

4.10: Conditional Expected Value Revisited

Conditional expected value is much more important than one might at first think. In fact, conditional expected value is at the core of modern probability theory because it provides the basic way of incorporating known information into a probability measure.

Basic Theory

Definition

As usual, our starting point is a random experiment modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, so that Ω is the set of outcomes, \mathcal{F} is the σ -algebra of events, and \mathbb{P} is the probability measure on the sample space (Ω, \mathcal{F}) . In our first elementary discussion, we studied the conditional expected value of a real-value random variable X given a general random variable Y . The more general approach is to condition on a sub σ -algebra \mathcal{G} of \mathcal{F} . The sections on σ -algebras and measure theory are essential prerequisites for this section.

Before we get to the definition, we need some preliminaries. First, all random variables mentioned are assumed to be real valued. Next the notion of equivalence plays a fundamental role in this section. Next recall that random variables X_1 and X_2 are *equivalent* if $\mathbb{P}(X_1 = X_2) = 1$. Equivalence really does define an equivalence relation on the collection of random variables defined on the sample space. Moreover, we often regard equivalent random variables as being essentially the same object. More precisely from this point of view, the objects of our study are not individual random variables but rather equivalence classes of random variables under this equivalence relation. Finally, for $A \in \mathcal{F}$, recall the notation for the expected value of X on the event A

$$\mathbb{E}(X; A) = \mathbb{E}(X \mathbf{1}_A) \quad (4.10.1)$$

assuming of course that the expected value exists. For the remainder of this subsection, suppose that \mathcal{G} is a sub σ -algebra of \mathcal{F} .

Suppose that X is a random variable with $\mathbb{E}(|X|) < \infty$. The *conditional expected value* of X given \mathcal{G} is the random variable $\mathbb{E}(X | \mathcal{G})$ defined by the following properties:

1. $\mathbb{E}(X | \mathcal{G})$ is measurable with respect to \mathcal{G} .
2. If $A \in \mathcal{G}$ then $\mathbb{E}[\mathbb{E}(X | \mathcal{G}); A] = \mathbb{E}(X; A)$

The basic idea is that $\mathbb{E}(X | \mathcal{G})$ is the expected value of X given the information in the σ -algebra \mathcal{G} . Hopefully this idea will become clearer during our study. The conditions above uniquely define $\mathbb{E}(X | \mathcal{G})$ up to equivalence. The proof of this fact is a simple application of the Radon-Nikodym theorem, named for Johann Radon and Otto Nikodym

Suppose again that X is a random variable with $\mathbb{E}(|X|) < \infty$.

1. There exists a random variable V satisfying the [definition](#).
2. If V_1 and V_2 satisfy the [definition](#), then $\mathbb{P}(V_1 = V_2) = 1$ so that V_1 and V_2 are equivalent.

Proof

1. Note that $\nu(A) = \mathbb{E}(X; A)$ for $A \in \mathcal{G}$ defines a (signed) measure on \mathcal{G} . Moreover, if $A \in \mathcal{G}$ and $\mathbb{P}(A) = 0$ then $\nu(A) = 0$. Hence ν is absolutely continuous with respect to the restriction of \mathbb{P} to \mathcal{G} . By the Radon-Nikodym theorem, there exists a random variable V that is measurable with respect to \mathcal{G} such that $\nu(A) = \mathbb{E}(V; A)$ for $A \in \mathcal{G}$. That is, V is the *density* or *derivative* of ν with respect to \mathbb{P} on \mathcal{G} .
2. This follows from the uniqueness of the Radon-Nikodym derivative, up to equivalence.

The following characterization might seem stronger but in fact is equivalent to the definition.

Suppose again that X is a random variable with $\mathbb{E}(|X|) < \infty$. Then $\mathbb{E}(X | \mathcal{G})$ is characterized by the following properties:

1. $\mathbb{E}(X | \mathcal{G})$ is measurable with respect to \mathcal{G}
2. If U is measurable with respect to \mathcal{G} and $\mathbb{E}(|UX|) < \infty$ then $\mathbb{E}[U\mathbb{E}(X | \mathcal{G})] = \mathbb{E}(UX)$.

Proof

We have to show that part (b) in the [definition](#) is equivalent to part (b) here. First (b) here implies (b) in the definition since $\mathbf{1}_A$ is \mathcal{G} -measurable if $A \in \mathcal{G}$. Conversely suppose that (b) in the definition holds. We will show that (b) here holds by a classical *bootstrapping argument*. First $\mathbb{E}[U\mathbb{E}(X | \mathcal{G})] = \mathbb{E}(UX)$ if $U = \mathbf{1}_A$ for some $A \in \mathcal{G}$. Next suppose that U is a simple random variable that is \mathcal{G} -measurable. That is, $U = \sum_{i \in I} a_i \mathbf{1}_{A_i}$ where I is a finite index set, $a_i \geq 0$ for $i \in I$, and $A_i \in \mathcal{G}$ for $i \in I$. then

$$\mathbb{E}[U\mathbb{E}(X | \mathcal{G})] = \mathbb{E}\left[\sum_{i \in I} a_i \mathbf{1}_{A_i} \mathbb{E}(X | \mathcal{G})\right] = \sum_{i \in I} a_i \mathbb{E}[\mathbf{1}_{A_i} \mathbb{E}(X | \mathcal{G})] = \sum_{i \in I} a_i \mathbb{E}(\mathbf{1}_{A_i} X) = \mathbb{E}\left(\sum_{i \in I} a_i \mathbf{1}_{A_i} X\right) = \mathbb{E}(UX) \quad (4.10.2)$$

