Conditional Expectation

Andreas Klappenecker

Texas A&M University

We are going to define the conditional expectation of a random variable given

1. an event,
2. another random variable,
3. a σ-algebra.

Conditional expectations can be convenient in some computations.

Conditional Expectation given an Event

Definition

The **conditional expectation** of a discrete random variable X given an event A is denoted as $\mathrm{E}[X \mid A]$ and is defined by

$$\mathrm{E}[X \mid A] = \sum_x x \Pr[X = x \mid A].$$

It follows that

$$\mathrm{E}[X \mid A] = \sum_x x \Pr[X = x \mid A] = \sum_x x \frac{\Pr[X = x \text{ and } A]}{\Pr[A]}.$$

Example

Problem

Suppose that X *and* Y *are discrete random variables with values in* $\{1, 2\}$ *s.t.*

$$\Pr[X = 1, Y = 1] = \tfrac{1}{2}, \quad \Pr[X = 1, Y = 2] = \tfrac{1}{10},$$
$$\Pr[X = 2, Y = 1] = \tfrac{1}{10}, \quad \Pr[X = 2, Y = 2] = \tfrac{3}{10}.$$

Calculate $\mathrm{E}[X \mid Y = 1]$.

Example

Problem

Suppose that X and Y are discrete random variables with values in $\{1, 2\}$ s.t.

$$\Pr[X = 1, Y = 1] = \tfrac{1}{2}, \quad \Pr[X = 1, Y = 2] = \tfrac{1}{10},$$
$$\Pr[X = 2, Y = 1] = \tfrac{1}{10}, \quad \Pr[X = 2, Y = 2] = \tfrac{3}{10}.$$

Calculate $\mathrm{E}[X \mid Y = 1]$.

By definition

$$\begin{aligned}\mathrm{E}[X \mid Y = 1] &= 1 \Pr[X = 1 \mid Y = 1] + 2 \Pr[X = 2 \mid Y = 1].\\ &= 1 \frac{\Pr[X = 1, Y = 1]}{\Pr[Y = 1]} + 2 \frac{\Pr[X = 2, Y = 1]}{\Pr[Y = 1]}.\end{aligned}$$

Example

Problem

Suppose that X and Y are discrete random variables with values in $\{1, 2\}$ s.t.

$$\Pr[X = 1, Y = 1] = \tfrac{1}{2}, \quad \Pr[X = 1, Y = 2] = \tfrac{1}{10},$$
$$\Pr[X = 2, Y = 1] = \tfrac{1}{10}, \quad \Pr[X = 2, Y = 2] = \tfrac{3}{10}.$$

Example

Problem

Suppose that X and Y are discrete random variables with values in $\{1, 2\}$ s.t.

$$\Pr[X = 1, Y = 1] = \tfrac{1}{2}, \quad \Pr[X = 1, Y = 2] = \tfrac{1}{10},$$
$$\Pr[X = 2, Y = 1] = \tfrac{1}{10}, \quad \Pr[X = 2, Y = 2] = \tfrac{3}{10}.$$

We have $\Pr[Y = 1] = \Pr[X = 1, Y = 1] + \Pr[X = 2, Y = 1] = \frac{1}{2} + \frac{1}{10} = \frac{3}{5}$.

Example

Problem

Suppose that X and Y are discrete random variables with values in $\{1, 2\}$ s.t.

$$\Pr[X = 1, Y = 1] = \tfrac{1}{2}, \quad \Pr[X = 1, Y = 2] = \tfrac{1}{10},$$
$$\Pr[X = 2, Y = 1] = \tfrac{1}{10}, \quad \Pr[X = 2, Y = 2] = \tfrac{3}{10}.$$

We have $\Pr[Y = 1] = \Pr[X = 1, Y = 1] + \Pr[X = 2, Y = 1] = \frac{1}{2} + \frac{1}{10} = \frac{3}{5}$.

$$\begin{aligned} E[X \mid Y = 1] &= 1\frac{\Pr[X = 1, Y = 1]}{\Pr[Y = 1]} + 2\frac{\Pr[X = 2, Y = 1]}{\Pr[Y = 1]} \\ &= 1\frac{1/2}{3/5} + 2\frac{1/10}{3/5} = \frac{5}{6} + 2\frac{1}{6} = \frac{7}{6} \end{aligned}$$

Interpretation

Let $\mathcal{F} = 2^\Omega$ with Ω finite. For a random variable X and an event A, we can interpret $E[X \mid A]$ as the average of $X(\omega)$ over all $\omega \in A$.

