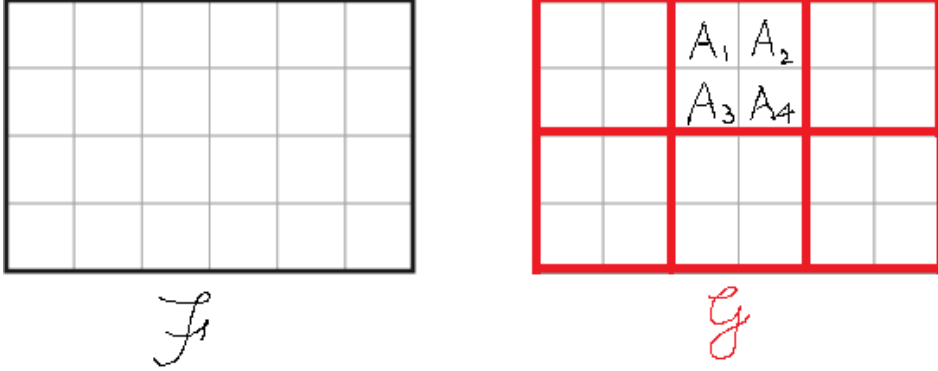
1. Z is \mathcal{G} -measurable and integrable;
2. Z satisfies the following equation:

$$\int_A Z dP = \int_A Y dP, \quad \forall A \in \mathcal{G}. \tag{1}$$

Remark 2. The existence of Z can be proved in a constructive way by four steps: Y is an indicator; Y is simple random variable; Y is nonnegative random variable; Y is integrable random variable. (See the paragraph after the definition of conditional expectation in Billinsley, Section 34, P445 for detail.) This routine has been used in constructing the Lebesgue integral. Also, one can show that the conditional expectation is a.s. unique, i.e., if Z and W both satisfy the above two conditions, then $Z = W$ a.s.

In the following example, we illustrate how the conditional expectation $E(Y | \mathcal{G})$ can be seen as the “smoothed version of Y over \mathcal{G} ”.

Example 2. Suppose Y is an integrable random variable on (Ω, \mathcal{F}, P) , and that \mathcal{G} is a sub σ -field of \mathcal{F} . The following two figures represent \mathcal{F} and \mathcal{G} , respectively. So \mathcal{F} contains all the 24 rectangles with grey edges and all combinations of them, while \mathcal{G} contains all the 6 rectangles with red edges and all combinations of them. Note that \mathcal{F} is a “finer” partition of Ω than \mathcal{G} .



Now, consider the 4 grey cells A_1, A_2, A_3, A_4 in the figure, and call their union $A = \bigcup_{i=1}^4 A_i$, note that $A \in \mathcal{G}$, and A can be seen as a “finest” piece of \mathcal{G} . Denote $Z := E(Y | \mathcal{G})$. By definition Z is \mathcal{G} -measurable, so Z need to be constant on A (as explained in Example 1). Also, by definition we need Z to satisfy the equation:

$$\int_A Z dP = \int_A Y dP = \sum_{i=1}^4 \int_{A_i} Y dP. \quad (2)$$

Using Z is constant on A and Y is constant on each A_i , we have (with a little abuse of notation, denote the constant value of Z on A as $Z(A)$, and similar to Y)

$$Z(A) P(A) = \sum_{i=1}^4 Y(A_i) P(A_i), \quad (3)$$

i.e.

$$Z(A) = \frac{1}{P(A)} \sum_{i=1}^4 P(A_i) Y(A_i) = \sum_{i=1}^4 \frac{P(A_i)}{P(A)} Y(A_i). \quad (4)$$

This means $Z(A)$ is the average of the 4 values that Y takes on A_i 's, with weight proportional to the probability of A_i . This observation extends to other cells of the grid as well.

To conclude what we get from this example, we find that on each cell of \mathcal{G} , $E(Y | \mathcal{G})$ is constant, and the constant value equals the weighted average of the values Y can take within this cell. So with this observation, we can say that $E(Y | \mathcal{G})$ is a “weighted average version of Y over \mathcal{G} ”.

More generally, if \mathcal{G} is a rather complicated σ -field than in this example, when considering

$E(Y | \mathcal{G})$, instead of calling it a “weighted average version of Y over \mathcal{G} ”, we can call it a “smoothed version of Y over \mathcal{G} ”. In other words, $E(Y | \mathcal{G})$ tries its best to perform just like Y , under the constraint that $E(Y | \mathcal{G})$ needs to be \mathcal{G} -measurable.

One can also substitute \mathcal{G} by $\sigma(X)$ for some X , assuming $\sigma(X) \subset \mathcal{F}$, and the logic follows the same way.

3 Relation to conditional expectation defined in introductory probability course

How is the above definition of conditional expectation related to the definition in introductory probability courses, i.e. $E(Y | X = x)$?

Given (Ω, \mathcal{F}, P) , let $Y : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and $X : (\Omega, \sigma(X)) \rightarrow (\Omega', \mathcal{F}')$, with $\sigma(X) \subset \mathcal{F}$. For the usual $E(Y|X = x)$, consider it as a function of x , and denote it by $g(x)$, i.e., $g : \Omega' \rightarrow \mathbb{R}, g(x) = E(Y|X = x)$. Now let $h := g \circ X : \Omega \rightarrow \mathbb{R}, h(\omega) = g(X(\omega))$, as in the following figure (figure adopted from R. Ash’s Probability and Measure Theory):

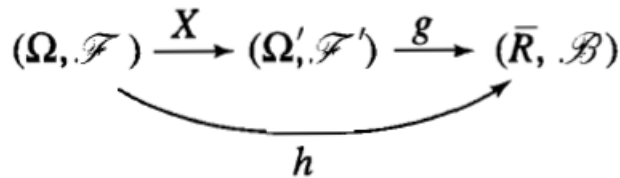


Figure 5.4.1

One can show that the random variable h here serves as the conditional expectation $E(Y | \sigma(X))$. For detailed proof, see R. Ash’s Probability and Measure Theory (2nd edition) Section 5.4, P215-216.

Reference

1. Billingsley, P. (2008). Probability and measure. John Wiley & Sons.
2. Ash, R. B., & Doléans-Dade, C. A. (2000). Probability and measure theory. Access Online via Elsevier.