Next suppose that U is nonnegative and \mathcal{G} -measurable. Then there exists a sequence of simple \mathcal{G} -measurable random variables (U_1, U_2, \dots) with $U_n \uparrow U$ as $n \rightarrow \infty$. Then by the previous step, $\mathbb{E}[U_n \mathbb{E}(X | \mathcal{G})] = \mathbb{E}(U_n X)$ for each n . Letting $n \rightarrow \infty$ and using the monotone convergence theorem we have $\mathbb{E}[U\mathbb{E}(X | \mathcal{G})] = \mathbb{E}(UX)$. Finally, suppose that U is a general \mathcal{G} -measurable random variable. Then $U = U^+ - U^-$ where U^+ and U^- are the usual positive and negative parts of U . These parts are nonnegative and \mathcal{G} -measurable, so by the previous step, $\mathbb{E}[U^+ \mathbb{E}(X | \mathcal{G})] = \mathbb{E}(U^+ X)$ and $\mathbb{E}[U^- \mathbb{E}(X | \mathcal{G})] = \mathbb{E}(U^- X)$. hence

$$\mathbb{E}[U\mathbb{E}(X | \mathcal{G})] = \mathbb{E}[(U^+ - U^-)\mathbb{E}(X | \mathcal{G})] = \mathbb{E}[U^+ \mathbb{E}(X | \mathcal{G})] - \mathbb{E}[U^- \mathbb{E}(X | \mathcal{G})] = \mathbb{E}(U^+ X) - \mathbb{E}(U^- X) = \mathbb{E}(UX) \quad (4.10.3)$$

Properties

Our next discussion concerns some fundamental properties of conditional expected value. All equalities and inequalities are understood to hold *modulo equivalence*, that is, with probability 1. Note also that many of the proofs work by showing that the right hand side satisfies the properties in the [definition](#) for the conditional expected value on the left side. Once again we assume that \mathcal{G} is a sub *sigma*-algebra of \mathcal{F} .

Our first property is a simple consequence of the definition: X and $\mathbb{E}(X | \mathcal{G})$ have the same mean.

Suppose that X is a random variable with $\mathbb{E}(|X|) < \infty$. Then $\mathbb{E}[\mathbb{E}(X | \mathcal{G})] = \mathbb{E}(X)$.

Proof

This follows immediately by letting $A = \Omega$ in the [definition](#).

The result above can often be used to compute $\mathbb{E}(X)$, by choosing the σ -algebra \mathcal{G} in a clever way. We say that we are computing $\mathbb{E}(X)$ by *conditioning on \mathcal{G}* . Our next properties are fundamental: every version of expected value must satisfy the *linearity* properties. The first part is the *additive property* and the second part is the *scaling property*.

Suppose that X and Y are random variables with $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|Y|) < \infty$, and that $c \in \mathbb{R}$. Then

1. $\mathbb{E}(X + Y | \mathcal{G}) = \mathbb{E}(X | \mathcal{G}) + \mathbb{E}(Y | \mathcal{G})$
2. $\mathbb{E}(cX | \mathcal{G}) = c\mathbb{E}(X | \mathcal{G})$

Proof

1. Note that $\mathbb{E}(|X + Y|) \leq \mathbb{E}(|X|) + \mathbb{E}(|Y|) < \infty$ so $\mathbb{E}(X + Y | \mathcal{G})$ is defined. We show that $\mathbb{E}(X | \mathcal{G}) + \mathbb{E}(Y | \mathcal{G})$ satisfies the conditions in the [definition](#) for $\mathbb{E}(X + Y | \mathcal{G})$. Note first that $\mathbb{E}(X | \mathcal{G}) + \mathbb{E}(Y | \mathcal{G})$ is \mathcal{G} -measurable since both terms are. If $A \in \mathcal{G}$ then

$$\mathbb{E}\{\mathbb{E}(X | \mathcal{G}) + \mathbb{E}(Y | \mathcal{G}); A\} = \mathbb{E}[\mathbb{E}(X | \mathcal{G}); A] + \mathbb{E}[\mathbb{E}(Y | \mathcal{G}); A] = \mathbb{E}(X; A) + \mathbb{E}(Y; A) = \mathbb{E}(X + Y; A) \quad (4.10.4)$$

2. Note that $\mathbb{E}(|cX|) = |c|\mathbb{E}(|X|) < \infty$ so $\mathbb{E}(cX | \mathcal{G})$ is defined. We show that $c\mathbb{E}(X | \mathcal{G})$ satisfy the conditions in the [definition](#) for $\mathbb{E}(cX | \mathcal{G})$. Note first that $c\mathbb{E}(X | \mathcal{G})$ is \mathcal{G} -measurable since the second factor is. If $A \in \mathcal{G}$ then

$$\mathbb{E}[c\mathbb{E}(X | \mathcal{G}); A] = c\mathbb{E}[\mathbb{E}(X | \mathcal{G}); A] = c\mathbb{E}(X; A) = \mathbb{E}(cX; A) \quad (4.10.5)$$

The next set of properties are also fundamental to every notion of expected value. The first part is the *positive property* and the second part is the *increasing property*.

Suppose again that X and Y are random variables with $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|Y|) < \infty$.

