σ-Fields, Conditional Expectation

Definition: Let Ω be a set, and let $\mathcal{F}$ be a *σ-field* of subsets of Ω, i.e., a collection of subsets of Ω whose elements satisfy the following properties:

(σ1) $\emptyset \in \mathcal{F}$ (where $\emptyset$ is the empty set, $\emptyset := \{\,\}$);

(σ2) if A_1, A_2, ... are elements of $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;

(σ3) if $A \in \mathcal{F}$, then its complement, $A^c := \Omega \setminus A$, belongs to $\mathcal{F}$.

Remark: In probability language, Ω is called the *sample space*, the elements of Ω are called *outcomes*, the elements of $\mathcal{F}$ are called *events*.

Remark: Note that $A \in \mathcal{F}$, but $A \subseteq \Omega$.

Definition: We say that the collection $\mathcal{G}$ of subsets of Ω is a *sub-σ-field* of $\mathcal{F}$ (and we write $\mathcal{G} \subseteq \mathcal{F}$) if all elements of $\mathcal{G}$ are also elements of $\mathcal{F}$, and $\mathcal{G}$ itself is a σ-field (i.e., the elements of $\mathcal{G}$ satisfy the properties (σ1)-(σ3)).

Definition: Let A_1, A_2, ... be an arbitrary collection of subsets of Ω. The smallest σ-field that contains all A_1, A_2, ... can be constructed by taking all possible unions of intersections of A_i's or their complements. The notation for this σ-field is $\sigma(A_1, A_2, \ldots)$; we say that $\sigma(A_1, A_2, \ldots)$ is *generated by* A_1, A_2, ...

Remark: The smallest σ-field of subsets of Ω (notation: $\sigma(\emptyset)$ or, equivalently, $\sigma(\Omega)$) consists of $\emptyset$ and Ω only; it is a sub-σ-field of any other σ-field of subsets of Ω.

Example A: Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, and let $E := \{2, 4, 6\}$ be the set of all even numbers, while $F := \{1, 2, 3, 4\}$. Then

$$\sigma(E) = \{\emptyset, E = \{2, 4, 6\}, E^c = \{1, 3, 5\}, \Omega\} ,$$

$$\sigma(F) = \{\emptyset, F = \{1, 2, 3, 4\}, F^c = \{5, 6\}, \Omega\} ,$$

and $\sigma(E, F)$ consists of the following 8 sets:

$$\emptyset,\ E = \{2, 4, 6\},\ E^c = \{1, 3, 5\},\ EF = \{2, 4\},\ EF^c = \{6\},\ E^cF = \{1, 3\},\ E^cF^c = \{5\},\ \Omega\ .$$

Note that $\sigma(E) \subseteq \sigma(E, F)$ and $\sigma(F) \subseteq \sigma(E, F)$. Also note that in this example the subset $\{2\}$ does not belong to $\sigma(E, F)$. The largest σ-field of subsets of Ω in this example – let us call it $\mathcal{F}$ – has 2^6 elements; can you write them down?
Hint: There are $\binom{6}{0} = 1$ elements of $\mathcal{F}$ consisting of 0 elements of Ω, $\binom{6}{1} = 6$ elements

of $\mathcal{F}$ consisting of 1 element of Ω, $\binom{6}{2} = 15$ elements of $\mathcal{F}$ consisting of 2 elements of Ω, $\binom{6}{3} = 20$ elements of $\mathcal{F}$ consisting of 3 elements of Ω, $\binom{6}{4} = 15$ elements of $\mathcal{F}$ consisting of 4 elements of Ω, $\binom{6}{5} = 6$ elements of $\mathcal{F}$ consisting of 5 elements of Ω, and $\binom{6}{6} = 1$ elements of $\mathcal{F}$ consisting of 6 elements of Ω.