Indeed, we have

$$E[X|A] = \sum_x x \Pr[X = x \mid A] = \sum_x x \frac{Pr[X = x \text{ and } A]}{\Pr[A]}$$
$$= \sum_{\omega \in A} X(\omega) \frac{\Pr[\omega]}{\Pr[A]}.$$

Interpretation

Let $\mathcal{F} = 2^{\Omega}$ with Ω finite. For a random variable X and an event A, we can interpret $E[X \mid A]$ as the average of $X(\omega)$ over all $\omega \in A$.

Indeed, we have

$$E[X|A] = \sum_x x \Pr[X = x \mid A] = \sum_x x \frac{Pr[X = x \text{ and } A]}{\Pr[A]}$$
$$= \sum_{\omega \in A} X(\omega) \frac{\Pr[\omega]}{\Pr[A]}.$$

Caveat

This interpretation does not work for all random variables, but it gives a better understanding of the meaning of $E[X \mid A]$.

Conditional Expectation

Proposition

We have

$$\mathrm{E}[X \mid A] = \frac{\mathrm{E}[X\, I_A]}{\Pr[A]}.$$

Proof.

As we have seen,

$$\mathrm{E}[X|A] = \sum_x x \frac{Pr[X = x \text{ and } A]}{\Pr[A]} = \frac{1}{\Pr[A]} \sum_x x Pr[X = x \text{ and } A].$$

We can rewrite the latter expression in the form

$$\mathrm{E}[X|A] = \frac{\mathrm{E}[X\, I_A]}{\Pr[A]}. \qquad \square$$

Definition

The **conditional expectation** $\mathrm{E}[X \mid A]$ of an arbitrary random variable X given an event A is defined by

$$\mathrm{E}[X|A] = \begin{cases} \dfrac{\mathrm{E}[X\, I_A]}{\Pr[A]} & \text{if } \Pr[A] > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Properties

Linearity

Proposition

If a and b are real numbers and X and Y are random variables, then

$$\mathrm{E}[aX + bY \mid A] = a\mathrm{E}[X \mid A] + b\mathrm{E}[Y \mid A].$$

Proof.

$$\begin{aligned}
\mathrm{E}[aX + bY \mid A] &= \frac{\mathrm{E}[(aX + bY)\, I_A]}{\Pr[A]} \\
&= a\frac{\mathrm{E}[X\, I_A]}{\Pr[A]} + b\frac{\mathrm{E}[Y\, I_A]}{\Pr[A]} \\
&= a\mathrm{E}[X \mid A] + b\mathrm{E}[Y \mid A].
\end{aligned}$$

□

Proposition

If X and Y are independent discrete random variables, then

$$E[Y \mid X = x] = E[Y].$$

Proof.

By definition,

$$\begin{aligned} E[Y \mid X = x] &= \sum_y y \Pr[Y = y \mid X = x] \\ &= \sum_y y \Pr[Y = y] = E[Y]. \end{aligned}$$

□

Important Application

Computing Expectations

We can compute the expected value of X as a sum of conditional expectations. This is similar to the law of total probability.

Proposition

If X and Y are discrete random variables, then

$$\mathrm{E}[X] = \sum_{y} \mathrm{E}[X \mid Y = y] \Pr[Y = y].$$

Proposition

If X and Y are discrete random variables, then

$$\mathrm{E}[X] = \sum_y \mathrm{E}[X \mid Y = y] \Pr[Y = y].$$

Proof.

$$\begin{aligned}
\sum_y \mathrm{E}[X \mid Y = y] \Pr[Y = y] &= \sum_y \left(\sum_x x \Pr[X = x | Y = y] \right) \Pr[Y = y] \\
&= \sum_x \sum_y x \Pr[X = x | Y = y] \Pr[Y = y] \\
&= \sum_x \sum_y x \Pr[X = x, Y = y] \\
&= \sum_x x \Pr[X = x] = \mathrm{E}[X]
\end{aligned}$$

$\square$

Why We Need More than One Type of Conditional Expectation

Conditional Expectation

We can also define conditional expectations for continuous random variables.