1. If $X \geq 0$ then $\mathbb{E}(X | \mathcal{G}) \geq 0$
2. If $X \leq Y$ then $\mathbb{E}(X | \mathcal{G}) \leq \mathbb{E}(Y | \mathcal{G})$

Proof

1. Let $A = \{\mathbb{E}(X | \mathcal{G}) < 0\}$. Note that $A \in \mathcal{G}$ and hence $\mathbb{E}(X; A) = \mathbb{E}[\mathbb{E}(X | \mathcal{G}); A]$. Since $X \geq 0$ with probability 1 we have $\mathbb{E}(X; A) \geq 0$. On the other hand, if $\mathbb{P}(A) > 0$ then $\mathbb{E}[\mathbb{E}(X | \mathcal{G}); A] < 0$ which is a contradiction. Hence we must have $\mathbb{P}(A) = 0$.
2. Note that if $X \leq Y$ then $Y - X \geq 0$. Hence by (a) and the additive property, $\mathbb{E}(Y - X | \mathcal{G}) = \mathbb{E}(Y | \mathcal{G}) - \mathbb{E}(X | \mathcal{G}) \geq 0$ so $\mathbb{E}(Y | \mathcal{G}) \geq \mathbb{E}(X | \mathcal{G})$.

The next few properties relate to the central idea that $\mathbb{E}(X | \mathcal{G})$ is the expected value of X given the information in the σ -algebra \mathcal{G} .

Suppose that X and V are random variables with $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|XV|) < \infty$ and that V is measurable with respect to \mathcal{G} . Then $\mathbb{E}(VX | \mathcal{G}) = V\mathbb{E}(X | \mathcal{G})$.

Proof

We show that $V\mathbb{E}(X | \mathcal{G})$ satisfy the in [properties that characterize \$\mathbb{E}\(VX | \mathcal{G}\)\$](#) . First, $V\mathbb{E}(X | \mathcal{G})$ is \mathcal{G} -measurable since both factors are. If U is \mathcal{G} -measurable with $\mathbb{E}(|UVX|) < \infty$ then UV is also \mathcal{G} -measurable and hence

$$\mathbb{E}[UV\mathbb{E}(X | \mathcal{G})] = \mathbb{E}(UVX) = \mathbb{E}[U(VX)] \quad (4.10.6)$$

Compare this result with the [scaling property](#). If V is measurable with respect to \mathcal{G} then V is like a constant in terms of the conditional expected value given \mathcal{G} . On the other hand, note that this result implies the [scaling property](#), since a constant can be viewed as a random variable, and as such, is measurable with respect to any σ -algebra. As a corollary to this result, note that if X itself is measurable with respect to \mathcal{G} then $\mathbb{E}(X | \mathcal{G}) = X$. The following result gives the other extreme.

Suppose that X is a random variable with $\mathbb{E}(|X|) < \infty$. If X and \mathcal{G} are independent then $\mathbb{E}(X | \mathcal{G}) = \mathbb{E}(X)$.

Proof

We show that $\mathbb{E}(X)$ satisfy the properties in the [definition](#) for $\mathbb{E}(X | \mathcal{G})$. First of course, $\mathbb{E}(X)$ is \mathcal{G} -measurable as a constant random variable. If $A \in \mathcal{G}$ then X and 1_A are independent and hence

$$\mathbb{E}(X; A) = \mathbb{E}(X)\mathbb{P}(A) = \mathbb{E}[\mathbb{E}(X); A] \quad (4.10.7)$$

Every random variable X is independent of the trivial σ -algebra $\{\emptyset, \Omega\}$ so it follows that $\mathbb{E}(X | \{\emptyset, \Omega\}) = \mathbb{E}(X)$.

The next properties are *consistency conditions*, also known as the *tower properties*. When conditioning twice, with respect to nested σ -algebras, the smaller one (representing the least amount of information) always prevails.

Suppose that X is a random variable with $\mathbb{E}(|X|) < \infty$ and that \mathcal{H} is a sub σ -algebra of \mathcal{G} . Then

1. $\mathbb{E}[\mathbb{E}(X | \mathcal{H}) | \mathcal{G}] = \mathbb{E}(X | \mathcal{H})$
2. $\mathbb{E}[\mathbb{E}(X | \mathcal{G}) | \mathcal{H}] = \mathbb{E}(X | \mathcal{H})$

Proof

1. Note first that $\mathbb{E}(X | \mathcal{H})$ is \mathcal{H} -measurable and hence also \mathcal{G} -measurable. Thus by (7), $\mathbb{E}[\mathbb{E}(X | \mathcal{H}) | \mathcal{G}] = \mathbb{E}(X | \mathcal{H})$.
2. We show that $\mathbb{E}(X | \mathcal{H})$ satisfies the conditions in the [definition](#) for $\mathbb{E}[\mathbb{E}(X | \mathcal{G}) | \mathcal{H}]$. Note again that $\mathbb{E}(X | \mathcal{H})$ is \mathcal{H} -measurable. If $A \in \mathcal{H}$ then $A \in \mathcal{G}$ and hence

$$\mathbb{E}[\mathbb{E}(X | \mathcal{G}); A] = \mathbb{E}(X; A) = \mathbb{E}[\mathbb{E}(X | \mathcal{H}); A] \quad (4.10.8)$$

The next result gives *Jensen's inequality* for conditional expected value, named for Johan Jensen.

Suppose that X takes values in an interval $S \subseteq \mathbb{R}$ and that $g: S \rightarrow \mathbb{R}$ is convex. If $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|g(X)|) < \infty$ then

$$\mathbb{E}[g(X) | \mathcal{G}] \geq g[\mathbb{E}(X | \mathcal{G})] \quad (4.10.9)$$

Proof

As with Jensen's inequality for ordinary expected value, the best proof uses the characterization of convex functions in terms of *supporting lines*: For each $t \in S$ there exist numbers a and b (depending on t) such that

- $a + bt = g(t)$
- $a + bx \leq g(x)$ for $x \in S$

A convex function and several supporting lines



Random variables X and $\mathbb{E}(X | \mathcal{G})$ takes values in S . We can construct a *random* supporting line at $\mathbb{E}(X | \mathcal{G})$. That is, there exist random variables A and B , measurable with respect to \mathcal{G} , such that

1. $A + B\mathbb{E}(X | \mathcal{G}) = g[\mathbb{E}(X | \mathcal{G})]$
2. $A + BX \leq g(X)$

We take conditional expected value through the inequality in (b) and then use properties of conditional expected value and property (a):

$$\mathbb{E}[g(X) | \mathcal{G}] \geq \mathbb{E}(A + BX | \mathcal{G}) = A + B\mathbb{E}(X | \mathcal{G}) = g[\mathbb{E}(X | \mathcal{G})] \quad (4.10.10)$$

Note that the second step uses the fact that A and B are measurable with respect to \mathcal{G} .