Definition: A function $\mathbb{P} : \mathcal{F} \to [0,1]$ is a *probability measure* on $(\Omega, \mathcal{F})$ if it satisfies the following properties:

($\mathbb{P}$1) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$;

($\mathbb{P}$2) if A_1, A_2, ... are elements of $\mathcal{F}$ that are *disjoint*, i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Example A (cont.): Since we have different σ-fields, there are different probabilities. If we know the probability $\mathbb{P} : \mathcal{F} \to [0,1]$ (where $\mathcal{F}$ was the largest σ-field of subsets of Ω), then we can construct the probabilities

$$\bar{\mathbb{P}} : \sigma(E) \to [0,1]\ , \quad \tilde{\mathbb{P}} : \sigma(F) \to [0,1]\ , \quad \mathbb{P}^* : \sigma(E,F) \to [0,1]\ , \quad \mathbb{P}^\dagger : \sigma(\emptyset) \to [0,1]$$

in such a way that all these probabilities are consistent with $\mathbb{P}$. For example,

$$\tilde{\mathbb{P}}(\emptyset) = 0\ , \quad \tilde{\mathbb{P}}(F) = \frac{2}{3}\ , \quad \tilde{\mathbb{P}}(F^c) = \frac{1}{3}\ , \quad \tilde{\mathbb{P}}(\Omega) = 1\ ,$$

while $\mathbb{P}^\dagger(\emptyset) = 0$, $\mathbb{P}^\dagger(\Omega) = 1$. (Note that $\mathbb{P}^\dagger(E)$, among many other things, is not defined!)

Definition: Let $\mathbb{R}$ be the real line, and let $\mathcal{B}(\mathbb{R})$ be the σ-field generated by all open sets of $\mathbb{R}$; the σ-field $\mathcal{B}(\mathbb{R})$ is called the *Borel σ-field* on $\mathbb{R}$.

Definition: A function $X : \Omega \to \mathbb{R}$ is called a *random variable* if the preimage of each Borel set in $\mathbb{R}$ is an element of $\mathcal{F}$, i.e., each subset of Ω of the form $X^{-1}(B) := \{\omega \in \Omega\ :\ X(\omega) \in B\}$ is an element of $\mathcal{F}$, for any Borel set $B \in \mathcal{B}(\mathbb{R})$.

Remark: The above definition is very simple although it may sound complicated. Think about the discrete case: if X is a function $X : \Omega \to \{x_1, x_2, \ldots\} \subset \mathbb{R}$ that takes a finite or a countable set of values, then it is a (discrete) random variable if and only if $X^{-1}(x_i) \in \mathcal{F}$ for all values x_i that X can possibly take.

Definition: The *σ-field $\sigma(X)$ generated by the discrete random variable X* is defined as the σ-field generated by all subsets of the form $X^{-1}(x_i)$ for all possible values x_i of X.

In general (not only for a discrete random variable), $\sigma(X)$ is defined as the σ-field generated by the set of $X^{-1}(B)$ for all $B \in \mathcal{B}(\mathbb{R})$.

Example A (cont.): Let X, Y, and Z be random variables defined as follows:

$$X(\omega) = \omega \qquad \text{for all } \omega \in \Omega \ ,$$

$$Y(\omega) = \begin{cases} 8 & \text{for } \omega \in E = \{2,4,6\} \ , \\ 9 & \text{for } \omega \in E^c = \{1,3,5\} \ , \end{cases}$$

$$Z(\omega) = \begin{cases} 10 & \text{for } \omega \in F = \{1,2,3,4\} \ , \\ 13 & \text{for } \omega \in F^c = \{5,6\} \ . \end{cases}$$

Then $\sigma(X) = \mathcal{F}$ (recall that $\mathcal{F}$ is the largest σ-field of subsets of Ω), $\sigma(Y) = \sigma(E)$, and $\sigma(Z) = \sigma(F)$.

If you know that $Y(\omega) = 9$, then you know that ω is an odd number (i.e., belongs to the set $Y^{-1}(9) = \{1,3,5\} \subset \sigma(Y)$), but you do not know the exact value of ω.