Definition

The conditional expectation of a discrete random variable Y given that $X = x$ is defined as

$$\mathrm{E}[Y \mid X = x] = \sum_{y} y \Pr[Y = y \mid X = x].$$

The conditional expectation of a continuous random variable Y given that $X = x$ is defined as

$$\mathrm{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy,$$

We assume absolute convergence in each case.

Problem

A stick of length one is broken at a random point, uniformly distributed over the stick. The remaining piece is broken once more.

Find the expected value of the piece that now remains.

Motivating Example

Let X denote the random variable giving the length of the first remaining piece. Then X is uniformly distributed over the unit interval $(0, 1)$.

Let Y denote the random variable giving the length of the second remaining piece. Then Y is uniformly distributed over the shorter interval $(0, X)$.

Motivating Example: Interpretation

Given that $X = x$, the random variable Y is uniformly distributed over the interval $(0, x)$. In other words,

$$Y \mid X = x$$

has the density function

$$f_{Y|X=x}(y) = \frac{1}{x}$$

for all y in $(0, x)$.

Motivating Example: Expectation

For a random variable Z that is uniformly distributed on the interval (a, b), we have

$$\begin{aligned} \mathrm{E}[Z] &= \int_a^b x \frac{1}{b-a} \, dx = \frac{1}{b-a} \frac{1}{2} x^2 \Big|_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}. \end{aligned}$$

Motivating Example: Expectation

For a random variable Z that is uniformly distributed on the interval (a, b), we have

$$
\begin{aligned}
\mathrm{E}[Z] &= \int_a^b x \frac{1}{b-a}\, dx = \frac{1}{b-a} \frac{1}{2} x^2 \Big|_a^b \\
&= \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}.
\end{aligned}
$$

Example

Since the random variable X is uniformly distributed over the interval $(0, 1)$, we have

$$\mathrm{E}[X] = \frac{1+0}{2} = \frac{1}{2}.$$

Motivating Example

Example

Since $Y|X = x$ is uniformly distributed over $(0, x)$, we get

$$\mathrm{E}[Y \mid X = x] = \int_0^x y\frac{1}{x}\,dy = \frac{x+0}{2} = \frac{x}{2}.$$

Example

Since $Y|X = x$ is uniformly distributed over $(0, x)$, we get

$$\mathrm{E}[Y \mid X = x] = \int_0^x y \frac{1}{x} \, dy = \frac{x + 0}{2} = \frac{x}{2}.$$

Does this solve the problem?

Now we know the expected length of the second remaining piece, **given** that we know the length x of the first remaining piece of the stick.

Motivating Example

We can also define a random variable $\mathrm{E}[Y \mid X]$ that satisfies

$$\mathrm{E}[Y \mid X](\omega) = \mathrm{E}[Y \mid X = X(\omega)].$$

We expect that

$$\mathrm{E}[\mathrm{E}[Y \mid X]] = \mathrm{E}[X/2] = \frac{1}{4}.$$

Now this solves the problem. The expected length of the remaining piece is $1/4$ of the length of the stick.

Conditional Expectation given a Random Variable

Motivation

Question

How should we think about $\mathrm{E}[X \mid Y]$?

Answer

Suppose that Y is a discrete random variable. If we **observe** one of the values y of Y, then the conditional expectation should be given by

$$\mathrm{E}[X \mid Y = y].$$

If we **do not know** the value y of Y, then we need to contend ourselves with the possible expectations

$$\mathrm{E}[X \mid Y = y_1], \quad \mathrm{E}[X \mid Y = y_2], \quad \mathrm{E}[X \mid Y = y_2], \ldots$$

So $\mathrm{E}[X \mid Y]$ should be a $\sigma(Y)$-measurable random variable itself.

Definition

Let X and Y be two discrete random variables.