Conditional Probability

For our next discussion, suppose as usual that \mathcal{G} is a sub σ -algebra of \mathcal{F} . The conditional probability of an event A given \mathcal{G} can be defined as a special case of conditional expected value. As usual, let $\mathbf{1}_A$ denote the indicator random variable of A .

For $A \in \mathcal{F}$ we define

$$\mathbb{P}(A | \mathcal{G}) = \mathbb{E}(\mathbf{1}_A | \mathcal{G}) \quad (4.10.11)$$

Thus, we have the following characterizations of conditional probability, which are special cases of the [definition](#) and the [alternate version](#):

If $A \in \mathcal{F}$ then $\mathbb{P}(A | \mathcal{G})$ is characterized (up to equivalence) by the following properties

1. $\mathbb{P}(A | \mathcal{G})$ is measurable with respect to \mathcal{G} .
2. If $B \in \mathcal{G}$ then $\mathbb{E}[\mathbb{P}(A | \mathcal{G}); B] = \mathbb{P}(A \cap B)$

Proof

For part (b), note that

$$\mathbb{E}[\mathbf{1}_B \mathbb{P}(A | \mathcal{G})] = \mathbb{E}[\mathbf{1}_B \mathbb{E}(\mathbf{1}_A | \mathcal{G})] = \mathbb{E}(\mathbf{1}_B \mathbf{1}_A) = \mathbb{E}(\mathbf{1}_{A \cap B}) = \mathbb{P}(A \cap B) \quad (4.10.12)$$

If $A \in \mathcal{F}$ then $\mathbb{P}(A | \mathcal{G})$ is characterized (up to equivalence) by the following properties

1. $\mathbb{P}(A | \mathcal{G})$ is measurable with respect to \mathcal{G} .
2. If U is measurable with respect to \mathcal{G} and $\mathbb{E}(|U|) < \infty$ then $\mathbb{E}[U \mathbb{P}(A | \mathcal{G})] = \mathbb{E}(U; A)$

The properties above for conditional expected value, of course, have special cases for conditional probability. In particular, we can compute the probability of an event by *conditioning* on a σ -algebra:

If $A \in \mathcal{F}$ then $\mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A | \mathcal{G})]$.

Proof

This is a direct result of the [mean property](#) since $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.

Again, the last theorem is often a good way to compute $\mathbb{P}(A)$ when we know the conditional probability of A given \mathcal{G} . This is a very compact and elegant version of the law of total probability given first in the section on Conditional Probability in the chapter on Probability Spaces and later in the section on Discrete Distributions in the Chapter on Distributions. The following theorem gives the conditional version of the axioms of probability.

The following properties hold (as usual, modulo equivalence):

1. $\mathbb{P}(A | \mathcal{G}) \geq 0$ for every $A \in \mathcal{F}$
2. $\mathbb{P}(\Omega | \mathcal{G}) = 1$
3. If $\{A_i : i \in I\}$ is a countable disjoint subset of \mathcal{F} then $\mathbb{P}(\bigcup_{i \in I} A_i | \mathcal{G}) = \sum_{i \in I} \mathbb{P}(A_i | \mathcal{G})$

Proof

1. This is a direct consequence of (6).
2. This is trivial since $\mathbf{1}_\Omega = 1$.
3. We show that the right side satisfies the conditions in (11) that define the left side. Note that $\sum_{i \in I} \mathbb{P}(A_i | \mathcal{G})$ is \mathcal{G} -measurable since each term in the sum has this property. Let $B \in \mathcal{G}$. then

$$\mathbb{E}\left[\sum_{i \in I} \mathbb{P}(A_i | \mathcal{G}); B\right] = \sum_{i \in I} \mathbb{E}[\mathbb{P}(A_i | \mathcal{G}); B] = \sum_{i \in I} \mathbb{P}(A_i \cap B) = \mathbb{P}\left(B \cap \bigcup_{i \in I} A_i\right) \quad (4.10.13)$$

From the last result, it follows that other standard probability rules hold for conditional probability given \mathcal{G} (as always, modulo equivalence). These results include

- the complement rule

- the increasing property
- Boole's inequality
- Bonferroni's inequality
- the inclusion-exclusion laws

However, it is not correct to state that $A \mapsto \mathbb{P}(A | \mathcal{G})$ is a probability measure, because the conditional probabilities are only defined up to equivalence, and so the mapping does not make sense. We would have to specify a particular version of $\mathbb{E}(A | \mathcal{G})$ for each $A \in \mathcal{F}$ for the mapping to make sense. Even if we do this, the mapping may not define a probability measure. In part (c), the left and right sides are random variables and the equation is an event that has probability 1. However this event depends on the collection $\{A_i : i \in I\}$. In general, there will be uncountably many such collections in \mathcal{F} , and the intersection of all of the corresponding events may well have probability less than 1 (if it's measurable at all). It turns out that if the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is sufficiently "nice" (and most probability spaces that arise in applications are nice), then there does in fact exist a *regular conditional probability*. That is, for each $A \in \mathcal{F}$, there exists a random variable $\mathbb{P}(A | \mathcal{G})$ satisfying the conditions in (12) and such that with probability 1, $A \mapsto \mathbb{P}(A | \mathcal{G})$ is a probability measure.