If, in addition to $Y(\omega) = 9$, I tell you that $Z(\omega) = 10$, then you can conclude that ω must belong to $Y^{-1}(9) \cap Z^{-1}(10) = \{1,3\} \subset \sigma(Y,Z)$, so the new information allowed us to narrow the possible set of values of ω (but we are still not sure whether ω is 1 or 3). This reasoning leads us to the following

Very Important Remark (as Winnie the Pooh would say):
If X_1, X_2, X_3, ... is a sequence of random variables on the same probability space, we can construct the σ-fields

$$\mathcal{F}_1 := \sigma(X_1) \ , \quad \mathcal{F}_2 := \sigma(X_1, X_2) \ , \quad \mathcal{F}_3 := \sigma(X_1, X_2, X_3) \ , \quad \ldots$$

each of which is clearly a sub-σ-field of the next one:

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \cdots \ .$$

Such a nested sequence of σ-fields is sometimes called a *filtration* of σ-fields. For an outcome $\omega \in \Omega$, to know $X_1(\omega)$ is the same as to know to which element of $\mathcal{F}_1 = \sigma(X_1)$ the outcome ω belongs. To know $X_1(\omega)$ and $X_2(\omega)$ is the same as knowing to which element of $\mathcal{F}_2 = \sigma(X_1, X_2)$ the outcome ω belongs, etc.

Now think of X_n as a piece of information about ω that you receive at time $t = n \in \mathbb{N}$, i.e., at time $t = 1$ somebody tells us the value of $X_1(\omega)$, at time $t = 2$ somebody tells us the value of $X_2(\omega)$, at time $t = 3$ somebody tells us the value of $X_3(\omega)$, etc. Then, obviously, the longer we wait, the more information about ω we will acquire, and the more "precisely" we will know it.

As an example of such a situation, think about the following: I will be telling you the consecutive digits of a number ω between 0 and 1, one digit per second. If at time $t = 1$, I tell you "5", which means that that $\omega \in [0.5, 0.6)$. At time $t = 2$, I tell you "1", and you learn that $\omega \in [0.51, 0.52)$. At time $t = 3$, I tell you "7", and you now know that $\omega \in [0.517, 0.518)$, etc. Clearly, in the notation introduced above, $\mathcal{F}_1$ is generated of all sets of the form $\left[\frac{i}{10^1}, \frac{i+1}{10^1}\right)$ for $i \in \{0,1,2,\ldots,10^1 - 1\}$, $\mathcal{F}_2$ is generated of all sets of the form $\left[\frac{i}{10^2}, \frac{i+1}{10^2}\right)$ for $i \in \{0,1,2,\ldots,10^2 - 1\}$, etc.

A subset $A \subseteq \Omega$ belongs to the σ-field $\mathcal{F}_n$ if and only if the knowledge of the values $X_1(\omega)$, $X_2(\omega)$, ..., $X_n(\omega)$ will allow you to determine whether $\omega \in A$ or $\omega \notin A$ for an arbitrary $\omega \in \Omega$.

Example: For the situation just considered, let A be the event

$$A := \{\text{there are at least two digits 4 among the first five digits of the number } \omega\} .$$

("Digits" here means "digits to the right of the decimal point".)
To which $\mathcal{F}_n$ does the event A belong? In general, to be able to answer this question, we have to know the first five digits of ω, which tells us that $A \in \mathcal{F}_5$, and, since $\mathcal{F}_5 \subseteq \mathcal{F}_6 \subseteq \mathcal{F}_7 \subseteq \cdots$, we have that $A \in \mathcal{F}_n$ for all $n \geq 5$. It may happen that for some particular ω we can determine whether ω belongs to A even *before* we know all the 5 first digits of ω – for example, if it happens that $X_1(\omega) = 4$ and $X_2(\omega) = 4$, then we would know that $\omega \in A$. This, however, does *not* imply that $A \in \mathcal{F}_2$ – think about the case $X_1(\omega) = 1$, $X_2(\omega) = 4$, $X_1(\omega) = 3$, $X_4(\omega) = 7$ – can you say whether there are two digits 4 in the first five digits of ω before you learn $X_5(\omega)$? Clearly not.
Once again, note that since the event A belongs to $\mathcal{F}_5$, it belongs to all events $\mathcal{F}_n$ with $n \geq 5$.
As an exercise, for each of the following events find the smallest n such that the event belongs to $\mathcal{F}_n$:

- $B :=$ {the first occurrence of the digit 7 in ω is preceded by no more than ten digits 2};
- $C :=$ {there is at least one digit 0 among the digits of ω};
- $D :=$ {the first one hundred digits of ω are the same};
- $E :=$ {there are no more than two digits 4 and two digits $\neq 4$ among the first five digits of ω}.

