Advanced Probability and Applications (Part II)

Olivier Lévéque, IC–LTHI, EPFL

April 24, 2018

Contents

1 Conditional expectation

1.1 Conditioning with respect to an event $B \in \mathcal{F}$ ................................. 2
1.2 Conditioning with respect to a discrete random variable $Y$ ................................. 2
1.3 Conditioning with respect to a continuous random variable $Y$? ................................. 3
1.4 Conditioning with respect to a sub-$\sigma$-field $\mathcal{G}$ ................................. 3
1.5 Conditioning with respect to a random variable $Y$ ................................. 4
1 Conditional expectation

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space.

1.1 Conditioning with respect to an event \(B \in \mathcal{F}\)

The conditional probability of an event \(A \in \mathcal{F}\) given another event \(B \in \mathcal{F}\) is defined as

\[
P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{provided that } P(B) > 0
\]

Notice that if \(A\) and \(B\) are independent, then \(P(A|B) = P(A)\); the conditioning does not affect the probability. This fact remains true in more generality (see below).

In a similar manner, the conditional expectation of an integrable random variable \(X\) given \(B \in \mathcal{F}\) is defined as

\[
E(X|B) = \frac{E(X 1_B)}{P(B)}, \quad \text{provided that } P(B) > 0
\]

1.2 Conditioning with respect to a discrete random variable \(Y\)

Let us assume that the random variable \(Y\) (is \(\mathcal{F}\)-measurable and) takes values in a countable set \(C\).

\[
P(A|Y) = \varphi(Y), \quad \text{where } \varphi(y) = P(A|\{Y = y\}), \quad y \in C
\]

\[
E(X|Y) = \psi(Y), \quad \text{where } \psi(y) = E(X|\{Y = y\}), \quad y \in C
\]

If \(X\) is also a discrete random variable with values in \(C\), then

\[
E(X|Y) = \psi(Y), \quad \text{where } \psi(y) = \frac{E(X 1_{\{Y = y\}})}{P(\{Y = y\})} = \sum_{x \in C} x \frac{E(1_{\{X = x\} \cap \{Y = y\}})}{P(\{Y = y\})} = \sum_{x \in C} x P(\{X = x\}|\{Y = y\})
\]

**Important remark.** \(\varphi(y)\) and \(\psi(y)\) are functions, while \(\varphi(Y) = P(A|Y)\) and \(\psi(Y) = E(X|Y)\) are random variables. They both are functions of the outcome of the random variable \(Y\), that is, they are \(\sigma(Y)\)-measurable random variables.

**Example.** Let \(X_1, X_2\) be two independent dice rolls and let us compute \(E(X_1 + X_2|X_2) = \psi(X_2)\), where

\[
\psi(y) = E(X_1 + X_2|\{X_2 = y\}) = \frac{E((X_1 + X_2) 1_{\{X_2 = y\}})}{P(\{X_2 = y\})}
\]

\[
= \frac{E(X_1 1_{\{X_2 = y\}}) + E(X_2 1_{\{X_2 = y\}})}{P(\{X_2 = y\})} = \frac{E(X_1) E(1_{\{X_2 = y\}}) + E(y 1_{\{X_2 = y\}})}{P(\{X_2 = y\})}
\]

\[
= E(X_1) \frac{P(\{X_2 = y\}) + y P(\{X_2 = y\})}{P(\{X_2 = y\})} = E(X_1) + y
\]

where the independence assumption between \(X_1\) and \(X_2\) has been used in equality (a). So finally (as one would expect), \(E(X_1 + X_2|X_2) = E(X_1) + X_2\), which can be explained intuitively as follows: the expectation of \(X_1\) conditioned on \(X_2\) is nothing but the expectation of \(X_1\), as the outcome of \(X_2\) provides no information on the outcome of \(X_1\) (\(X_1\) and \(X_2\) being independent); on the other hand, the expectation of \(X_2\) conditioned on \(X_2\) is exactly \(X_2\), as the outcome of \(X_2\) is known.
1.3 Conditioning with respect to a continuous random variable \( Y \)?

In this case, one faces the following problem: if \( Y \) is a continuous random variable, \( P(\{ Y = y \}) = 0 \) for all \( y \in \mathbb{R} \). So a direct generalization of the above formulas to the continuous case is impossible at first sight. A possible solution to this problem is to replace the event \( \{ Y = y \} \) by \( \{ y \leq Y < y + \varepsilon \} \) and to take the limit \( \varepsilon \to 0 \) for the definition of conditional expectation. This actually works, but also leads to a paradox in the multidimensional setting (known as Borel’s paradox). In addition, some random variables are neither discrete, nor continuous. It turns out that the cleanest way to define conditional expectation in the general case is through \( \sigma \)-fields.