The following theorem gives a version of *Bayes' theorem*, named for the inimitable Thomas Bayes.

Suppose that $A \in \mathcal{G}$ and $B \in \mathcal{F}$. then

$$\mathbb{P}(A | B) = \frac{\mathbb{E}[\mathbb{P}(B | \mathcal{G}); A]}{\mathbb{E}[\mathbb{P}(B | \mathcal{G})]} \quad (4.10.14)$$

Proof

The proof is absolutely trivial. By definition of conditional probability given \mathcal{G} , the numerator is $\mathbb{P}(A \cap B)$ and the denominator is $\mathbb{P}(B)$. Nonetheless, Bayes' theorem is useful in settings where the expected values in the numerator and denominator can be computed directly

Basic Examples

The purpose of this discussion is to tie the general notions of conditional expected value that we are studying here to the more elementary concepts that you have seen before. Suppose that A is an event (that is, a member of \mathcal{F}) with $\mathbb{P}(A) > 0$. If B is another event, then of course, the conditional probability of B given A is

$$\mathbb{P}(B | A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \quad (4.10.15)$$

If X is a random variable then the conditional distribution of X given A is the probability measure on \mathbb{R} given by

$$R \mapsto \mathbb{P}(X \in R | A) = \frac{\mathbb{P}(\{X \in R\} \cap A)}{\mathbb{P}(A)} \text{ for measurable } R \subseteq \mathbb{R} \quad (4.10.16)$$

If $\mathbb{E}(|X|) < \infty$ then the *conditional expected value* of X given A , denoted $\mathbb{E}(X | A)$, is simply the mean of this conditional distribution.

Suppose now that $\mathcal{A} = \{A_i : i \in I\}$ is a countable partition of the sample space Ω into events with positive probability. To review the jargon, $\mathcal{A} \subseteq \mathcal{F}$; the index set I is countable; $A_i \cap A_j = \emptyset$ for distinct $i, j \in I$; $\bigcup_{i \in I} A_i = \Omega$; and $\mathbb{P}(A_i) > 0$ for $i \in I$. Let $\mathcal{G} = \sigma(\mathcal{A})$, the σ -algebra generated by \mathcal{A} . The elements of \mathcal{G} are of the form $\bigcup_{j \in J} A_j$ for $J \subseteq I$. Moreover, the random variables that are measurable with respect to \mathcal{G} are precisely the variables that are constant on A_i for each $i \in I$. The σ -algebra \mathcal{G} is said to be *countably generated*.

If $B \in \mathcal{F}$ then $\mathbb{P}(B | \mathcal{G})$ is the random variable whose value on A_i is $\mathbb{P}(B | A_i)$ for each $i \in I$.

Proof

Let U denote the random variable that takes the value $\mathbb{P}(B | A_i)$ on A_i for each $i \in I$. First, U is measurable with respect to \mathcal{G} since U is constant on A_i for each $i \in I$. So we just need to show that $\mathbb{E}(U; A) = \mathbb{P}(A \cap B)$ for each $A \in \mathcal{G}$. Thus, let $A = \bigcup_{j \in J} A_j$ where $J \subseteq I$. Then

$$\mathbb{E}(U; A) = \sum_{j \in J} \mathbb{E}(U; A_j) = \sum_{j \in J} \mathbb{P}(B | A_j) \mathbb{P}(A_j) = \mathbb{P}(A \cap B) \quad (4.10.17)$$

In this setting, the version of Bayes' theorem in (15) reduces to the usual elementary formulation: For $i \in I$, $\mathbb{E}[\mathbb{P}(B | \mathcal{G}); A_i] = \mathbb{P}(A_i) \mathbb{P}(B | A_i)$ and $\mathbb{E}[\mathbb{P}(B | \mathcal{G})] = \sum_{j \in I} \mathbb{P}(A_j) \mathbb{P}(B | A_j)$. Hence

$$\mathbb{P}(A_i | B) = \frac{\mathbb{P}(A_i) \mathbb{P}(B | A_i)}{\sum_{j \in I} \mathbb{P}(A_j) \mathbb{P}(B | A_j)} \quad (4.10.18)$$

If X is a random variable with $\mathbb{E}(|X|) < \infty$, then $\mathbb{E}(X | \mathcal{G})$ is the random variable whose value on A_i is $\mathbb{E}(X | A_i)$ for each $i \in I$.

Proof

Let U denote the random variable that takes the value $\mathbb{E}(X | A_i)$ on A_i for each $i \in I$. First, U is measurable with respect to \mathcal{G} since U is constant on A_i for each $i \in I$. So we just need to show that $\mathbb{E}(U; A) = \mathbb{E}(X; A)$ for each $A \in \mathcal{G}$. Thus, let $A = \bigcup_{j \in J} A_j$ where $J \subseteq I$. Then

$$\mathbb{E}(U; A) = \sum_{j \in J} \mathbb{E}(U; A_j) = \sum_{j \in J} \mathbb{E}(X | A_j) \mathbb{P}(A_j) = \mathbb{E}(X; A) \quad (4.10.19)$$

The previous examples would apply to $\mathcal{G} = \sigma(Y)$ if Y is a discrete random variable taking values in a countable set T . In this case, the partition is simply $\mathcal{A} = \{\{Y = y\} : y \in T\}$. On the other hand, suppose that Y is a random variable taking values in a general set T with σ -algebra \mathcal{T} . The real-valued random variables that are measurable with respect to $\mathcal{G} = \sigma(Y)$ are (up to equivalence) the measurable, real-valued functions of Y .