The **conditional expectation** $\mathrm{E}[X \mid Y]$ of X given Y is the random variable defined by

$$\mathrm{E}[X \mid Y](\omega) = \mathrm{E}[X \mid Y = Y(\omega)].$$

Definition

Let X and Y be two discrete random variables.

The **conditional expectation** $E[X \mid Y]$ of X given Y is the random variable defined by

$$E[X \mid Y](\omega) = E[X \mid Y = Y(\omega)].$$

Caveat

Sometimes $E[X \mid Y]$ is defined differently as a $\mathcal{B}(\mathbf{R})$-measurable function $y \mapsto E[X \mid Y = y]$. We prefer to think about $E[X \mid Y]$ as a function $\Omega \to \mathbf{R}$. The two definitions are obviously not equivalent. Our choice generalizes nicely.

A Pair of Fair Coin Flips

Example

Suppose that X and Y are random variables describing independent fair coin flips with values 0 and 1. Then the sample space of (X, Y) is given by

$$\Omega = \{(0,0),(0,1),(1,0),(1,1)\}.$$

Let Z denote the random variable $Z = X + Y$. Then we have

$$Z(0,0) = 0, \quad Z(0,1) = 1, \quad Z(1,0) = 1, \quad Z(1,1) = 2.$$

A Pair of Fair Coin Flips

Example (Continued.)

Suppose that we want to know $\mathrm{E}[Z \mid X]$. We calculate

$$\mathrm{E}[Z \mid X = 0] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2},$$
$$\mathrm{E}[Z \mid X = 1] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}.$$

Then

$$\mathrm{E}[Z \mid X](0,0) = \tfrac{1}{2}, \quad \mathrm{E}[Z \mid X](0,1) = \tfrac{1}{2},$$
$$\mathrm{E}[Z \mid X](1,0) = \tfrac{3}{2}, \quad \mathrm{E}[Z \mid X](1,1) = \tfrac{3}{2}.$$

Example (Continued.)

Suppose that we now want to know $E[Z \mid Y]$. We calculate

$$\begin{aligned} E[Z \mid Y = 0] &= 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}, \\ E[Z \mid Y = 1] &= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}. \end{aligned}$$

Then

$$\begin{aligned} E[Z \mid Y](0,0) &= \tfrac{1}{2}, & E[Z \mid Y](0,1) &= \tfrac{3}{2}, \\ E[Z \mid Y](1,0) &= \tfrac{1}{2}, & E[Z \mid Y](1,1) &= \tfrac{3}{2}. \end{aligned}$$

Example (Continued.)

Suppose that we now want to know $\mathrm{E}[X \mid Z]$. We calculate

$$\begin{aligned}
\mathrm{E}[X \mid Z = 0] &= 0 \\
\mathrm{E}[X \mid Z = 1] &= 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}, \\
\mathrm{E}[X \mid Z = 2] &= 1
\end{aligned}$$

Then

$$\mathrm{E}[X \mid Z](0,0) = 0, \quad \mathrm{E}[X \mid Z](0,1) = \tfrac{1}{2},$$
$$\mathrm{E}[X \mid Z](1,0) = \tfrac{1}{2}, \quad \mathrm{E}[X \mid Z](1,1) = 1.$$

Properties of the Conditional Expectation

Proposition

If X is a function of Y, then $\mathrm{E}[X \mid Y] = X$.

Proof.

Suppose that $X = f(Y)$. Then

$$\begin{aligned} \mathrm{E}[X \mid Y](\omega) &= \mathrm{E}[X \mid Y = Y(\omega)] \\ &= \mathrm{E}[f(Y(\omega)) \mid Y = Y(\omega)] \\ &= f(Y(\omega)) = X(\omega). \end{aligned}$$

$\square$

Proposition

If X and Y are independent, then $\mathrm{E}[X \mid Y] = \mathrm{E}[X]$.

Proof.

