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 (Radon-Nikodym theorem)**.** 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 $1 \le p \le \infty$. 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 $\sigma(\mathcal{G}, \mathcal{H})$

$\square$

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 $Y - X_n$.

□

2 Martingales

Definition ((Discrete) stochastic process)**.** A *stochastic process* (in discrete time) is a collection of random variables $(X_n)_{n\in\mathbb{N}}$. A stochastic process is is *integrable* if $X_n \in L^1$ for all n.

Definition (Filtration)**.** A *filtration* is a sequence of σ-algebras $\mathcal{F}_n \subseteq \mathcal{F}$ such that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all n. We define $F_\infty = \sigma(\bigcup_{n=1}^\infty \mathcal{F}_n)$. The *natural filtration* of a stochastic process X is the filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. A stochastic process is *adapted* to a filtration $\mathcal{F}_n$ if X_n is $\mathcal{F}_n$-measurable for all n.

Definition (Martingale)**.** An integrable adapted process $(X_n)_{n\geq 0}$ is a *martingale* if for all $n \geq m$, we have

$$\mathbb{E}(X_n \mid \mathcal{F}_m) = X_m.$$

We say it is a *super-martingale* if

$$\mathbb{E}(X_n \mid \mathcal{F}_m) \leq X_m,$$

and a *sub-martingale* if

$$\mathbb{E}(X_n \mid \mathcal{F}_m) \geq X_m,$$

By the tower property, it is sufficient to check the martingale property for $n = m + 1$.

Theorem 2.1 (Doob decomposition, non-examinable)**.** *Let X_n be an integrable adapted process. Then there exists a martingale M_n and an integrable predictable process A_n such that $X_n = M_n + A_n$ and $A_0 = 0$, where predictable means that A_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$. Moreover, M_n and A_n are unique up to a.s. equivalence.*

Proof. (Existence) Add up the 'known' bits to get A and the 'surprises' to get M. Formally,

$$\begin{aligned} A_n &= A_{n-1} + \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} \\ M_n &= M_{n-1} + \underbrace{X_n - \mathbb{E}(X_n \mid \mathcal{F}_{n-1})}_{\text{surprise}} \end{aligned}$$

(Uniqueness) Let $X_n = M_n + A_n = M_n' + A_n'$. Then $M_n - M_n' = A_n' - A_n$ is $\mathcal{F}_{n-1}$-measurable. But $M_n - M_n'$ is a martingale, so $\mathbb{E}(M_n - M_n' \mid \mathcal{F}_{n-1}) = 0$ so $M_n = M_n'$ almost surely. Similarly, $A_n = A_n'$ almost surely. □

Definition (Stopping time)**.** A random variable $T : \Omega \to \mathbb{N} \cup \{\infty\}$ is a *stopping time* if $\{T \leq n\} \in \mathcal{F}_n$ for all n.

In the discrete case, we can equivalently require that $\{T = n\} \in \mathcal{F}_n$ for all n.

Definition (X_T)**.** Let X be a stochastic process and T a stopping time. Then $X_T : \Omega \to \mathbb{R}$ is defined by cases

$$X_T(\omega) = \begin{cases} X_n(\omega) & T(\omega) = n \\ 0 & T(\omega) = \infty \end{cases}$$

Definition (Stopped σ-algebra)**.** Let T be a stopping time. Then the *stopped σ-algebra* is

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n \text{ for all } n\}.$$

Example. Let $N = \#$ of times a random walk hits -5 before it first hits 10 and T be the first time the random walk hits 10. N is $\mathcal{F}_T$-measurable

Definition (Stopped process)**.** Let X be a stochastic process and T a stopping time. Then the *stopped process* is $X_n^T = X_{T\wedge n}$

Proposition.

1. If $T, S, (T_n)_{n\geq 0}$ are all stopping times, then

$$T \vee S, T \wedge S, \sup_n T_n, \inf T_n, \limsup T_n, \liminf T_n$$

   are all stopping times.

2. $\mathcal{F}_T$ is a σ-algebra
3. If $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.
4. $X_T \mathbf{1}_{T<\infty}$ is $\mathcal{F}_T$-measurable.
5. If (X_n) is an adapted process, then so is $(X_n^T)_{n\geq 0}$ for any stopping time T.
6. If (X_n) is an integrable process, then so is $(X_n^T)_{n\geq 0}$ for any stopping time T.

Proof.

