Conditional expectations, filtration and martingales

Content.

1. Conditional expectations
2. Martingales, sub-martingales and super-martingales

1 Conditional Expectations

1.1 Definition

Recall how we define conditional expectations. Given a random variable $X$ and an event $A$ we define $E[X | A] = \frac{E[X 1(A)]}{P(A)}$.

Also we can consider conditional expectations with respect to random variables. For simplicity say $Y$ is a simple random variable on $\Omega$ taking values $y_1, y_2, \ldots, y_n$ with some probabilities $P(\omega : Y(\omega) = y_i) = p_i$.

Now we define conditional expectation $E[X | Y]$ as a random variable which takes value $E[X | Y = y_i]$ with probability $p_i$, where $E[X | Y = y_i]$ should be understood as expectation of $X$ conditioned on the event $\{ \omega \in \Omega : Y(\omega) = y_i \}$.

It turns out that one can define conditional expectation with respect to a $\sigma$-field. This notion will include both conditioning on events and conditioning on random variables as special cases.

**Definition 1.** Given $\Omega$, two $\sigma$-fields $G \subset F$ on $\Omega$, and a probability measure $\mathbb{P}$ on $(\Omega, F)$. Suppose $X$ is a random variable with respect to $F$ but not necessarily with respect to $G$, and suppose $X$ has a finite $L_1$ norm (that is $E[|X|] < \infty$).

The conditional expectation $E[X | G]$ is defined to be a random variable $Y$ which satisfies the following properties:

(a) $Y$ is measurable with respect to $G$. 

1
(b) For every $A \in \mathcal{G}$, we have $E[X1\{A\}] = E[Y1\{A\}]$.

For simplicity, from now on we write $Z \in \mathcal{F}$ to indicate that $Z$ is measurable with respect to $\mathcal{F}$. Also let $\mathcal{F}(Z)$ denote the smallest $\sigma$-field such with respect to which $Z$ is measurable.

**Theorem 1.** The conditional expectation $E[X|\mathcal{G}]$ exists and is unique.

Uniqueness means that if $Y' \in \mathcal{G}$ is any other random variable satisfying conditions (a),(b), then $Y' = Y$ a.s. (with respect to measure $\mathbb{P}$). We will prove this theorem using the notion of Radon-Nikodym derivative, the existence of which we state without a proof below. But before we do this, let us develop some intuition behind this definition.

### 1.2 Simple properties

- Consider the trivial case when $\mathcal{G} = \{\emptyset, \Omega\}$. We claim that the constant value $c = E[X]$ is $E[X|\mathcal{G}]$. Indeed, any constant function is measurable with respect to any $\sigma$-field So (a) holds. For (b), we have $E[X1\{\Omega\}] = E[X] = c$ and $E[c1\{\Omega\}] = c$; and $E[X1\{\emptyset\}] = 0$ and $E[c1\{\emptyset\}] = 0$.

- As the other extreme, suppose $\mathcal{G} = \mathcal{F}$. Then we claim that $X = E[X|\mathcal{G}]$. The condition (b) trivially holds. The condition (a) also holds because of the equality between two $\sigma$-fields.

- Let us go back to our example of conditional expectation with respect to an event $A \subset \Omega$. Consider the associated $\sigma$-fields $\mathcal{G} = \{\emptyset, A, A^c, \Omega\}$ (we established in the first lecture that this is indeed a $\sigma$-field). Consider a random variable $Y : \Omega \to \mathbb{R}$ defined as

$$Y(\omega) = E[X|A] = \frac{E[X1\{A\}]}{P(A)} \triangleq c_1$$

for $\omega \in A$ and

$$Y(\omega) = E[X|A^c] = \frac{E[X1\{A^c\}]}{P(A^c)} \triangleq c_2$$

for $\omega \in A^c$. We claim that $Y = E[X|\mathcal{G}]$. First $Y \in \mathcal{G}$. Indeed, assume for simplicity $c_1 < c_2$. Then $\{\omega : Y(\omega) \leq x\} = \emptyset$ when $x < c_1$, $= A$...
for $c_1 \leq x < c_2 = \Omega$ when $x \geq c_2$. Thus $Y \in \mathcal{G}$. Then we need to check equality $\mathbb{E}[X1\{B\}] = \mathbb{E}[Y1\{B\}]$ for every $B = \emptyset, A, A^\complement, \Omega$, which is straightforward to do. For example say $B = A$. Then

$$\mathbb{E}[X1\{A\}] = \mathbb{E}[X|A]P(A) = c_1P(A).$$

On the other hand we defined $Y(\omega) = c_1$ for all $\omega \in A$. Thus

$$\mathbb{E}[Y1\{A\}] = c_1\mathbb{E}[1\{A\}] = c_1P(A).$$

And the equality checks.