For all ω in Ω, we have

$$\mathrm{E}[X \mid Y](\omega) = \mathrm{E}[X \mid Y = Y(\omega)] = \mathrm{E}[X]. \qquad \square$$

Linearity

Proposition

If a and b are real numbers and X, Y, and Z discrete random variables, then

$$\mathrm{E}[aX + bY \mid Z] = a\mathrm{E}[X \mid Z] + b\mathrm{E}[Y \mid Z].$$

A Pair of Fair Coin Flips

Example

Suppose that X and Y are independent random variables describing fair coin flips with values 0 and 1. Let $Z = X + Y$. We determined $\mathrm{E}[Z|X]$, but it was a bit cumbersome. Here is an easier way:

$$\begin{aligned} \mathrm{E}[Z \mid X] &= \mathrm{E}[X + Y \mid X] \text{ by definition} \\ &= \mathrm{E}[X \mid X] + \mathrm{E}[Y \mid X] \text{ by linearity} \\ &= X + \mathrm{E}[Y] \text{ by function and by independence} \\ &= X + \frac{1}{2}. \end{aligned}$$

Proposition

$$\mathrm{E}[\mathrm{E}[X \mid Y]] = \mathrm{E}[X].$$

Proof.

$$\begin{aligned}
\mathrm{E}[\mathrm{E}[X \mid Y]] &= \sum_y \mathrm{E}[\,\mathrm{E}[X \mid Y] | Y = y] \Pr[Y = y] \\
&= \sum_y \mathrm{E}[X \mid Y = y] \Pr[Y = y] \\
&= \mathrm{E}[X]
\end{aligned}$$

$\square$

Applications

Wald's Theorem

Theorem

Suppose that $X_1, X_2, \ldots$ *are independent random variables, all with the same mean. Suppose that* N *is a nonnegative, integer-valued random variable that is independent of the* X_i*'s. If* $\mathrm{E}[X_1] < \infty$ *and* $\mathrm{E}[N] < \infty$*, then*

$$\mathrm{E}\left[\sum_{k=1}^{N} X_i\right] = \mathrm{E}[N]\mathrm{E}[X_1].$$

Proof.

By double expectation, we have

$$
\begin{aligned}
\mathrm{E}\left[\sum_{k=1}^{N} X_i\right] &= \mathrm{E}\left[\mathrm{E}\left[\sum_{k=1}^{N} X_i \middle| N\right]\right] \\
&= \sum_{n=1}^{\infty} \mathrm{E}\left[\sum_{k=1}^{N} X_i \middle| N = n\right] \Pr[N = n] \\
&= \sum_{n=1}^{\infty} \mathrm{E}\left[\sum_{k=1}^{n} X_i \middle| N = n\right] \Pr[N = n]
\end{aligned}
$$

Proof. (Continued)

$$
\begin{aligned}
\mathrm{E}\left[\sum_{k=1}^{N} X_i\right] &= \sum_{n=1}^{\infty} \mathrm{E}\left[\sum_{k=1}^{n} X_i \middle| N = n\right] \Pr[N = n] \\
&= \sum_{n=1}^{\infty} \mathrm{E}\left[\sum_{k=1}^{n} X_i\right] \Pr[N = n] \\
&= \sum_{n=1}^{\infty} n\mathrm{E}\left[X_1\right] \Pr[N = n] \\
&= \mathrm{E}[X_1] \sum_{n=1}^{\infty} n \Pr[N = n] = \mathrm{E}[X_1]\mathrm{E}[N]. \square
\end{aligned}
$$

Example

Suppose that we roll a navy die. The face value N of the die ranges from 1 to 6. Depending on the face value of the navy die, we roll N ivory dice and sum their values.

On average, what is the resulting value of the sum face values of the N ivory dice?

Dice: Solution

Solution

Let $X_1, \ldots, X_6$ denote the random variables describing the face values of the ivory dice. By Wald's theorem, we have

$$
\begin{aligned}
\mathrm{E}\left[\sum_{k=1}^{N} X_i\right] &= \mathrm{E}[N]\mathrm{E}[X_1] \\
&= \left(\frac{1+2+3+4+5+6}{6}\right)\left(\frac{1+2+3+4+5+6}{6}\right) \\
&= \left(\frac{7}{2}\right)^2 = \frac{49}{4} = 12.25
\end{aligned}
$$

Conditional Expectation Given a σ-Algebra

Motivation

Suppose that a sample space Ω is partitioned into measurable sets

$$B_1, B_2, \ldots, B_n.$$

We know know the expectation of a random variable X given that one of the events B_k will happen, but we do not know which one.