Specializing further, Suppose that X takes values in $S \subseteq \mathbb{R}$, Y takes values in $T \subseteq \mathbb{R}^n$ (where S and T are Lebesgue measurable) and that (X, Y) has a joint continuous distribution with probability density function f . Then Y has probability density function h given by

$$h(y) = \int_S f(x, y) dx, \quad y \in T \quad (4.10.20)$$

Assume that $h(y) > 0$ for $y \in T$. Then for $y \in T$, a conditional probability density function of X given $Y = y$ is defined by

$$g(x | y) = \frac{f(x, y)}{h(y)}, \quad x \in S \quad (4.10.21)$$

This is precisely the setting of our elementary discussion of conditional expected value. If $\mathbb{E}(|X|) < \infty$ then we usually write $\mathbb{E}(X | Y)$ instead of the clunkier $\mathbb{E}[X | \sigma(Y)]$.

In this setting above suppose that $\mathbb{E}(|X|) < \infty$. Then

$$\mathbb{E}(X | Y) = \int_S x g(x | Y) dx \quad (4.10.22)$$

Proof

Once again, we show that the integral on the right satisfies the properties in the [definition](#) for $\mathbb{E}(X | Y) = \mathbb{E}[X | \sigma(Y)]$. First, $y \mapsto \int_S x g(x | y) dx$ is measurable as a function from T into \mathbb{R} and hence the random variable $\int_S x g(x | Y) dx$ is a measurable function of Y and so is measurable with respect to $\sigma(Y)$. Next suppose that $B \in \sigma(Y)$. Then $B = \{Y \in A\}$ for some $A \in \mathcal{F}$. Then

$$\begin{aligned} \mathbb{E} \left[\int_S x g(x | Y) dx; B \right] &= \mathbb{E} \left[\int_S x g(x | Y) dx; Y \in A \right] \\ &= \mathbb{E} \left[\int_S x \frac{f(x, y)}{h(y)} dx; Y \in A \right] = \int_A \int_S x \frac{f(x, y)}{h(y)} h(y) dx dy \\ &= \int_{S \times A} x f(x, y) d(x, y) = \mathbb{E}(X; Y \in A) = \mathbb{E}(X; B) \end{aligned}$$

Best Predictor

In our elementary treatment of conditional expected value, we showed that the conditional expected value of a real-valued random variable X given a general random variable Y is the best predictor of X , in the least squares sense, among all real-valued functions of Y . A more careful statement is that $\mathbb{E}(X | Y)$ is the best predictor of X among all real-valued random variables that are *measurable* with respect to $\sigma(Y)$. Thus, it should come as not surprise that if \mathcal{G} is a sub σ -algebra of \mathcal{F} , then $\mathbb{E}(X | \mathcal{G})$ is the best predictor of X , in the least squares sense, among all real-valued random variables that are measurable with respect to \mathcal{G} . We will show that this is indeed the case in this subsection. The proofs are very similar to the ones given in the elementary section. For the rest of this discussion, we assume that \mathcal{G} is a sub σ -algebra of \mathcal{F} and that all random variables mentioned are real valued.

Suppose that X and U are random variables with $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|XU|) < \infty$ and that U is measurable with respect to \mathcal{G} . Then $X - \mathbb{E}(X | \mathcal{G})$ and U are uncorrelated.

Proof

Note that $X - \mathbb{E}(X | \mathcal{G})$ has mean 0 by the [mean property](#). Using the [properties that characterize](#) $\mathbb{E}(X | \mathcal{G})$ we have

$$\text{cov}[X - \mathbb{E}(X | \mathcal{G}), U] = \mathbb{E}[U[X - \mathbb{E}(X | \mathcal{G})]] = \mathbb{E}(UX) - \mathbb{E}[U\mathbb{E}(X | \mathcal{G})] = \mathbb{E}(UX) - \mathbb{E}(UX) = 0 \quad (4.10.23)$$

The next result is the main one: $\mathbb{E}(X | \mathcal{G})$ is closer to X in the mean square sense than any other random variable that is measurable with respect to \mathcal{G} . Thus, if \mathcal{G} represents the information that we have, then $\mathbb{E}(X | \mathcal{G})$ is the best we can do in estimating X .

Suppose that X and U are random variables with $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(U^2) < \infty$ and that U is measurable with respect to \mathcal{G} . Then

1. $\mathbb{E}([X - \mathbb{E}(X | \mathcal{G})]^2) \leq \mathbb{E}([X - U]^2)$.
2. Equality holds if and only if $\mathbb{P}[U = \mathbb{E}(X | \mathcal{G})] = 1$, so U and $\mathbb{E}(X | \mathcal{G})$ are equivalent.

Proof

1. Note that

$$\mathbb{E}([X - U]^2) = \mathbb{E}([X - \mathbb{E}(X | \mathcal{G}) + \mathbb{E}(X | \mathcal{G}) - U]^2) \quad (4.10.24)$$

$$= \mathbb{E}([X - \mathbb{E}(X | \mathcal{G})]^2) + 2\mathbb{E}([X - \mathbb{E}(X | \mathcal{G})][\mathbb{E}(X | \mathcal{G}) - U]) + \mathbb{E}([\mathbb{E}(X | \mathcal{G}) - U]^2) \quad (4.10.25)$$

By [mean property](#), $X - \mathbb{E}(X | \mathcal{G})$ has mean 0, so the middle term in the displayed equation is $2\text{cov}[X - \mathbb{E}(X | \mathcal{G}), \mathbb{E}(X | \mathcal{G}) - U]$. But $\mathbb{E}(X | \mathcal{G}) - U$ is \mathcal{G} -measurable and hence this covariance is 0 by [uncorrelated property](#). Therefore

$$\mathbb{E}([X - U]^2) = \mathbb{E}([X - \mathbb{E}(X | \mathcal{G})]^2) + \mathbb{E}([\mathbb{E}(X | \mathcal{G}) - U]^2) \geq \mathbb{E}([X - \mathbb{E}(X | \mathcal{G})]^2) \quad (4.10.26)$$

2. Equality holds if and only if $\mathbb{E}([\mathbb{E}(X | \mathcal{G}) - U]^2) = 0$ if and only if $\mathbb{P}[U = \mathbb{E}(X | \mathcal{G})] = 1$

Conditional Variance

Once again, we assume that \mathcal{G} is a sub σ -algebra of \mathcal{F} and that all random variables mentioned are real valued, unless otherwise noted. It's natural to define the conditional variance of a random variable given \mathcal{G} in the same way as ordinary variance, but with all expected values conditioned on \mathcal{G} .