- Suppose now $\mathcal{G}$ corresponds to some partition $A_1, \ldots, A_m$ of the sample space $\Omega$. Given $X \in \mathcal{F}$, using a similar analysis, we can check that $Y = \mathbb{E}[X|\mathcal{G}]$ is a random variable which takes values $\mathbb{E}[X|A_j]$ for all $\omega \in A_j$, for $j = 1, 2, \ldots, m$. You will recognize that this is one of our earlier examples where we considered conditioning on a simple random variable $Y$ to get $\mathbb{E}[X|Y]$. In fact this generalizes as follows:

- Given two random variables $X, Y : \Omega \rightarrow \mathbb{R}$, suppose both $\in \mathcal{F}$. Let $\mathcal{G} = \mathcal{G}(Y) \subset \mathcal{F}$ be the field generated by $Y$. We define $\mathbb{E}[X|Y]$ to be $\mathbb{E}[X|\mathcal{G}]$.

### 1.3 Proof of existence

We now give a proof sketch of Theorem 1.

**Proof.** Given two probability measures $\mathbb{P}_1, \mathbb{P}_2$ defined on the same $(\Omega, \mathcal{F})$, $\mathbb{P}_2$ is defined to be absolutely continuous with respect to $\mathbb{P}_1$ if for every set $A \in \mathcal{F}$, $\mathbb{P}_1(A) = 0$ implies $\mathbb{P}_2(A) = 0$.

The following theorem is the main technical part for our proof. It involves using the familiar idea of change of measures.

**Theorem 2 (Radon-Nikodym Theorem).** Suppose $\mathbb{P}_2$ is absolutely continuous with respect to $\mathbb{P}_1$. Then there exists a non-negative random variable $Y : \Omega \rightarrow \mathbb{R}_+$ such that for every $A \in \mathcal{F}$

$$\mathbb{P}_2(A) = \mathbb{E}_{\mathbb{P}_1}[Y1\{A\}].$$

Function $Y$ is called Radon-Nikodym (RN) derivative and sometimes is denoted $d\mathbb{P}_2/d\mathbb{P}_1$. 


Problem 1. Prove that \( Y \) is unique up-to measure zero. That is if \( Y' \) is also RN derivative, then \( Y = Y' \) a.s. w.r.t. \( P_1 \) and hence \( P_2 \).

We now use this theorem to establish the existence of conditional expectations. Thus we have \( G \subset F \), \( P \) is a probability measure on \( F \) and \( X \) is measurable with respect to \( F \). We will only consider the case \( X \geq 0 \) such that \( E[X] < \infty \). We also assume that \( X \) is not constant, so that \( E[X] > 0 \). Consider a new probability measure \( P_2 \) on \( G \) defined as follows:

\[
P_2(A) = \frac{E_P[X \mathbb{1}_{\{A\}}]}{E_P[X]}, \quad A \in G,
\]

where we write \( E_P \) in place of \( E \) to emphasize that the expectation operator is with respect to the original measure \( P \). Check that this is indeed a probability measure on \((\Omega, G)\). Now \( P \) also induces a probability measure on \((\Omega, G)\). We claim that \( P_2 \) is absolutely continuous with respect to \( P \). Indeed if \( P(A) = 0 \) then the numerator is zero. By the Radon-Nikodym Theorem then there exists \( Z \) which is measurable with respect to \( G \) such that for any \( A \in G \)

\[
P_2(A) = E_P[Z \mathbb{1}_{\{A\}}].
\]

We now take \( Y = Z E_P[X] \). Then \( Y \) satisfies the condition (b) of being a conditional expectation, since for every set \( B \)

\[
E_P[Y \mathbb{1}_{\{B\}}] = E_P[X] E_P[Z \mathbb{1}_{\{B\}}] = E_P[X \mathbb{1}_{\{B\}}].
\]

The second part, corresponding to the uniqueness property is proved similarly to the uniqueness of the RN derivative (Problem 1).

\[ \Box \]

2 Properties

Here are some additional properties of conditional expectations.

Linearity. \( E[aX + Y \mid G] = a E[X \mid G] + E[Y \mid G] \).

Monotonicity. If \( X_1 \leq X_2 \) a.s, then \( E[X_1 \mid G] \leq E[X_2 \mid G] \). Proof idea is similar to the one you need to use for Problem 1.

Independence.

Problem 2. Suppose \( X \) is independent from \( G \). Namely, for every measurable \( A \subset \mathbb{R}, B \in G \) \( P(\{X \in A\} \cap B) = P(X \in A)P(B) \). Prove that \( E[X \mid G] = E[X] \).
Conditional Jensen’s inequality. Let \( \phi \) be a convex function and \( \mathbb{E}[|X|], \mathbb{E}[|\phi(X)|] < \infty \). Then \( \phi(\mathbb{E}[X|G]) \leq \mathbb{E}[\phi(X)|G] \).

**Proof.** We use the following representation of a convex function, which we do not prove (see Durrett [1]). Let

\[
A = \{(a, b) \in \mathbb{Q} : ax + b \leq \phi(x), \forall x\}.
\]

Then \( \phi(x) = \sup\{ax + b : (a, b) \in A\} \).