We want to form a conditional expectation $\mathrm{E}[X \mid \mathcal{G}]$ with $\mathcal{G} = \sigma(B_1, B_2, \ldots, B_n)$ such that

$$\mathrm{E}[X \mid \mathcal{G}](\omega) = \mathrm{E}[X \mid B_k] = \frac{\mathrm{E}[X\, I_{B_k}]}{\Pr[B_k]}$$

for $\omega \in B_k$. Then $\mathrm{E}[\,\mathrm{E}[X \mid \mathcal{G}]\,] = \mathrm{E}[X]$.

Definition

Let $\mathcal{F}$ be a σ-algebra with sub-σ-algebra $\mathcal{G}$. A random variable Y is called a **conditional expectation** of X given $\mathcal{G}$, written $Y = \mathrm{E}[X \mid \mathcal{G}]$ if and only if

1. Y is $\mathcal{G}$-measurable
2. $\mathrm{E}[Y I_G] = \mathrm{E}[X I_G]$ for all $G \in \mathcal{G}$.

Example

Let A and B be events with $0 < \Pr[A] < 1$. If we define $\mathcal{G} = \sigma(B)$, then $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$. Then

$$\mathrm{E}[X \mid \mathcal{G}] = \frac{\mathrm{E}[X\, I_B]}{\Pr[B]} I_B + \frac{\mathrm{E}[X\, I_{B^c}]}{\Pr[B^c]} I_{B^c}.$$

Indeed, the right-hand side is clearly $\mathcal{G}$-measurable. We have

$$\mathrm{E}[\,\mathrm{E}[X \mid \mathcal{G}] I_B\,] = \mathrm{E}[X\, I_B]$$

and

$$\mathrm{E}[\,\mathrm{E}[X \mid \mathcal{G}] I_{B^c}\,] = \mathrm{E}[X\, I_{B^c}].$$

Interpretation

We would like to think of $\mathrm{E}[X \mid \mathcal{G}]$ as the average of $X(\omega)$ over all ω which is consistent with the information encoded in $\mathcal{G}$.

Example

Suppose that $(\Omega, \mathcal{F}, \Pr)$ is a probability space with $\Omega = \{a, b, c, d, e, f\}$, $\mathcal{F} = 2^{\Omega}$, and Pr uniform. Define a random variable X by

$$X(a) = 1,\ X(b) = 3,\ X(c) = 3,\ X(d) = 5,\ X(e) = 5,\ X(f) = 7.$$

Suppose that another random variable Z is given by

$$Z(a) = 3,\ Z(b) = 3,\ Z(c) = 3,\ Z(d) = 3,\ Z(e) = 2,\ Z(f) = 2.$$

We want to determine $\mathrm{E}[X \mid \mathcal{G}]$ with $\mathcal{G} = \sigma(Z)$.

Example

Since

$$Z(a) = 3,\ Z(b) = 3,\ Z(c) = 3,\ Z(d) = 3,\ Z(e) = 2,\ Z(f) = 2,$$

the σ-algebra $\sigma(Z)$ is generated by the event $Z^{-1}(3)$ and its complement

$$Z^{-1}(3) = \{a, b, c, d\} \quad \text{and} \quad Z^{-1}(2) = \{e, f\}.$$

Example

Now consider X on $Z^{-1}(3) = \{a, b, c, d\}$ and its complement

$$X(a) = 1,\ X(b) = 3,\ X(c) = 3,\ X(d) = 5,\ X(e) = 5,\ X(f) = 7.$$

Since the distribution is uniform, we have

$$\mathrm{E}[X \mid \sigma(Z)](\omega) = \begin{cases} 3 & \text{if } \omega \in \{a, b, c, d\}, \\ 6 & \text{if } \omega \in \{e, f\} \end{cases}$$

Example

Suppose that $\mathcal{G}$ is generated by a finite partition

$$B_1, B_2, \ldots, B_n$$

of the sample space Ω. Then

$$\mathrm{E}[X \mid \mathcal{G}](\omega) = \sum_{k=1}^{n} a_k I_{B_k},$$

where

$$a_k = \frac{\mathrm{E}[X\, I_{B_k}]}{\Pr[B_k]} = \mathrm{E}[X \mid B_k].$$

Example (Continued.)

