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

1 Conditional expectation

Theorem 1.1 (Existence and uniqueness of conditional expectation)**.** *Let $X \in L^1$, and $\mathcal{G} \subseteq \mathcal{F}$. Then there exists a random variable Y such that*

- *Y is $\mathcal{G}$-measurable*
- *$Y \in L^1$, and $\mathbb{E}X\mathbf{1}_A = \mathbb{E}Y\mathbf{1}_A$ for all $A \in \mathcal{G}$.*

Moreover, if Y' is another random variable satisfying these conditions, then $Y' = Y$ almost surely.

We call Y a (version of) the conditional expectation given $\mathcal{G}$.

Proof. (Existence)

Case 1: $X \in L^2$.

Recall that L^2 is a Hilbert space, and that the set of $\mathcal{G}$-measurable random variables is a closed subspace of L^2 (it is closed because the space $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is complete). The projection theorem then gives us the existence and uniqueness of $Y \in L^2 \subseteq L^1$.

Case 2: $X \geq 0 \in L^1$.

Let $X_n = X \wedge n \in L^2$. Then by case 1, we can define $Y_n = \mathbb{E}(X_n \mid \mathcal{G}) \in L^2$. We make the following observation

Lemma 1.1.1. Suppose (X, Y) and (X', Y') are two pairs of random variables satisfying the conditions of the theorem, then $X \geq X'$ implies $Y \geq Y'$ almost surely.

Proof. Let $A = \{Y < Y'\}$. Then $\mathbb{E}Y\mathbf{1}_A = \mathbb{E}X\mathbf{1}_A \geq \mathbb{E}X'\mathbf{1}_A = \mathbb{E}Y'\mathbf{1}_A$, so $\mathbb{E}(Y - Y')\mathbf{1}_A \geq 0$ and $\mathbb{P}(A) = 0$. $\square$

It follows that there is some random variable Y such that $Y_n \uparrow Y$. Clearly Y is $\mathcal{G}$-measurable. For any $A \in \mathcal{G}$, we have

$$\begin{aligned}
\mathbb{E}Y\mathbf{1}_A &= \lim_{n\to\infty} \mathbb{E}Y_n\mathbf{1}_A && \text{(MCV)} \\
&= \lim_{n\to\infty} \mathbb{E}X_n\mathbf{1}_A && \text{(MCV)} \\
&= \mathbb{E}X\mathbf{1}_A
\end{aligned}$$

Case 3: $X \in L^1$.

Write $X = X^+ - X^-$, and apply case 2 to X^+ and X^-.

(Uniqueness) Suppose Y and Y' are two random variables satisfying the conditions of the theorem. The $\{Y > Y'\}$ is in $\mathcal{G}$ so $\mathbb{E}Y\mathbf{1}_{\{Y>Y'\}} = \mathbb{E}Y'\mathbf{1}_{\{Y>Y'\}} \implies \mathbb{E}(Y - Y')\mathbf{1}_{\{Y>Y'\}} = 0 \implies \mathbb{P}(Y > Y') = 0$. Similarly, $\mathbb{P}(Y' > Y) = 0$.

Remark. The above can also be proved using the Radon-Nikodym theorem.

(Proof via Radon-Nikodym) First recall the Radon-Nikodym theorem

Proposition. Let μ, ν be two σ-finite measures on $(\Omega, \mathcal{F})$ such that $\nu \ll \mu$. Then there exists a unique (up to a.e. equivalence) $f \in L^1(\Omega, \mathcal{F}, \mu)$ such that $\nu(A) = \int_A f \, d\mu$ for all $A \in \mathcal{F}$.

Consider the measure on $(\Omega, \mathcal{G})$ given by

$$\mu(A) = \mathbb{E}X\mathbf{1}_A, \quad A \in \mathcal{G}$$

so $\mu \ll \mathbb{P}$. By the Radon-Nikodym theorem, there exists a unique $Y \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ such that $\mu(A) = \int_A Y \, d\mathbb{P}$ for all $A \in \mathcal{G}$.

