1 Conditional Expectation

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $(X_i)_{i\in I}$ be a collection of random variables defined on this space. Then we define $\sigma(X_i : i \in I) \subseteq \mathcal{F}$ to be the smallest σ-algebra such that all of the X_i are measurable, i.e

$$\sigma(X_i : i \in I) = \sigma(X_i^{-1}(B) : i \in I,\ B \in \mathcal{B}(\mathbb{R})).$$

Definition. If $B \in \mathcal{F}$ has $\mathbb{P}(B) > 0$ then we define

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

for any $A \in \mathcal{F}$. Furthermore, if X is an integrable random variable we define

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbb{1}(B)]}{\mathbb{P}(B)}.$$

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say a σ-algebra $\mathcal{G}$ is countably generated if there exist $(B_i)_{i \in I}$ pairwise disjoint (with I countable) such that $\bigcup_{i \in I} B_i = \Omega$ and $\mathcal{G} = \sigma(B_i : i \in I)$.

Let X be an integrable random variable and $\mathcal{G}$ a countably generated σ-algebra. We want to define $X' = \mathbb{E}[X|\mathcal{G}]$. So define

$$X'(\omega) = \mathbb{E}[X|B_i] \text{ whenever } \omega \in B_i.$$

Or equivalently,

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

where we use the convention that $\mathbb{E}[X|B_i] = 0$ if $\mathbb{P}(B_i) = 0$. Then X' is indeed $\mathcal{G}$-measurable (note $\mathcal{G}$ is the set of $\bigcup_{j \in J} B_j$ for $J \subseteq I$).

Note that for any $G \in \mathcal{G}$ we have $\mathbb{E}[X \mathbb{1}(G)] = \mathbb{E}[X' \mathbb{1}(G)]$. Also

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

so X' is integrable.

Theorem (Monotone convergence theorem)**.** *Let $(X_n)_{n \geq 1}$ be a sequence of non-negative random variables with $X_n \uparrow X$ as $n \to \infty$ almost surely. Then $\mathbb{E}X_n \uparrow \mathbb{E}X$ as $n \to \infty$.*

Proof. See Part II Probability & Measure. □

Theorem (Dominated convergence theorem)**.** *Let $(X_n)_{n \geq 1}$ be a sequence of random variables with $X_n \to X$ as $n \to \infty$ almost surely and $|X_n| \leq Y$ almost surely for some Y integrable. Then $\mathbb{E}X_n \to \mathbb{E}X$ as $n \to \infty$.*

Proof. See Part II Probability & Measure. □

Definition (L^p)**.** Let $p \in [1, \infty]$ and f be a measurable function. Define the *L^p-norm*

$$||f||_p = \left(\mathbb{E}[|f|^p]\right)^{1/p} \text{ for } p \in [1, \infty)$$
$$||f||_\infty = \inf\{\lambda : |f| \leq \lambda \text{ a.e}\}.$$

Furthermore write $f \sim g$ if $f = g$ almost everywhere. Then define the L^p-space $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) = \{f : ||f||_p < \infty\}/\sim$.

Theorem ($\mathcal{L}^2$ is a Hilbert space)**.** *$\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space with inner product $\langle U, V \rangle = \mathbb{E}[UV]$. For a closed subspace $\mathcal{H}$, if $f \in \mathcal{L}^2$ there exists a unique $g \in \mathcal{H}$ with $||f - g||_2 = \inf\{||f - h||_2 : h \in \mathcal{H}\}$ and $\langle f - g, h \rangle = 0$ for all $h \in \mathcal{H}$. g is called the orthogonal projection of f on $\mathcal{H}$.*

Proof. See Part II Probability & Measure. □

Theorem (Conditional expectation)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra, $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists an integrable random variable Y satisfying*

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

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

Moreover Y is unique, in the sense that if Y' also satisfies (a) and (b), then $Y = Y'$ almost surely. We call Y a version of the conditional expectation of X given $\mathcal{G}$. We write $Y = \mathbb{E}[X|\mathcal{G}]$ almost surely. If $\mathcal{G} = \sigma(Z)$ for a random variable Z, then we write $\mathbb{E}[X|Z] = \mathbb{E}[X|\mathcal{G}]$.

Remark. (b) could be replaced by $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ for all Z bounded and $\mathcal{G}$-measurable.

