Advanced Probability

Perla Sousi*

October 10, 2011

Contents

1 Conditional expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, i.e. Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$.

Definition 1.1. $\mathcal{F}$ is a σ-algebra on Ω if it satisfies:

1. $\Omega \in \mathcal{F}$
2. If $A \in \mathcal{F}$, then also the complement is in $\mathcal{F}$, i.e.,$A^c \in \mathcal{F}$.
3. If $(A_n)_{n\geq 1}$ is a collection of sets in $\mathcal{F}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definition 1.2. $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$ if it satisfies:

1. $\mathbb{P} : \mathcal{F} \to [0, 1]$, i.e. it is a set function
2. $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\varnothing) = 0$
3. If $(A_n)_{n\geq 1}$ is a collection of pairwise disjoint sets in $\mathcal{F}$, then $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

Let $A, B \in \mathcal{F}$ be two events with $\mathbb{P}(B) > 0$. Then the conditional probability of A given the event B is defined by

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

*University of Cambridge, Cambridge, UK; p.sousi@statslab.cam.ac.uk

Definition 1.3. The Borel σ-algebra, $\mathcal{B}(\mathbb{R})$, is the σ-algebra generated by the open sets in $\mathbb{R}$, i.e., it is the intersection of all σ-algebras containing the open sets of $\mathbb{R}$. More formally, let $\mathcal{O}$ be the open sets of $\mathbb{R}$, then

$$\mathcal{B}(\mathbb{R}) = \cap\{\mathcal{E} : \mathcal{E} \text{ is a } \sigma\text{-algebra containing } \mathcal{O}\}.$$

Informally speaking, consider the open sets of $\mathbb{R}$, do all possible operations, i.e., unions, intersections, complements. and take the smallest σ-algebra that you get.

Definition 1.4. X is a random variable, i.e., a measurable function with respect to $\mathcal{F}$, if $X : \Omega \to \mathbb{R}$ is a function with the property that for all open sets V the inverse image $X^{-1}(V) \in \mathcal{F}$.

Remark 1.5. If X is a random variable, then the collection of sets

$$\{B \subseteq \mathbb{R} : X^{-1}(B) \in \mathcal{F}\}$$

is a σ-algebra (check!) and hence it must contain $\mathcal{B}(\mathbb{R})$.

Let $A \in \mathcal{F}$. The indicator function $\mathbf{1}(A)$ is defined via

$$\mathbf{1}(A)(x) = \mathbf{1}(x \in A) = \begin{cases} 1, & \text{if } x \in A; \\ 0, & \text{otherwise.} \end{cases}$$

Recall the definition of expectation. First for positive simple random variables, i.e., linear combinations of indicator random variables, we define

$$\mathbb{E}\left[\sum_{i=1}^{n} c_i \mathbf{1}(A_i)\right] := \sum_{i=1}^{n} c_i \mathbb{P}(A_i),$$

where c_i are positive constants and A_i are measurable events. Next, let X be a positive random variable. Then X is the increasing limit of positive simple variables. For example

$$X_n(\omega) = 2^{-n}\lfloor 2^n X(\omega)\rfloor \wedge n \uparrow X(\omega) \text{ as } n \to \infty.$$

So we define

$$\mathbb{E}[X] :=\uparrow \lim \mathbb{E}[X_n].$$

Finally, for a general random variable X, we can write $X = X^+ - X^-$, where $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$ and we define

$$\mathbb{E}[X] := \mathbb{E}[X^+] - \mathbb{E}[X^-],$$

if at least one of $\mathbb{E}[X^+], \mathbb{E}[X^-]$ is finite. A random variable X is called *integrable*, if $\mathbb{E}[|X|] < \infty$.

Let X be a random variable with $\mathbb{E}[|X|] < \infty$. Let A be an event in $\mathcal{F}$ with $\mathbb{P}(A) > 0$. Then the conditional expectation of X given A is defined by

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X\mathbf{1}(A)]}{\mathbb{P}(A)},$$

Our goal is to extend the definition of conditional expectation to σ-algebras. So far we only have defined it for events and it was a number. Now, the conditional expectation is going to be a random variable, measurable with respect to the σ-algebra wrt which we are conditioning.