If

$$\mathrm{E}[X \mid \mathcal{G}] = \sum_{k=1}^{n} \frac{\mathrm{E}[X\, I_{B_k}]}{\Pr[B_k]} I_{B_k},$$

then it is certainly $\mathcal{G}$-measurable and

$$\mathrm{E}[\,\mathrm{E}[X \mid \mathcal{G}]\,] I_{B_k}\,] = \mathrm{E}[X\, I_{B_k}].$$

Therefore,

$$\mathrm{E}[\mathrm{E}[X \mid \mathcal{G}]\,] = \sum_{k=1}^{n} \mathrm{E}[X\, I_{B_k}] = \mathrm{E}[X\, I_{\Omega}] = \mathrm{E}[X].$$

Definition

Let $\mathcal{F}$ be a σ-algebra with sub-σ-algebra $\mathcal{G}$. A random variable Y is called a **conditional expectation** of X given $\mathcal{G}$, written $Y = \mathrm{E}[X \mid \mathcal{G}]$ if and only if

1. Y is $\mathcal{G}$-measurable
2. $\mathrm{E}[Y I_G] = \mathrm{E}[X I_G]$ for all $G \in \mathcal{G}$.

Questions

1. Is the conditional expectation unique?
2. Does conditional expectation always exist?

Suppose that Y and Y' are $\mathcal{G}$-measurable random variables such that

$$\mathrm{E}[Y\, I_G] = \mathrm{E}[X\, I_G] = \mathrm{E}[Y'\, I_G]$$

holds for all $G \in \mathcal{G}$. Then $G = \{Y > Y'\}$ is an event in $\mathcal{G}$. We have

$$0 = \mathrm{E}[Y\, I_A] - \mathrm{E}[Y'\, I_A] = \mathrm{E}[(Y - Y')I_A].$$

Since $(Y - Y')I_A \geqslant 0$, we have $\Pr[A] = 0$.

We can conclude that $Y \leqslant Y'$ almost surely (meaning with probability 1). Similarly, $Y' \leqslant Y$ almost surely.

So $Y' = Y$ almost surely.

Existence (Sketch for those who know integration on measures)

Let $X^+ = \max\{X, 0\}$ and $X^- = X^+ - X$. We can define two finite measures on $(\Omega, \mathcal{F})$ by

$$Q^{\pm}(A) := \mathrm{E}[X^{\pm} I_A]$$

for all $A \in \mathcal{F}$.

If A satisfies $\Pr[A] = 0$, then $Q^{\pm}(A) = 0$.

Therefore, it follows from the Radon-Nikodym theorem that there exist densities $Y^{\pm}$ such that

$$Q^{\pm}(A) = \int_A Y^{\pm} d\Pr = \mathrm{E}[Y^{\pm} I_A].$$

Now define the conditional expectation by $Y = Y^+ - Y^-$.

Linearity

Proposition

$$\mathrm{E}[aX + bY \mid \mathcal{G}] = a\mathrm{E}[X \mid \mathcal{G}] + b\mathrm{E}[Y \mid \mathcal{G}].$$

Proof.

The right-hand side is $\mathcal{G}$-measurable by definition, hence, for $G \in \mathcal{G}$

$$\begin{aligned}
\mathrm{E}[I_G(a\mathrm{E}[X \mid \mathcal{G}] + b\mathrm{E}[Y \mid \mathcal{G}])] &= a\mathrm{E}[I_G\mathrm{E}[X \mid \mathcal{G}]] + b\mathrm{E}[I_G\mathrm{E}[Y \mid \mathcal{G}]] \\
&= a\mathrm{E}[I_GX] + b\mathrm{E}[I_GY] \\
&= \mathrm{E}[I_G(aX + bY)].
\end{aligned}$$

□

Proposition

If $X \geqslant Y$ *almost surely, then*

$$E[X \mid \mathcal{G}] \geqslant E[Y \mid \mathcal{G}].$$

Proof.