Proof. First we show uniqueness. Suppose Y and Y' both satisfy (a) and (b) and let $A = \{Y > Y'\} \in \mathcal{G}$. Then

$$\mathbb{E}[Y\mathbb{1}(A)] = \mathbb{E}[Y'\mathbb{1}(A)] \Rightarrow \mathbb{E}[(Y-Y')\mathbb{1}(A)] = 0 \Rightarrow \mathbb{P}(Y > Y') = 0 \Rightarrow Y \leq Y' \text{ a.s}$$

and similarly $Y \geq Y'$ a.s.

Now we show existence. First assume $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $\mathcal{L}^2(\mathcal{F})$. Hence

$$\mathcal{L}^2(\mathcal{F}) = \mathcal{L}^2(\mathcal{G}) \oplus \mathcal{L}^2(\mathcal{G})^\perp$$

so we can write $X = Y + Z$ for $Y \in \mathcal{L}^2(\mathcal{G})$ and $Z \in \mathcal{L}^2(\mathcal{G})^\perp$. Define $\mathbb{E}[X|\mathcal{G}] = Y$, so Y is $\mathcal{G}$-measurable and for all $A \in \mathcal{G}$

$$\mathbb{E}[X\mathbb{1}(A)] = \mathbb{E}[Y\mathbb{1}(A)] + \underbrace{\mathbb{E}[Z\mathbb{1}(A)]}_{=0} = \mathbb{E}[Y\mathbb{1}(A)].$$

We claim that if $X \geq 0$ almost surely, then $Y \geq 0$ almost surely. Indeed, let $A = \{Y < 0\} \in \mathcal{G}$ so $0 \leq \mathbb{E}[X\mathbb{1}(Y < 0)] = \mathbb{E}[Y\mathbb{1}(Y < 0)] \leq 0$ which implies $\mathbb{P}(Y < 0) = 0$.

Assume now that $X \geq 0$ almost surely. Define $X_n = X \wedge n \leq n$, so $X_n \in \mathcal{L}^2$ for all n. Let $Y_n = \mathbb{E}[X_n|\mathcal{G}]$. Then X_n is an increasing sequence and by the above claim, Y_n is also an increasing sequence almost surely. Define $Y = \limsup_{n\to\infty} Y_n$, so Y is $\mathcal{G}$-measurable. Also $Y =\uparrow \lim_{n\to\infty} Y_n$ almost surely. For any $A \in \mathcal{G}$ we have

$$\mathbb{E}[X\mathbb{1}(A)] = \lim_{n\to\infty} \mathbb{E}[X_n\mathbb{1}(A)] = \lim_{n\to\infty} \mathbb{E}[Y_n\mathbb{1}(A)] = \mathbb{E}[Y\mathbb{1}(A)]$$

by the Monotone Convergence Theorem.

Finally, for general X write $X = X^+ - X^-$ and define $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}]$. $\square$

Remark. From the last proof we can see that we can define $\mathbb{E}[X|\mathcal{G}]$ for $X \geq 0$ without assuming integrability of X. It satisfies all the conditions apart from integrability.

Definition. Let $(\mathcal{G}_n)_{n\geq 1}$ be sub σ-algebras of $\mathcal{F}$. We call them *independent* if whenever $G_i \in \mathcal{G}_i$ and $i_1 < i_2 < \ldots < i_k$ we have

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

For a random variable X and a σ-algebra $\mathcal{G}$, we say they are *independent* if $\sigma(X)$ is independent of $\mathcal{G}$.

Properties of conditional expectation

Let $X, Y \in \mathcal{L}^1$, $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra. Then

1. $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$ (take $A = \Omega$);
2. If X is $\mathcal{G}$-measurable then $\mathbb{E}[X|\mathcal{G}] = X$ almost surely (X clearly satisfies the conditions);
3. If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ almost surely;
4. If $X \geq 0$ almost surely then $\mathbb{E}[X|\mathcal{G}] \geq 0$ almost surely;
5. For $\alpha, \beta \in \mathbb{R}$, $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\mathbb{E}[X|\mathcal{G}] + \beta\mathbb{E}[Y|\mathcal{G}]$ almost surely;
6. $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$ almost surely.