1.1 Discrete case

Let $X \in \mathcal{L}^1$. Let's start with a σ-algebra which is generated by a countable family of disjoint events $(B_i)_{i\in I}$ with $\cup_i B_i = \Omega$, i.e., $\mathcal{G} = \sigma(B_i, i \in I)$. It is easy to check that $\mathcal{G} = \{\cup_{i\in J} B_i : J \subseteq \mathbb{N}\}$.
The natural thing to do is to define a new random variable $X' = \mathbb{E}[X|\mathcal{G}]$ as follows

$$X' = \sum_{i\in I} \mathbb{E}[X|B_i]\mathbf{1}(B_i).$$

What does this mean? Let $\omega \in \Omega$. Then $X'(\omega) = \sum_{i\in I} \mathbb{E}[X|B_i]\mathbf{1}(\omega \in B_i)$. Note that we use the convention that $\mathbb{E}[X|B_i] = 0$, if $\mathbb{P}(B_i) = 0$
It is very easy to check that

$$X' \text{ is } \mathcal{G} - \text{measurable} \tag{1.1}$$

and integrable, since

$$\mathbb{E}[|X'|] \leq \sum_{i\in I} E|X\mathbf{1}(B_i)| = \mathbb{E}[|X|] < \infty,$$

since $X \in \mathcal{L}^1$.
Let $G \in \mathcal{G}$. Then it is straightforward to check that

$$\mathbb{E}[X\mathbf{1}(G)] = \mathbb{E}[X'\mathbf{1}(G)]. \tag{1.2}$$

1.2 Existence and uniqueness

Before stating the existence and uniqueness theorem on conditional expectation, let us quickly recall the notion of an event happening almost surely (a.s.), the Monotone convergence theorem and $\mathcal{L}^p$ spaces.
Let $A \in \mathcal{F}$. We will say that A happens a.s., if $\mathbb{P}(A) = 1$.

Theorem 1.6. [**Monotone convergence theorem**] Let $(X_n)_n$ be random variables such that $X_n \geq 0$ for all n and $X_n \uparrow X$ as $n \to \infty$ a.s. Then

$$\mathbb{E}[X_n] \uparrow \mathbb{E}[X] \text{ as } n \to \infty.$$

Let $p \in [1, \infty)$ and f a measurable function in $(\Omega, \mathcal{F}, \mathbb{P})$. We define the norm

$$\|f\|_p = (\mathbb{E}[|f|^p])^{1/p}$$

and we denote by $\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ the set of measurable functions f with $\|f\|_p < \infty$. For $p = \infty$, we let

$$\|f\|_\infty = \inf\{\lambda : |f| \leq \lambda \text{ a.e.}\}$$

and $\mathcal{L}^\infty$ the set of measurable functions with $\|f\|_\infty < \infty$.
Formally, $\mathcal{L}^p$ is the collection of equivalence classes, where two functions are equivalent if they are equal almost everywhere (a.e.). In practice, we will represent an element of $\mathcal{L}^p$ by a function, but remember that equality in $\mathcal{L}^p$ means equality a.e..

Theorem 1.7. *The space $(\mathcal{L}^2, \|\cdot\|_2)$ is a Hilbert space with $\langle f, g \rangle = \mathbb{E}[fg]$. If $\mathcal{H}$ is a closed subspace, then for all $f \in \mathcal{L}^2$, there exists a unique (in the sense of a.e.) $g \in \mathcal{H}$ such that $\|f - g\|_2 = \inf_{h \in \mathcal{H}} \|f - h\|_2$ and $\langle g, f - g \rangle = 0$.*

Remark 1.8. We call g the orthogonal projection of f on $\mathcal{H}$.

Theorem 1.9. *Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Then there exists a random variable Y such that:*

(a) *Y is $\mathcal{G}$-measurable;*

(b) *Y is integrable and $\mathbb{E}[X\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)]$ for all $A \in \mathcal{G}$.*

Moreover, if Y' also satisfies (a) *and* (b)*, then $Y = Y'$ a.s..*

We call Y (*a version of*) *the conditional expectation of X given $\mathcal{G}$* and write $Y = \mathbb{E}[X|\mathcal{G}]$ a.s.. In the case $\mathcal{G} = \sigma(G)$ for some random variable G, we also write $Y = \mathbb{E}[X|G]$ a.s..

Remark 1.10. We could replace (b) in the statement of the theorem by requiring that for all bounded $\mathcal{G}$-measurable random variables Z we have

$$\mathbb{E}[XZ] = \mathbb{E}[YZ].$$

Remark 1.11. In a later section we will show how to construct explicit versions of the conditional expectation in certain simple cases. In general, we have to live with the indirect approach provided by the theorem.