Let A denote the event $\{E[X \mid \mathcal{G}] < E[Y \mid \mathcal{G}]\} \in \mathcal{G}$.
Since we have $X \geqslant Y$, we get

$$E[I_A(X - Y)] \geqslant 0.$$

Therefore, $\Pr[A] = 0$. $\square$

For this proof, make sure that you understand what the event A encodes.

Products

Proposition

If $\mathsf{E}[|XY|] < \infty$ *and* Y *is* $\mathcal{G}$*-measurable, then*

$$\mathsf{E}[XY \mid \mathcal{G}] = Y\mathsf{E}[X \mid \mathcal{G}] \quad \textit{and} \quad \mathsf{E}[Y \mid \mathcal{G}] = \mathsf{E}[Y \mid Y] = Y.$$

The proof is a bit more involved.

Tower Property

Proposition

Let $\mathcal{G} \subseteq \mathcal{F} \subseteq \mathcal{A}$ be σ-algebras. Let X be an $\mathcal{A}$-measurable random variable. Then

$$\mathrm{E}[\mathrm{E}[X \mid \mathcal{F}] \mid \mathcal{G}] = \mathrm{E}[\mathrm{E}[X \mid \mathcal{G}] \mid \mathcal{F}] = \mathrm{E}[X \mid \mathcal{G}].$$

Proof.

The second equality follows from the product property with $X = 1$ and $Y = \mathrm{E}[X \mid \mathcal{G}]$, since Y is $\mathcal{F}$-measurable.

If $A \in \mathcal{G}$, then $A \in \mathcal{F}$ and

$$\begin{aligned} \mathrm{E}[I_A\, \mathrm{E}[\mathrm{E}[X \mid \mathcal{F}] \mid \mathcal{G}]] &= \mathrm{E}[I_A \mathrm{E}[X \mid \mathcal{F}]] \\ &= \mathrm{E}[I_A\, X] \\ &= \mathrm{E}[I_A\, \mathrm{E}[X \mid \mathcal{G}]]. \end{aligned}$$

□

Triangle Inequality

Proposition

$$\mathrm{E}[|X| \mid \mathcal{G}] \geqslant |\mathrm{E}[X \mid \mathcal{G}]|$$

Proposition

If $\sigma(X)$ *and* $\mathcal{G}$ *are independent* σ*-algebras, so*

$$\Pr[A \cap B] = \Pr[A] \Pr[B]$$

for all $A \in \sigma(X)$ *and* $B \in \mathcal{G}$*, then*

$$\mathrm{E}[X \mid \mathcal{G}] = \mathrm{E}[X].$$

Lack of Information

Proposition

If $\Pr[A] \in \{0, 1\}$ *for all* $A \in \mathcal{G}$*, then*

$$\mathrm{E}[X \mid \mathcal{G}] = \mathrm{E}[X].$$

Best Prediction

The conditional expectation $\mathsf{E}[X \mid \mathcal{G}]$ is supposed to be the "best" prediction one can make about X if we only have the information contained in σ-algebra $\mathcal{G}$.

Extremal Case 1

If $\sigma(X) \subseteq \mathcal{G}$, then

$$\mathsf{E}[X \mid \mathcal{G}] = X.$$

Extremal Case 2

If $\sigma(X)$ and $\mathcal{G}$ are independent, then

$$\mathsf{E}[X \mid \mathcal{G}] = \mathsf{E}[X].$$

Proposition

Let $\mathcal{G} \subseteq \mathcal{A}$ be σ-algebras. Let X be an $\mathcal{A}$-measurable random variable with $\mathrm{E}[X^2] < \infty$. Then for any $\mathcal{G}$-measurable random variable Y with $\mathrm{E}[Y^2] < \infty$, we have

$$\mathrm{E}[(X - Y)^2] \geqslant \mathrm{E}[(X - \mathrm{E}[X \mid \mathcal{G}])^2]$$

with equality if and only if $Y = \mathrm{E}[X \mid \mathcal{G}]$.