1.4 Conditioning with respect to a sub-\( \sigma \)-field \( G \)

In order to define the conditional expectation in the general case, one needs the following proposition.

**Proposition 1.1.** Let \((\Omega, \mathcal{F}, P)\) be a probability space, \( G \) be a sub-\( \sigma \)-field of \( \mathcal{F} \) and \( X \) be an integrable random variable on \((\Omega, \mathcal{F}, P)\). There exists then an integrable random variable \( Z \) such that

(i) \( Z \) is \( G \)-measurable,

(ii) \( E(ZU) = E(XU) \) for any random variable \( U \) \( G \)-measurable and bounded.

Moreover, if \( Z_1, Z_2 \) are two integrable random variables satisfying (i) and (ii), then \( Z_1 = Z_2 \) a.s.

**Definition 1.2.** The above random variable \( Z \) is called the conditional expectation of \( X \) given \( G \) and is denoted as \( E(\cdot|G) \). Because of the last part of the above proposition, it is defined up to a negligible set.

**Definition 1.3.** One further defines \( P(A|G) = E(1_A|G) \) for \( A \in \mathcal{F} \).

**Remark.** Notice that as before, both \( P(A|G) \) and \( E(X|G) \) are \( (G \)-measurable) random variables.

**Properties.** The above definition does not give a computation rule for the conditional expectation; it is only an existence theorem. The properties listed below will therefore be of help for computing conditional expectations. The proofs of the first two are omitted, while the next five are left as (important!) exercises.

- Linearity. \( E(cX + Y|G) = cE(X|G) + E(Y|G) \) a.s.

- Monotonicity. If \( X \geq Y \) a.s., then \( E(X|G) \geq E(Y|G) \) a.s. (so if \( X \geq 0 \) a.s., then \( E(X|G) \geq 0 \) a.s.)

- \( E(E(X|G)) = E(X) \).

- If \( X \) is independent of \( G \), then \( E(X|G) = E(X) \) a.s.

- If \( X \) is \( G \)-measurable, then \( E(X|G) = X \) a.s.

- If \( Y \) is \( G \)-measurable and bounded (or if \( Y \) is \( G \)-measurable and both \( X \) and \( Y \) are square-integrable; what actually matters here is that the random variable \( XY \) is integrable), then \( E(XY|G) = E(X|G)Y \) a.s.

- If \( H \) is a sub-\( \sigma \)-field of \( G \), then \( E(E(X|H)|G) = E(E(X|G)|H) = E(X|H) \) a.s. (in other words, the smallest \( \sigma \)-field always “wins”.)
Some of the above properties are illustrated below with an example.

**Example.** Let $\Omega = \{1, \ldots, 6\}$, $F = \mathcal{P}(\Omega)$ and $P(\{\omega\}) = \frac{1}{6}$ for $\omega = 1, \ldots, 6$ (the probability space of the die roll). Let also $X(\omega) = \omega$ be the outcome of the die roll and consider the two sub-$\sigma$-fields:

$$G = \sigma(\{1, 3\}, \{2\}, \{5\}, \{4, 6\}) \quad \text{and} \quad H = \sigma(\{1, 3, 5\}, \{2, 4, 6\})$$

Then $E(X) = 3.5$,

$$E(X|G)(\omega) = \begin{cases} 2 & \text{if } \omega \in \{1, 3\} \text{ or } \omega = 2 \\ 5 & \text{if } \omega \in \{4, 6\} \text{ or } \omega = 5 \end{cases} \quad \text{and} \quad E(X|H)(\omega) = \begin{cases} 3 & \text{if } \omega \in \{1, 3, 5\} \\ 4 & \text{if } \omega \in \{2, 4, 6\} \end{cases}$$

So $E(E(X|G)) = E(E(X|H)) = E(X)$. Moreover,

$$E(E(X|G)|H)(\omega) = \begin{cases} \frac{1}{3}(2 + 2 + 5) = 3 & \text{if } \omega \in \{1, 3, 5\} \\ \frac{1}{3}(2 + 5 + 5) = 4 & \text{if } \omega \in \{2, 4, 6\} \end{cases} = E(X|H)(\omega)$$

and

$$E(E(X|H)|G)(\omega) = \begin{cases} 3 & \text{if } \omega \in \{1, 3\} \text{ or } \omega = 5 \\ 4 & \text{if } \omega \in \{4, 6\} \text{ or } \omega = 2 \end{cases} = E(X|H)(\omega)$$

The proposition below (given here without proof) is an extension of some of the above properties.