Answers: $B \in \mathcal{F}_{11}$ but does not belong to $\mathcal{F}_{10}$; C does not belong to $\mathcal{F}_n$ for any $n \in \mathbb{N}$; $D \in \mathcal{F}_{100}$ but not to $\mathcal{F}_{99}$; $E \in \mathcal{F}_n$ for all $n \in \mathbb{N}$ because $E = \emptyset$, and the empty set belongs to all $\mathcal{F}_n$.

Definition: Let X_n be a sequence of a random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \cdots$$

be a filtration of sub-σ-fields of $\mathcal{F}$, where $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$. Let S_n be a sequence of random variables on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that the sequence S_n is *adapted to the filtration* $\mathcal{F}_n$ if we can find the value of $S_n(\omega)$ from the values of $X_1(\omega)$, $X_2(\omega)$, ..., $X_n(\omega)$ (in a more mathematical language, the sequence S_n is adapted to the filtration $\mathcal{F}_n$ if S_n is $\mathcal{F}_n$-measurable).

Example: Let $S_n := X_1 + X_2 + \cdots + X_n$ be the random variable equal to the sum of the first n digits of the random number $\omega \in [0, 1)$. Then $S_n(\omega) = \sum_{i=1}^{n} X_i(\omega)$, i.e., if we know

$X_1(\omega)$, $X_2(\omega)$, …, $X_n(\omega)$, we will be able to find $S_n(\omega)$. This means that the sequence of random variables S_n is adapted to the filtration $\mathcal{F}_n$, where $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.

Definition: Let $\mathbb{E}(Y|X=x)$ be the average value of X over all elements of $X^{-1}(x)$:

$$\mathbb{E}(Y|X=x) := \sum_y y\, f_{Y|X}(y|x) = \sum_y y\, \mathbb{P}(Y=y|X=x) = \sum_x x \frac{\mathbb{P}(Y=y, X=x)}{\mathbb{P}(X=x)} .$$

We define the *conditional expectation*, $\mathbb{E}(Y|X)$, of Y given X as a random variable $\mathbb{E}(Y|X) : \Omega \to \mathbb{R}$ such that

$$\mathbb{E}(Y|X)(\omega) = \mathbb{E}(Y|X = X(\omega)) = \sum_y y\, f_{Y|X}(y|X(\omega)) ,$$

i.e., $\mathbb{E}(Y|X)(\omega)$ is the average of Y over the event $X^{-1}(X(\omega))$.

Remark: Note that if $X(\omega) = X(\omega')$, then $\mathbb{E}(Y|X)(\omega) = \mathbb{E}(Y|X)(\omega')$. This means that $\mathbb{E}(Y|X)$ – which is a random variable on the sample space Ω endowed with the σ-field $\mathcal{F}$ – can also be considered as a random variable on the sample space Ω endowed with the σ-field $\sigma(X)$ (which obviously is a sub-σ-field of $\mathcal{F}$).

Remark: Convince yourself that the discrete random variables X and $\mathbb{E}(Y|X)$ can be written as

$$X = \sum_x x\, I_{X^{-1}(x)} , \qquad \mathbb{E}(Y|X) = \sum_x \mathbb{E}(Y|X=x)\, I_{X^{-1}(x)} .$$

This can be interpreted as the fact that $\mathbb{E}(Y|X)$ is a function of X, $\mathbb{E}(Y|X) = \psi(X)$ for some (Borel-measurable) function $\psi : \mathbb{R} \to \mathbb{R}$:

$$\mathbb{E}(Y|X)(\omega) = \psi(X(\omega)) .$$

Remark: Recall that the event $X^{-1}(X(\omega)) = \{\omega' \in \Omega : X(\omega') = X(\omega)\}$ is an element of the σ-field $\sigma(X)$ generated by the random variable X. Note that the value of $\mathbb{E}(Y|X)(\omega)$ does not depend on the numerical value of $X(\omega)$, but only on the element $X^{-1}(X(\omega))$ of the $\sigma(X)$ to which ω belongs. This means that instead of $\mathbb{E}(Y|X)$, we can write $\mathbb{E}(Y|\sigma(X))$. Now recall that in the Very Important Remark above we constructed a sequence of nested σ-fields $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \cdots$ generated by the sequence of random variables X_1, X_2, X_3, …, where $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Taking into account what we have just discussed, we introduce the notation (which is standard in the theory of martingales)