For general $X \in L^1$, we can write $X = X^+ - X^-$ and apply the above to X^+ and X^-. □

Proposition (Equivalent definition for conditional expectation)**.** Let $X, \mathcal{G}$ be as above. Then there exists a random variable Y such that

- Y is $\mathcal{G}$-measurable
- $Y \in L^1$ and $\mathbb{E}XZ = \mathbb{E}YZ$ for all $Z \in L^\infty(\mathcal{G})$

Moreover, $Y = \mathbb{E}(X \mid \mathcal{G})$ almost surely.

Proof. (Existence) Set $Y = \mathbb{E}(X \mid \mathcal{G})$. It is straightforward to see that Y satisfies the conditions of the proposition for simple functions Z. Note that simple functions that are in L^p are dense in L^p for $1 \leq p \leq \infty$. Let $Z_n \in L^\infty(\mathcal{G})$ be a sequence of simple functions such that $Z_n \to Z$ in L^∞ (in particular, we have almost sure pointwise convergence). Then

$$\begin{aligned} \mathbb{E}XZ &= \lim_{n\to\infty} \mathbb{E}XZ_n && \text{(DCT)} \\ &= \lim_{n\to\infty} \mathbb{E}YZ_n \\ &= \mathbb{E}YZ && \text{(DCT)} \end{aligned}$$

□

(Uniqueness) Note that any two random variables satisfying the conditions of the proposition are versions of the conditional expectation given $\mathcal{G}$, which was shown to be unique.

Lemma 1.1.2 (Conditional expectation as a function)**.** Let $X, Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then Y is measurable with respect to $\sigma(X)$ if and only if there exists a Borel-measurable function $f : \mathbb{R} \to \mathbb{R}$ such that $Y(\omega) = f(X(\omega))$ for all $\omega \in \Omega$.

Proposition (Properties of conditional expectation)**.** All (in)equality relations below hold almost surely.

1. If $X \geq 0$ a.s., then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$
2. If X and $\mathcal{G}$ are independent, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}[X]$

3. If $\alpha, \beta \in \mathbb{R}$ and $X_1, X_2 \in L^1$, then

$$\mathbb{E}(\alpha X_1 + \beta X_2 \mid \mathcal{G}) = \alpha\mathbb{E}(X_1 \mid \mathcal{G}) + \beta\mathbb{E}(X_2 \mid \mathcal{G}).$$

4. *Tower property*: If $\mathcal{H} \subseteq \mathcal{G}$, then

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}).$$

5. If Z is bounded and $\mathcal{G}$-measurable, then

$$\mathbb{E}(ZX \mid \mathcal{G}) = Z\mathbb{E}(X \mid \mathcal{G}).$$

6. Let $X \in L^1$ and $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$. Assume that $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$. Then

$$\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X \mid \sigma(\mathcal{G}, \mathcal{H})).$$

Proof. 1. Follows from the proof of existence and uniqueness of conditional expectation, or just use monotonicity.

2. Let $A \in \mathcal{G}$. Then $\mathbb{E}(\mathbb{E}(X)\mathbf{1}_A) = \mathbb{E}X\mathbb{E}1_A = \mathbb{E}(X1_A)$

3. Use linearity of conditional expectation.

4. Let $A \in \mathcal{H}$. Then $\mathbb{E}\left[\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H})\mathbf{1}_A\right] = \mathbb{E}[\mathbb{E}(X \mid \mathcal{G})\mathbf{1}_A] = \mathbb{E}(X\mathbf{1}_A)$

5. Easy if Z is an indicator function. Then use linearity and covergence theorems.

6. Note $\mathbb{E}(X \mid \mathcal{G})$ is $\sigma(\mathcal{G}, \mathcal{H})$-measurable and $\sigma(\mathcal{G}, \mathcal{H})$ is generated by the π-system $\{A \cap B : A \in \mathcal{G}, B \in \mathcal{H}\}$. We show that $\mathbb{E}(X \mid \mathcal{G})$ satisfies the defining property of $\mathbb{E}(X \mid \sigma(\mathcal{G}, \mathcal{H}))$. Let $A \in \mathcal{G}$ and $B \in \mathcal{H}$. Then for any element of the π-system, we have

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})\mathbf{1}_{A\cap B}) = \mathbb{E}[\mathbb{E}(X \mid \mathcal{G})\mathbf{1}_A\mathbf{1}_B] = \mathbb{E}[\mathbb{E}(X\mathbf{1}_A \mid \mathcal{G})\mathbf{1}_B] = \mathbb{E}(\underbrace{X\mathbf{1}_A}_{\in\sigma(\mathcal{G},X)})\mathbb{E}(\mathbf{1}_B) = \mathbb{E}(X\mathbf{1}_{A\cap B})$$

Since finite measures extend uniquely from π-systems, the above holds if $A \cap B$ is replaced by any element of $\sigma(\mathcal{G}, \mathcal{H})$

□

Proposition (Properties of conditional expectation)**.** All (in)equality relations below hold almost surely.