**Proposition 1.4.** Let $G$ be a sub-$\sigma$-field of $F$, $X$, $Y$ be two random variables such that $X$ is independent of $G$ and $Y$ is $G$-measurable, and let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be a Borel-measurable function such that $E(|\varphi(X,Y)|) < \infty$. Then

$$E(\varphi(X,Y)|G) = \psi(Y) \quad a.s., \quad \text{where } \psi(y) = E(\varphi(X,y))$$

This proposition has the following consequence: when computing the expectation of a function $\varphi$ of two independent random variables $X$ and $Y$, one can always divide the computation in two steps by writing

$$E(\varphi(X,Y)) = E(E(\varphi(X,Y)|G)) = E(\psi(Y))$$

where $\psi(y) = E(\varphi(X,y))$ (this is actually nothing but Fubini’s theorem).

Finally, the proposition below (given again without proof) shows that Jensen’s inequality also holds for conditional expectation.

**Proposition 1.5.** Let $X$ be a random variable, $G$ be a sub-$\sigma$-field of $F$ and $\psi : \mathbb{R} \to \mathbb{R}$ be Borel-measurable, convex and such that $E(|\psi(X)|) < \infty$. Then

$$\psi(E(X|G)) \leq E(\psi(X)|G) \quad a.s.$$ 

In particular, $|E(X|G)| \leq E(|X||G)$ a.s.

### 1.5 Conditioning with respect to a random variable $Y$

Once the definition of conditional expectation with respect to a $\sigma$-field is set, it is natural to define it for a generic random variable $Y$:

$$E(X|Y) = E(X|\sigma(Y)) \quad \text{and} \quad P(A|Y) = P(A|\sigma(Y))$$

**Remark.** Since any $\sigma(Y)$-measurable random variable may be written as $g(Y)$, where $g$ is a Borel-measurable function, the definition of $E(X|Y)$ may be rephrased as follows.

**Definition 1.6.** $E(X|Y) = \psi(Y)$, where $\psi : \mathbb{R} \to \mathbb{R}$ is the unique Borel-measurable function such that $E(\psi(Y)g(Y)) = E(Xg(Y))$ for any function $g : \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded.
In two particular cases, the function $\psi$ can be made explicit, which allows for concrete computations.

- If $X$, $Y$ are two discrete random variables with values in a countable set $C$, then
  
  \[ E(X|Y) = \psi(Y), \quad \text{where} \quad \psi(y) = \sum_{x \in C} x \ P(\{X = x\}|\{Y = y\}), \quad y \in C \]

  which matches the formula given in Section 1.2. The proof that it also matches the theoretical definition of conditional expectation is left as an exercise.

- If $X,Y$ are two jointly continuous random variables with joint pdf $p_{X,Y}$, then
  
  \[ E(X|Y) = \psi(Y), \quad \text{where} \quad \psi(y) = \int_{\mathbb{R}} x \frac{p_{X,Y}(x,y)}{p_Y(y)} \, dx, \quad y \in \mathbb{R} \]

  and $p_Y$ is the marginal pdf of $Y$ given by $p_Y(y) = \int_{\mathbb{R}} p_{X,Y}(x,y) \, dx$, assumed here to be strictly positive. Let us check that the random variable $\psi(Y)$ is indeed the conditional expectation of $X$ given $Y$ according to Definition 1.6: for any function $g : \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded, one has

  \[
  \mathbb{E}(\psi(Y) \, g(Y)) = \int_{\mathbb{R}} \psi(y) \, g(y) \, p_Y(y) \, dy \\
  = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} x \frac{p_{X,Y}(x,y)}{p_Y(y)} \, dx \right) g(y) \, p_Y(y) \, dy \\
  = \int_{\mathbb{R}^2} x \, g(y) \frac{p_{X,Y}(x,y)}{p_Y(y)} \, dx \, dy = \mathbb{E}(X \, g(Y)).
  \]

  $\square$

Finally, the conditional expectation satisfies the following proposition when $X$ is a square-integrable random variable.

**Proposition 1.7.** Let $X$ be a square-integrable random variable, $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$ and $G$ be the linear subspace of square-integrable and $\mathcal{G}$-measurable random variables. Then the conditional expectation of $X$ with respect to $\mathcal{G}$ is equal a.s. to the random variable $Z$ satisfying

\[ Z = \operatorname{argmin}_{Y \in G} \mathbb{E}((X - Y)^2) \]

In other words, this is saying that $Z$ is the orthogonal projection of $X$ onto the linear subspace $G$ of square-integrable and $\mathcal{G}$-measurable random variables (the scalar product considered here being $\langle X, Y \rangle = \mathbb{E}(XY)$). In particular,

\[ \mathbb{E}((X - Z) \, U) = 0, \quad \text{for any} \ U \in G \]

which is nothing but a variant of condition (ii) in the definition of conditional expectation.