$$\mathbb{E}(Y|\mathcal{F}_n) := \mathbb{E}(Y|\sigma(X_1, \ldots, X_n)) .$$

If $X_1(\omega) = x_1$, …, $X_n(\omega) = x_n$, then the value of $\mathbb{E}(Y|\mathcal{F}_n)(\omega)$ is equal to the average of Y over the event $\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}$.

Example A (cont.): Show that $\mathbb{E}(X) = 3.5$, $\mathbb{E}(Y) = 8.5$, $\mathbb{E}(Z) = 11$, while for the conditional expectations we have

$$\begin{aligned}
\mathbb{E}(X|Y)(\omega) &= \mathbb{E}(X|Y=8) = 2\cdot\frac{1}{3} + 4\cdot\frac{1}{3} + 6\cdot\frac{1}{3} = 4 \qquad \text{for } \omega \in Y^{-1}(8) = \{2,4,6\}\ , \\
\mathbb{E}(X|Y)(\omega) &= \mathbb{E}(X|Y=9) = 1\cdot\frac{1}{3} + 3\cdot\frac{1}{3} + 5\cdot\frac{1}{3} = 3 \qquad \text{for } \omega \in Y^{-1}(9) = \{1,3,5\}\ ;
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(X|Z)(\omega) &= \mathbb{E}(X|Z=10) = 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{4} + 4\cdot\frac{1}{4} = 2.5 \qquad \text{for } \omega \in Z^{-1}(10) = \{1,2,3,4\}\ , \\
\mathbb{E}(X|Z)(\omega) &= \mathbb{E}(X|Z=13) = 5\cdot\frac{1}{2} + 6\cdot\frac{1}{2} = 5.5 \qquad \text{for } \omega \in Z^{-1}(13) = \{5,6\}\ .
\end{aligned}$$

Now we compute the expectation of the random variables $\mathbb{E}(X|Y)$ and $\mathbb{E}(X|Z)$:

$$\mathbb{E}\left[\mathbb{E}(X|Y)\right] = \sum_{y=8,9} \mathbb{E}(X|Y=y)\,\mathbb{P}(Y=y) = 4\cdot\frac{1}{2} + 3\cdot\frac{1}{2} = 3.5\ ,$$

$$\mathbb{E}\left[\mathbb{E}(X|Z)\right] = \sum_{z=10,13} \mathbb{E}(X|Z=z)\,\mathbb{P}(Z=z) = 2.5\cdot\frac{2}{3} + 5.5\cdot\frac{1}{3} = 3.5\ ,$$

and notice that $\mathbb{E}\left[\mathbb{E}(X|Y)\right] = \mathbb{E}(X)$ and $\mathbb{E}\left[\mathbb{E}(X|Z)\right] = \mathbb{E}(X)$, as it should, according to Theorem 3.7.6. As an exercise, show that:

- $\mathbb{E}(Y|X)(\omega) = 8$ for $\omega \in \{2,4,6\}$, $\mathbb{E}(Y|X)(\omega) = 9$ for $\omega \in \{1,3,5\}$, $\mathbb{E}\left[\mathbb{E}(Y|X)\right] = 8.5$;
- $\mathbb{E}(Y|Z)(\omega) = 8.5$ for $\omega \in \Omega$, $\mathbb{E}\left[\mathbb{E}(Y|Z)\right] = 8.5$;
- $\mathbb{E}(Z|X)(\omega) = 10$ for $\omega \in \{1,2,3,4\}$, $\mathbb{E}(Z|X)(\omega) = 13$ for $\omega \in \{5,6\}$, $\mathbb{E}\left[\mathbb{E}(Z|X)\right] = 11$;
- $\mathbb{E}(Z|Y)(\omega) = 11$ for $\omega \in \Omega$, $\mathbb{E}\left[\mathbb{E}(Z|Y)\right] = 11$.