Proof of Theorem 1.9. (*Uniqueness.*) Suppose that both Y and Y' satisfy (a) and (b). Then, clearly the event $A = \{Y > Y'\} \in \mathcal{G}$ and by (b) we have

$$\mathbb{E}[(Y - Y')\mathbf{1}(A)] = \mathbb{E}[X\mathbf{1}(A)] - \mathbb{E}[X\mathbf{1}(A)] = 0,$$

hence we get that $Y \leq Y'$ a.s. Similarly we can get $Y \geq Y'$ a.s.
(*Existence.*) We will prove existence in three steps.
1st step: Suppose that $X \in \mathcal{L}^2$. The space $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ with inner product defined by $\langle U, V\rangle = \mathbb{E}[UV]$ is a Hilbert space by Theorem 1.7 and $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace. (*Remember that $\mathcal{L}^2$ convergence implies convergence in probability and convergence in probability implies convergence a.s. along a subsequence (see for instance [1, A13.2])*).
Thus $\mathcal{L}^2(\mathcal{F}) = \mathcal{L}^2(\mathcal{G}) + \mathcal{L}^2(\mathcal{G})^{\perp}$, and hence, we can write X as $X = Y + Z$, where $Y \in \mathcal{L}^2(\mathcal{G})$ and $Z \in \mathcal{L}^2(\mathcal{G})^{\perp}$. If we now set $Y = \mathbb{E}[X|\mathcal{G}]$, then (a) is clearly satisfied. Let $A \in \mathcal{G}$. Then

$$\mathbb{E}[X\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)] + \mathbb{E}[Z\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)],$$

since $\mathbb{E}[Z\mathbf{1}(A)] = 0$.
Note that from the above definition of conditional expectation for random variables in $\mathcal{L}^2$, we get that

$$\text{if } X \geq 0, \text{ then } Y = \mathbb{E}[X|\mathcal{G}] \geq 0 \text{ a.s.}, \tag{1.3}$$

since note that $\{Y < 0\} \in \mathcal{G}$ and

$$\mathbb{E}[X\mathbf{1}(Y < 0)] = \mathbb{E}[Y\mathbf{1}(Y < 0)].$$

Notice that the left hand side is nonnegative, while the right hand side is non-positive, implying that $\mathbb{P}(Y < 0) = 0$.
2nd step: Suppose that $X \geq 0$. For each n we define the random variables $X_n = X \wedge n \leq n$, and hence $X_n \in \mathcal{L}^2$. Thus from the first part of the existence proof we have that for each n there exists a $\mathcal{G}$-measurable random variable Y_n satisfying for all $A \in \mathcal{G}$

$$\mathbb{E}[Y_n\mathbf{1}(A)] = \mathbb{E}[(X \wedge n)\mathbf{1}(A)]. \tag{1.4}$$

Since the sequence $(X_n)_n$ is increasing, from (1.3) we get that also $(Y_n)_n$ is increasing. If we now set $Y =\uparrow \lim_{n\to\infty} Y_n$, then clearly Y is $\mathcal{G}$-measurable and by the monotone convergence theorem in (1.4) we get for all $A \in \mathcal{G}$

$$\mathbb{E}[Y\mathbf{1}(A)] = \mathbb{E}[X\mathbf{1}(A)], \tag{1.5}$$

since $X_n \uparrow X$, as $n \to \infty$.
In particular, if $\mathbb{E}[X]$ is finite, then $\mathbb{E}[Y]$ is also finite.
3rd step: Finally, for a general random variable $X \in \mathcal{L}^1$ (not necessarily positive) we can apply the above construction to $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$ and then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}]$ satisfies (a) and (b). □

Remark 1.12. Note that the 2nd step of the above proof gives that if $X \geq 0$, then there exists a $\mathcal{G}$-measurable random variable Y such that

$$\text{for all } A \in \mathcal{G},\ \mathbb{E}[X\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)],$$

i.e., all the conditions of Theorem 1.9 are satisfied except for the integrability one.

Definition 1.13. Sub-σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots$ of $\mathcal{F}$ are called independent, if whenever $G_i \in \mathcal{G}_i$ $(i \in \mathbb{N})$ and $i_1, \ldots, i_n$ are distinct, then

$$\mathbb{P}(G_{i_1} \cap \ldots \cap G_{i_n}) = \prod_{k=1}^{n} \mathbb{P}(G_{i_k}).$$

When we say that a random variable X is independent of a σ-algebra $\mathcal{G}$, it means that $\sigma(X)$ is independent of $\mathcal{G}$.

The following properties are immediate consequences of Theorem 1.9 and its proof.

Proposition 1.14. *Let $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. Then*

1. $\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\right] = \mathbb{E}[X]$
2. *If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$ a.s..*
3. *If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ a.s..*
4. *If $X \geq 0$, then $\mathbb{E}[X|\mathcal{G}] \geq 0$ a.s..*
5. *For any $\alpha, \beta \in \mathbb{R}$ we have $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\mathbb{E}[X|\mathcal{G}] + \beta\mathbb{E}[Y|\mathcal{G}]$ a.s..*
6. $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$ *a.s..*

The basic convergence theorems for expectation have counterparts for conditional expectation. We first recall the theorems for expectation.