Now we prove the Jensen’s inequality. For any pair of rationals \( a, b \in \mathbb{Q} \) satisfying the bound above, we have, by monotonicity that \( \mathbb{E}[\phi(X)|G] \geq a\mathbb{E}[X|G] + b \), a.s., implying \( \mathbb{E}[\phi(X)|G] \geq \sup\{a\mathbb{E}[X|G] + b : (a, b) \in A\} = \phi(\mathbb{E}[X|G]) \) a.s.

Tower property. Suppose \( G_1 \subset G_2 \subset F \). Then \( \mathbb{E}[\mathbb{E}[X|G_1]|G_2] = \mathbb{E}[X|G_1] \) and \( \mathbb{E}[\mathbb{E}[X|G_2]|G_1] = \mathbb{E}[X|G_1] \). That is the smaller field wins.

**Proof.** By definition \( \mathbb{E}[X|G_1] \) is \( G_2 \) measurable. Therefore it is \( G_1 \) measurable. Then the first equality follows from the fact \( \mathbb{E}[X|G] = X \), when \( X \in G \), which we established earlier. Now fix any \( A \in G_1 \). Denote \( \mathbb{E}[X|G_1] \) by \( Y_1 \) and \( \mathbb{E}[X|G_2] \) by \( Y_2 \). Then \( Y_1 \in G_1, Y_2 \in G_2 \). Then

\[
\mathbb{E}[Y_11\{A\}] = \mathbb{E}[X1\{A\}],
\]

simply by the definition of \( Y_1 = \mathbb{E}[X|G_1] \). On the other hand, we also have \( A \in G_2 \). Therefore

\[
\mathbb{E}[X1\{A\}] = \mathbb{E}[Y_21\{A\}].
\]

Combining the two equalities we see that \( \mathbb{E}[Y_21\{A\}] = \mathbb{E}[Y_11\{A\}] \) for every \( A \in G_1 \). Therefore, \( \mathbb{E}[Y_2|G_1] = Y_1 \), which is the desired result.

An important special case is when \( G_1 \) is a trivial \( \sigma \)-field \( \{\emptyset, \Omega\} \). We obtain that for every field \( G \)

\[
\mathbb{E}[\mathbb{E}[X|G]] = \mathbb{E}[X].
\]
3 Filtration and martingales

3.1 Definition

A family of \( \sigma \)-fields \( \{F_t\} \) is defined to be a filtration if \( F_{t_1} \subset F_{t_2} \) whenever \( t_1 \leq t_2 \). We will consider only two cases when \( t \in \mathbb{Z}_+ \) or \( t \in \mathbb{R}_+ \). A stochastic process \( \{X_t\} \) is said to be adapted to filtration \( \{F_t\} \) if \( X_t \in F_t \) for every \( t \).

**Definition 2.** A stochastic process \( \{X_t\} \) adapted to a filtration \( \{F_t\} \) is defined to be a martingale if

1. \( \mathbb{E}[|X_t|] < \infty \) for all \( t \).
2. \( \mathbb{E}[X_t|F_s] = X_s \), for all \( s < t \).

When equality is substituted with \( \leq \), the process is called a **supermartingale**. When it is substituted with \( \geq \), the process is called a **submartingale**.

Suppose we have a stochastic process \( \{X_t\} \) adapted to filtration \( \{F_t\} \) and suppose for some \( s' < s < t \) we have \( \mathbb{E}[X_t|F_{s'}] = X_s \) and \( \mathbb{E}[X_s|F_{s'}] = X_{s'} \).

Then using Tower property of conditional expectations

\[
\mathbb{E}[X_t|F_{s'}] = \mathbb{E}[\mathbb{E}[X_t|F_{s}]|F_{s'}] = \mathbb{E}[X_s|F_{s'}] = X_{s'}.
\]

This means that when the stochastic process \( \{X_n\} \) is discrete time it suffices to check \( \mathbb{E}[X_{n+1}|F_n] = X_n \) for all \( n \) in order to make sure that it is a martingale.

3.2 Simple examples

1. **Random walk.** Let \( X_n, n = 1, \ldots \) be an i.i.d. sequence with mean \( \mu \) and variance \( \sigma^2 < \infty \). Let \( F_n \) be the Borel \( \sigma \)-algebra on \( \mathbb{R}^n \). Then \( S_n - \mu n = \sum_{0 \leq k \leq n} X_k - \mu n \) is a martingale. Indeed \( S_n \) is adapted to \( F_n \), and

\[
\mathbb{E}[S_{n+1} - (n + 1)\mu|F_n] = \mathbb{E}[X_{n+1} - \mu + S_n - n\mu|F_n] = \mathbb{E}[X_{n+1} - \mu|F_n] + \mathbb{E}[S_n - n\mu|F_n] = \mathbb{E}[X_{n+1} - \mu] + S_n - n\mu = S_n - n\mu.
\]

Here in (a) we used the fact that \( X_{n+1} \) is independent from \( F_n \) and \( S_n \in F_n \).

2. **Random walk squared.** Under the same setting, suppose in addition \( \mu = 0 \). Then \( S^2_n - n\sigma^2 \) is a martingale. The proof of this fact is very similar.
4 Additional reading materials

• Durrett [1] Section 4.1, 4.2.

References