Recall:

Theorem (Fatou's Lemma)**.** *If $X_n \geq 0$ for all n almost surely, then*

$$\mathbb{E}[\liminf_{n\geq 1} X_n] \leq \liminf_{n\geq 1} \mathbb{E}X_n.$$

Proof. See Part II Probability & Measure. □

Theorem (Jensen's Inequality)**.** *If X is integrable, $\varphi : \mathbb{R} \to \mathbb{R}$ is convex, then*

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

We consider any analogues of our convergence theorems for conditional expectation.

Theorem (Conditional Monotone Convergence Theorem)**.** *Suppose $X_n \geq 0$ for all n and $X_n \uparrow X$ almost surely as $n \to \infty$. Let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$ almost surely.*

Remark. Note that $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$ in the almost-sure sense, as these are random variables.

Proof. Let $Y_n = \mathbb{E}[X_n|\mathcal{G}]$ almost surely. Then Y_n is increasing. Set $Y = \limsup_{n\geq 1} Y_n$. Since Y_n is $\mathcal{G}$-measurable, Y is $\mathcal{G}$-measurable. Also $Y =\uparrow \lim_{n\geq 1} Y_n$ almost surely. We need to show $\mathbb{E}[Y\mathbb{1}(A)] = \mathbb{E}[X\mathbb{1}(A)]$ for all $A \in \mathcal{G}$. This follows from the usual Monotone Convergence Theorem as

$$\mathbb{E}[Y\mathbb{1}(A)] = \lim_{n\geq 1} \mathbb{E}[Y_n\mathbb{1}(A)] = \lim_{n\geq 1} \mathbb{E}[X_n\mathbb{1}(A)] = \mathbb{E}[X\mathbb{1}(A)].$$

□

Theorem (Conditional Fatou's Lemma)**.** *Let $(X_n)_{n\geq 1}$ be a non-negative sequence of random variables. Then*

$$\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \text{ almost surely.}$$

Proof. Note that $\inf_{k\geq n} X_k \uparrow \liminf_{n\to\infty} X_n$ so by the conditional MCT

$$\lim_{n\to\infty} \mathbb{E}[\inf_{k\geq n} X_k|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}].$$

We also have

$$\mathbb{E}[\inf_{k\geq n} X_k|\mathcal{G}] \leq \mathbb{E}[X_k|\mathcal{G}] \quad \forall k \geq n \text{ almost surely.}$$

Which implies

$$\mathbb{E}[\inf_{k\geq n} X_k|\mathcal{G}] \leq \inf_{k\geq n} \mathbb{E}[X_k|\mathcal{G}] \quad \forall k \geq n \text{ almost surely}$$

since k takes countable values (intersection of countable sets of full measure also has full measure). Now taking limits as $n \to \infty$ we are done. □

Theorem (Conditional Dominated Convergence Theorem)**.** *Suppose $X_n \to X$ almost surely, $|X_n| \leq Y$ almost surely with Y integrable. Then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ almost surely.*

Proof. We apply the Conditional Fatou's Lemma. Indeed $-Y \leq X_n \leq Y$ so $X_n + Y \geq 0$ and $Y - X_n \geq 0$ for all n. By Conditional Fatou's Lemma

$$\mathbb{E}[X|\mathcal{G}]+\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[X+Y|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty}(X_n+Y)] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]+\mathbb{E}[Y|\mathcal{G}]$$

and

$$\mathbb{E}[Y|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty}(Y - X_n)|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}] + \liminf_{n\to\infty}\left(-\mathbb{E}[X_n|\mathcal{G}]\right).$$

Hence $\limsup_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \leq \mathbb{E}[X|\mathcal{G}]$ and $\liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \geq \mathbb{E}[X|\mathcal{G}]$ almost surely.

□

Theorem (Conditional Jensen's Inequality)**.** *Let X be integrable, $\varphi : \mathbb{R} \to \mathbb{R}$ a convex function such that $\varphi(X)$ is integrable or $\varphi(X) \geq 0$. Then $\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbb{E}[X|\mathcal{G}])$ almost surely.*

Proof. We claim that $\varphi(x) = \sup_{i\in\mathbb{N}}(a_i x + b_i)$, $a_i, b_i \in \mathbb{R}$.

Then $\varphi(X) = \sup_{i\in\mathbb{N}}(a_i X + b_i)$. So

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \sup_{n\geq 1}(a_i \mathbb{E}[X|\mathcal{G}] + b_i) \quad \forall i \in \mathbb{N} \text{ almost surely.}$$

□

Note. We need the supremum in the claim to be over a countable set so we can preserve the almost-sue property of an inequality.