1. Elementary
2. Elementary
3. Let $A \in \mathcal{F}_S$. For any n, we have $A \cap \{S \leq n\} \in \mathcal{F}_n$ and $A \cap \{T \leq n\} = A \cap \{S \leq n\} \cap \{T \leq n\} \in \mathcal{F}_n$.
4. $X_T \mathbf{1}_{T<\infty} = \sum_{n=1}^{\infty} X_n \mathbf{1}_{\{T=n\}}$ where each of the terms is $\mathcal{F}_T$-measurable.
5. $X_n^T = X_n \mathbf{1}_{\{T\geq n\}} + X_T \mathbf{1}_{\{T<n\}} \mathbf{1}_{\{T<\infty\}}$.
6. $X_n^T = X_n \mathbf{1}_{\{T\geq n\}} + \sum_{k=1}^{n-1} X_k \mathbf{1}_{\{T=k\}}$ so $E|X_n^T| \leq E|X_n| + \sum_{k=1}^{n-1} E|X_k| < \infty$.

$\square$

Theorem 2.2 (Equivalent definitions for super-martingales)**.** *Let $(X_n)_{n\geq 0}$ be an integrable and adapted process. Then the following are equivalent:*

1. *$(X_n)_{n\geq 0}$ is a super-martingale.*
2. *For any bounded stopping times T and any stopping time S,*

$$\mathbb{E}(X_T \mid \mathcal{F}_S) \leq X_{S\wedge T}.$$

3. (X_n^T) *is a super-martingale for any stopping time* T.
4. *For bounded stopping times* S, T *such that* $S \le T$, *we have*

$$\mathbb{E}X_T \le \mathbb{E}X_S.$$

Proof.

- (2) $\Rightarrow$ (1): Let $n \ge m$ and set $T = n, S = m$.
- (2) $\Rightarrow$ (4): Tower rule
- (2) $\Rightarrow$ (3): Let $n \ge m$

$$\mathbb{E}(X_n^T \mid \mathcal{F}_m) = \mathbb{E}(X_{T\wedge n} \mid \mathcal{F}_m) \le X_{T\wedge m\wedge n} = X_m^T.$$

- (1) $\Rightarrow$ (2) Let $T \le N$

$$X_T = X_{S\wedge T} + \sum_{k=0}^{N} (X_{k+1} - X_k)\mathbf{1}_{S\le k<T} \qquad (*)$$

  Let $A \in \mathcal{F}_S$.

$$\begin{aligned}
\mathbb{E}\left[(X_{k+1} - X_k)\mathbf{1}_{S\le k<T}\mathbf{1}_A\right] &= \mathbb{E}\left[\mathbb{E}\left[(X_{k+1} - X_k)\underbrace{\mathbf{1}_{S\le k<T}\mathbf{1}_A}_{\in\mathcal{F}_k} \mid \mathcal{F}_k\right]\right] \\
&= \mathbb{E}\left[\mathbf{1}_{S\le k<T}\mathbf{1}_A \underbrace{\mathbb{E}\left[(X_{k+1} - X_k) \mid \mathcal{F}_k\right]}_{\le 0}\right] \\
&\le 0
\end{aligned}$$

  so $\mathbb{E}X_T\mathbf{1}_A \le \mathbb{E}X_{S\wedge T}\mathbf{1}_A$. By Radon-Nikodym, $\mathbb{E}(X_{S\wedge T} - X_T \mid \mathcal{F}_S) \ge 0$. But $X_{S\wedge T}$ is $\mathcal{F}_S$-measurable, so $X_{S\wedge T} - X_T \ge 0$ almost surely.

- (4) $\Rightarrow$ (2) Let $n \ge m$ and $A \in \mathcal{F}_m$. One can check that $T = m\mathbf{1}_A + n\mathbf{1}_{A^c} \le n$ is a stopping time such that

$$\mathbb{E}((X_n - X_m)\mathbf{1}_A) = \mathbb{E}(X_n - X_T) \le 0$$

  By Radon-Nikodym, $\mathbb{E}(X_m - X_n \mid \mathcal{F}_m) \ge 0$ so $\mathbb{E}(X_n \mid \mathcal{F}_m) \le X_m$.

- (3) $\Rightarrow$ (1) Let $T = \infty$

$\square$

Theorem 2.3 (Optional stopping)**.** *Let* $(X_n)_{n\ge 0}$ *be a martingale and* T *a stopping time. Then* $E(X_T) = E(X_0)$ *if any of the following conditions hold:*

1. T *is almost surely bounded, i.e. there is some* N *such that* $T \le N$ *almost surely.*

2. *X has bounded increments, i.e. there is some K such that $|X_{n+1} - X_n| \leq K$ for all n almost surely and T is integrable*

3. *There exists an integrable random variable Y such that $|X_n| \leq Y$ for all n almost surely and T is finite almost surely, i.e. $\mathbb{P}(T < \infty) = 1$.*

Proof.

1. Use (4) of the previous theorem with $S = 0$, or prove directly.
2. placeholder
3. placeholder

$\square$