Theorem 1.15. [**Fatou's lemma**] If $X_n \geq 0$ for all n, then

$$\mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n].$$

Theorem 1.16. [**Dominated convergence theorem**] If $X_n \to X$ and $|X_n| \leq Y$ for all n a.s., for some integrable random variable Y, then

$$\mathbb{E}[X_n] \to \mathbb{E}[X].$$

Theorem 1.17. [**Jensen's inequality**] Let X be an integrable random variable and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function. Then

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]).$$

Proposition 1.18. *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra and let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.*

1. ***Conditional monotone convergence theorem:*** *If $(X_n)_{n \geq 0}$ is an increasing sequence of nonnegative random variables with a.s. limit X, then*
$$\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}] \text{ as } n \to \infty, \text{ a.s.}.$$
2. ***Conditional Fatou's lemma:*** *If $X_n \geq 0$ for all n, then*
$$\mathbb{E}\left[\liminf_{n\to\infty} X_n|\mathcal{G}\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \text{ a.s.}.$$
3. ***Conditional dominated convergence theorem:*** *If $X_n \to X$ and $|X_n| \leq Y$ for all n a.s., for some integrable random variable Y, then*
$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}] \text{ a.s.}.$$
4. ***Conditional Jensen's inequality:*** *If X is integrable random and $\varphi : \mathbb{R} \to (-\infty, \infty]$ is a convex function such that either $\varphi(X)$ is integrable or φ is non-negative, then*
$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbb{E}[X|\mathcal{G}]) \text{ a.s.}.$$
In particular, for all $1 \leq p < \infty$
$$\|\mathbb{E}[X|\mathcal{G}]\|_p \leq \|X\|_p.$$

Proof. 1. Since $X_n \nearrow X$ as $n \to \infty$, we have that $\mathbb{E}[X_n|\mathcal{G}]$ is an increasing sequence. Let $Y = \lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]$. We want to show that $Y = \mathbb{E}[X|\mathcal{G}]$ a.s.. Clearly Y is $\mathcal{G}$-measurable, as an a.s. limit of $\mathcal{G}$-measurable random variables. Also, by the monotone convergence theorem we have
$$\mathbb{E}[X\mathbf{1}(A)] = \lim_{n\to\infty} \mathbb{E}[X_n\mathbf{1}(A)] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[X_n|\mathcal{G}]\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)].$$

2. The sequence $\inf_{k\geq n} X_k$ is increasing in n and $\lim_{n\to\infty} \inf_{k\geq n} X_k = \liminf_{n\to\infty} X_n$. Thus, by the conditional monotone convergence theorem we get
$$\lim_{n\to\infty} \mathbb{E}[\inf_{k\geq n} X_k|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}].$$
Clearly, $\mathbb{E}[\inf_{k\geq n} X_k|\mathcal{G}] \leq \inf_{k\geq n} \mathbb{E}[X_k|\mathcal{G}]$. Passing to the limit gives the desired inequality.

3. Since X_n+Y and $Y-X_n$ are positive random variables for all n, applying conditional Fatou's lemma we get
$$\mathbb{E}[X+Y|\mathcal{G}] = \mathbb{E}[\liminf(X_n+Y)|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n+Y|\mathcal{G}] \text{ and}$$
$$\mathbb{E}[Y-X|\mathcal{G}] = \mathbb{E}[\liminf(Y-X_n)|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[Y-X_n|\mathcal{G}].$$
Hence, we obtain that
$$\liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \geq \mathbb{E}[X|\mathcal{G}] \text{ and } \limsup_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \leq \mathbb{E}[X|\mathcal{G}].$$

4. A convex function is the supremum of countably many affine functions: (see for instance [1, §6.6])

$$\varphi(x) = \sup_i (a_i x + b_i), x \in \mathbb{R}.$$

So for all i we have $\mathbb{E}[\varphi(X)|\mathcal{G}] \geq a_i \mathbb{E}[X|\mathcal{G}] + b_i$ a.s. Now using the fact that the supremum is over a countable set we get that

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \sup_i (a_i \mathbb{E}[X|\mathcal{G}] + b_i) = \varphi(\mathbb{E}[X|\mathcal{G}]) \text{ a.s.}$$

In particular, for $1 \leq p < \infty$,

$$\|\mathbb{E}[X|\mathcal{G}]\|_p^p = \mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|^p] \leq \mathbb{E}[\mathbb{E}[|X|^p|\mathcal{G}]] = \mathbb{E}[|X|^p] = \|X\|_p^p.$$

□

References

[1] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.