1. *Jensen's inequality*: If $c : \mathbb{R} \to \mathbb{R}$ is convex, then

$$\mathbb{E}(c(X) \mid \mathcal{G}) \geq c(\mathbb{E}(X) \mid \mathcal{G}).$$

2. For $p \geq 1$,

$$\|\mathbb{E}(X \mid \mathcal{G})\|_p \leq \|X\|_p.$$

3. *Monotone convergence theorem* Suppose $X_n \uparrow X$ is a sequence of non-negative random variables. Then

$$\mathbb{E}(X_n \mid \mathcal{G}) \uparrow \mathbb{E}(X \mid \mathcal{G}).$$

4. *Fatou's lemma*: If X_n are non-negative measurable, then

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n \mid \mathcal{G}\right) \le \liminf_{n\to\infty} \mathbb{E}(X_n \mid \mathcal{G}).$$

5. *Dominated convergence theorem*: If $X_n \to X$ and $Y \in L^1$ such that $Y \ge |X_n|$ for all n, then

$$\mathbb{E}(X_n \mid \mathcal{G}) \to \mathbb{E}(X \mid \mathcal{G}).$$

Proof. 1. Note that a convex function is the supremum of countably many affine functions $c(x) = \sup_{i\in I} a_i x + b_i$. Then

$$\begin{aligned}\mathbb{E}(c(X) \mid \mathcal{G}) &= \mathbb{E}\left(\sup_{i\in I}(a_i X + b_i) \mid \mathcal{G}\right)\\ &\ge \mathbb{E}(a_i X + b_i \mid \mathcal{G}) \quad \forall i \in I && \text{(monotonicity)}\end{aligned}$$

So $\mathbb{E}(c(X) \mid \mathcal{G}) \ge \sup_{i\in I} \mathbb{E}(a_i X + b_i \mid \mathcal{G}) = c(\mathbb{E}(X \mid \mathcal{G}))$.

2. Jensen

3. By monotonicity, $\mathbb{E}(X_n \mid \mathcal{G}) \uparrow Y$ for some Y. By the usual monotone convergence theorem, $\mathbb{E}\mathbb{E}(X_n \mid \mathcal{G}) = \mathbb{E}X_n \to \mathbb{E}Y \le \mathbb{E}X$ so $Y \in L^1$. Since each of the $\mathbb{E}(X_n \mid \mathcal{G})$ are $\mathcal{G}$-measurable, so is Y. Finally, for any $A \in \mathcal{G}$,

$$\begin{aligned}\mathbb{E}Y\mathbf{1}_A &= \lim_{n\to\infty} \mathbb{E}\mathbb{E}(X_n \mid \mathcal{G})\mathbf{1}_A && \text{(MCV)}\\ &= \lim_{n\to\infty} \mathbb{E}X_n\mathbf{1}_A\\ &= \mathbb{E}X\mathbf{1}_A && \text{(MCV)}\end{aligned}$$

4.

$$\begin{aligned}\mathbb{E}\left(\liminf_{n\to\infty} X_n \mid \mathcal{G}\right) &= \mathbb{E}\left(\lim_{n\to\infty} \underbrace{\inf_{m\ge n} X_m}_{increasing} \mid \mathcal{G}\right)\\ &= \lim_{n\to\infty} \mathbb{E}\left(\inf_{m\ge n} X_m \mid \mathcal{G}\right) && \text{(MCV)}\\ &= \liminf_{n\to\infty} \mathbb{E}\left(\underbrace{\inf_{m\ge n} X_m}_{\le X_n} \mid \mathcal{G}\right)\\ &\le \liminf_{n\to\infty} \mathbb{E}(X_n \mid \mathcal{G}) && \text{(monotonicity)}\end{aligned}$$

5. Use Fatou's lemma on $Y + X_n$ and $Y - X_n$.

□