Suppose that X is a random variable with $\mathbb{E}(X^2) < \infty$. The *conditional variance* of X given \mathcal{G} is

$$\text{var}(X | \mathcal{G}) = \mathbb{E} \left([X - \mathbb{E}(X | \mathcal{G})]^2 \middle| \mathcal{G} \right) \quad (4.10.27)$$

Like all conditional expected values relative to \mathcal{G} , $\text{var}(X | \mathcal{G})$ is a random variable that is measurable with respect to \mathcal{G} and is unique up to equivalence. The first property is analogous to the computational formula for ordinary variance.

Suppose again that X is a random variable with $\mathbb{E}(X^2) < \infty$. Then

$$\text{var}(X | \mathcal{G}) = \mathbb{E}(X^2 | \mathcal{G}) - [\mathbb{E}(X | \mathcal{G})]^2 \quad (4.10.28)$$

Proof

Expanding the square in the definition and using basic properties of conditional expectation, we have

$$\text{var}(X | \mathcal{G}) = \mathbb{E}(X^2 - 2X\mathbb{E}(X | \mathcal{G}) + [\mathbb{E}(X | \mathcal{G})]^2 | \mathcal{G}) = \mathbb{E}(X^2 | \mathcal{G}) - 2\mathbb{E}[X\mathbb{E}(X | \mathcal{G}) | \mathcal{G}] + \mathbb{E}([\mathbb{E}(X | \mathcal{G})]^2 | \mathcal{G}) \quad (4.10.29)$$

$$= \mathbb{E}(X^2 | \mathcal{G}) - 2\mathbb{E}(X | \mathcal{G})\mathbb{E}(X | \mathcal{G}) + [\mathbb{E}(X | \mathcal{G})]^2 = \mathbb{E}(X^2 | \mathcal{G}) - [\mathbb{E}(X | \mathcal{G})]^2 \quad (4.10.30)$$

Next is a formula for the ordinary variance in terms of conditional variance and expected value.

Suppose again that X is a random variable with $\mathbb{E}(X^2) < \infty$. Then

$$\text{var}(X) = \mathbb{E}[\text{var}(X | \mathcal{G})] + \text{var}[\mathbb{E}(X | \mathcal{G})] \quad (4.10.31)$$

Proof

From the previous theorem and properties of conditional expected value we have $\mathbb{E}[\text{var}(X | \mathcal{G})] = \mathbb{E}(X^2) - \mathbb{E}([\mathbb{E}(X | \mathcal{G})]^2)$. But $\mathbb{E}(X^2) = \text{var}(X) + [\mathbb{E}(X)]^2$ and similarly, $\mathbb{E}([\mathbb{E}(X | \mathcal{G})]^2) = \text{var}[\mathbb{E}(X | \mathcal{G})] + (\mathbb{E}[\mathbb{E}(X | \mathcal{G})])^2$. But also, $\mathbb{E}[\mathbb{E}(X | \mathcal{G})] = \mathbb{E}(X)$ so substituting we get $\mathbb{E}[\text{var}(X | \mathcal{G})] = \text{var}(X) - \text{var}[\mathbb{E}(X | \mathcal{G})]$.

So the variance of X is the expected conditional variance plus the variance of the conditional expected value. This result is often a good way to compute $\text{var}(X)$ when we know the conditional distribution of X given \mathcal{G} . In turn, this property leads to a formula for the mean square error when $\mathbb{E}(X | \mathcal{G})$ is thought of as a predictor of X .

Suppose again that X is a random variable with $\mathbb{E}(X^2) < \infty$.

$$\mathbb{E}([X - \mathbb{E}(X | \mathcal{G})]^2) = \text{var}(X) - \text{var}[\mathbb{E}(X | \mathcal{G})] \quad (4.10.32)$$

Proof

From the [definition](#) and from the [mean property](#) and [variance formula](#),

$$\mathbb{E}([X - \mathbb{E}(X | \mathcal{G})]^2) = \mathbb{E}[\text{var}(X | \mathcal{G})] = \text{var}(X) - \text{var}[\mathbb{E}(X | \mathcal{G})] \quad (4.10.33)$$

Let us return to the study of predictors of the real-valued random variable X , and compare them in terms of mean square error.

Suppose again that X is a random variable with $\mathbb{E}(X^2) < \infty$.

1. The best constant predictor of X is $\mathbb{E}(X)$ with mean square error $\text{var}(X)$.
2. If Y is another random variable with $\mathbb{E}(Y^2) < \infty$, then the best predictor of X among linear functions of Y is

$$L(X | Y) = \mathbb{E}(X) + \frac{\text{cov}(X, Y)}{\text{var}(Y)}[Y - \mathbb{E}(Y)] \quad (4.10.34)$$

with mean square error $\text{var}(X)[1 - \text{cor}^2(X, Y)]$.

3. If Y is a (general) random variable, then the best predictor of X among all real-valued functions of Y with finite variance is $\mathbb{E}(X | Y)$ with mean square error $\text{var}(X) - \text{var}[\mathbb{E}(X | Y)]$.
4. If \mathcal{G} is a sub σ -algebra of \mathcal{F} , then the best predictor of X among random variables with finite variance that are measurable with respect to \mathcal{G} is $\mathbb{E}(X | \mathcal{G})$ with mean square error $\text{var}(X) - \text{var}[\mathbb{E}(X | \mathcal{G})]$.

Of course, (a) is a special case of (d) with $\mathcal{G} = \{\emptyset, \Omega\}$ and (c) is a special case of (d) with $\mathcal{G} = \sigma(Y)$. Only (b), the linear case, cannot be interpreted in terms of conditioning with respect to a σ -algebra.

Conditional Covariance

Suppose again that \mathcal{G} is a sub σ -algebra of \mathcal{F} . The conditional covariance of two random variables is defined like the ordinary covariance, but with all expected values conditioned on \mathcal{G} .

Suppose that X and Y are random variables with $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$. The *conditional covariance* of X and Y given \mathcal{G} is defined as

$$\text{cov}(X, Y | \mathcal{G}) = \mathbb{E} \left([X - \mathbb{E}(X | \mathcal{G})][Y - \mathbb{E}(Y | \mathcal{G})] \mid \mathcal{G} \right) \quad (4.10.35)$$

So $\text{cov}(X, Y | \mathcal{G})$ is a random variable that is measurable with respect to \mathcal{G} and is unique up to equivalence. As should be the case, conditional covariance generalizes conditional variance.

Suppose that X is a random variable with $\mathbb{E}(X^2) < \infty$. Then $\text{cov}(X, X | \mathcal{G}) = \text{var}(X | \mathcal{G})$.

Proof

This follows immediately from the two definitions.

Our next result is a computational formula that is analogous to the one for standard covariance—the covariance is the mean of the product minus the product of the means, but now with all expected values conditioned on \mathcal{G} :

Suppose again that X and Y are random variables with $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$. Then

$$\text{cov}(X, Y | \mathcal{G}) = \mathbb{E}(XY | \mathcal{G}) - \mathbb{E}(X | \mathcal{G})\mathbb{E}(Y | \mathcal{G}) \quad (4.10.36)$$

Proof

Expanding the product in the definition and using basic properties of conditional expectation, we have

$$\text{cov}(X, Y | \mathcal{G}) = \mathbb{E} \left(XY - X\mathbb{E}(Y | \mathcal{G}) - Y\mathbb{E}(X | \mathcal{G}) + \mathbb{E}(X | \mathcal{G})\mathbb{E}(Y | \mathcal{G}) \mid \mathcal{G} \right) = \mathbb{E}(XY | \mathcal{G}) - \mathbb{E}[X\mathbb{E}(Y | \mathcal{G}) | \mathcal{G}] - \mathbb{E}[Y\mathbb{E}(X | \mathcal{G}) | \mathcal{G}] \quad (4.10.37)$$

$$= \mathbb{E}(XY | \mathcal{G}) - \mathbb{E}(X | \mathcal{G})\mathbb{E}(Y | \mathcal{G}) - \mathbb{E}(X | \mathcal{G})\mathbb{E}(Y | \mathcal{G}) + \mathbb{E}(X | \mathcal{G})\mathbb{E}(Y | \mathcal{G}) = \mathbb{E}(XY | \mathcal{G}) - \mathbb{E}(X | \mathcal{G})\mathbb{E}(Y | \mathcal{G}) \quad (4.10.38)$$

Our next result shows how to compute the ordinary covariance of X and Y by *conditioning on X* .

Suppose again that X and Y are random variables with $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$. Then

$$\text{cov}(X, Y) = \mathbb{E}[\text{cov}(X, Y | \mathcal{G})] + \text{cov}[\mathbb{E}(X | \mathcal{G}), \mathbb{E}(Y | \mathcal{G})] \quad (4.10.39)$$

Proof

From (29) and properties of conditional expected value we have

$$\mathbb{E}[\text{cov}(X, Y | \mathcal{G})] = \mathbb{E}(XY) - \mathbb{E}[\mathbb{E}(X | \mathcal{G})\mathbb{E}(Y | \mathcal{G})] \quad (4.10.40)$$

But $\mathbb{E}(XY) = \text{cov}(X, Y) + \mathbb{E}(X)\mathbb{E}(Y)$ and similarly,

$$\mathbb{E}[\mathbb{E}(X | \mathcal{G})\mathbb{E}(Y | \mathcal{G})] = \text{cov}[\mathbb{E}(X | \mathcal{G}), \mathbb{E}(Y | \mathcal{G})] + \mathbb{E}[\mathbb{E}(X | \mathcal{G})]\mathbb{E}[\mathbb{E}(Y | \mathcal{G})] \quad (4.10.41)$$

But also, $\mathbb{E}[\mathbb{E}(X | \mathcal{G})] = \mathbb{E}(X)$ and $\mathbb{E}[\mathbb{E}(Y | \mathcal{G})] = \mathbb{E}(Y)$ so substituting we get

$$\mathbb{E}[\text{cov}(X, Y | \mathcal{G})] = \text{cov}(X, Y) - \text{cov}[\mathbb{E}(X | \mathcal{G}), \mathbb{E}(Y | \mathcal{G})] \quad (4.10.42)$$

Thus, the covariance of X and Y is the expected conditional covariance plus the covariance of the conditional expected values. This result is often a good way to compute $\text{cov}(X, Y)$ when we know the conditional distribution of (X, Y) given \mathcal{G} .

This page titled [4.10: Conditional Expected Value Revisited](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.