Corollary. *For all $p \in [1, \infty)$ we have*

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

Proof. Apply conditional Jensen ($x \mapsto x^p$ is convex). □

Theorem (Tower property)**.** *Let X be integrable and $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ sub σ-algebras. Then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] \text{ almost surely.}$$

Proof. $\mathbb{E}[X|\mathcal{H}]$ is certainly $\mathcal{H}$-measurable so it remains to check

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(A)] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbb{1}(A)] \quad \forall A \in \mathcal{H}.$$

But since $A \in \mathcal{G}$ whenever $A \in \mathcal{H}$ we have

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(A)] = \mathbb{E}[X\mathbb{1}(A)] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbb{1}(A)].$$

□

Proposition. Let $X \in \mathcal{L}^1$, $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra, Y bounded and $\mathcal{G}$-measurable. Then

$$\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}] \text{ almost surely.}$$

Proof. $Y\mathbb{E}[X\mathcal{G}]$ is certainly $\mathcal{G}$-measurable. Also for any $A \in \mathcal{G}$

$$\mathbb{E}[XY\mathbb{1}(A)] = \mathbb{E}[X \underbrace{(Y\mathbb{1}(A))}_{\substack{\text{bounded,}\\ \mathcal{G}\text{-measurable}}}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}](Y\mathbb{1}(A))].$$

□

Definition. Let $\mathcal{A}$ be a collection of sets. It is called a *π-system* if whenever $A, B \in \mathcal{A}$ we have $A \cap B \in \mathcal{A}$.

Recall

Theorem (Uniqueness of extension)**.** *Let $(E, \mathcal{E})$ be a measurable space and let $\mathcal{A}$ be a π-system generating $\mathcal{E}$. Let μ, ν be two measures on $(E, \mathcal{E})$ with $\mu(E) = \nu(E) < \infty$. If $\mu = \nu$ on $\mathcal{A}$. then $\mu = \nu$ on $\mathcal{E}$.*

Proof. See Part II Probability & Measure. $\square$

Theorem. *Let $X \in \mathcal{L}^1$, $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ sub σ-algebras. Assume $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$. Then*

$$\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}] \text{ almost-surely.}$$

Proof. We need to show $\mathbb{E}[X\mathbb{1}(F)] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(F)]$ for all $F \in \sigma(\mathcal{G}, \mathcal{H})$. Define $\mathcal{A} = \{A \cap B : A \in \mathcal{G}, B \in \mathcal{H}\}$. This is a π-system generating $\sigma(\mathcal{G}, \mathcal{H})$. If $F = A \cap B$, $A \in \mathcal{G}, B \in \mathcal{H}$ then

$$\begin{aligned}\mathbb{E}[X\mathbb{1}(A \cap B)] &= \mathbb{E}[\underbrace{(X\mathbb{1}(A))}_{\sigma(X,\mathcal{G})\text{measurable}} \mathbb{1}(B)] \\ &= \mathbb{E}[X\mathbb{1}(A)]\mathbb{P}(B) \\ &= \mathbb{E}[\underbrace{\mathbb{E}[X|\mathcal{G}]\mathbb{1}(A)}_{\substack{\mathcal{G} \\ \text{measurable}}}]\mathbb{P}(B) \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(A)\mathbb{1}(B)].\end{aligned}$$

Assume $X \geq 0$. Define $\mu(F) = \mathbb{E}[X\mathbb{1}(F)]$ and $\nu(F) = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(F)]$ for $F \in \sigma(\mathcal{G}, \mathcal{H})$. Then $\mu = \nu$ on $\mathcal{A}$ by the above and $\mu(\Omega) = \nu(\Omega) < \infty$. Therefore $\mu = \nu$ on $\sigma(\mathcal{G}, \mathcal{H})$. $\square$

Definition. We say $(X_1, \ldots, X_n) \in \mathbb{R}^n$ has the *Gaussian distribution* iff for all $a_1, \ldots, a_n \in \mathbb{R}$

$$a_1X_1 + \ldots + a_nX_n$$

has the Gaussian distribution in $\mathbb{R}$.

A process $(X_t)_{t\geq 0}$ is called a *Gaussian process* if $\forall t_1 < t_2 < \ldots < t_n$, the vector $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian random vector.

Example. Let (X, Y) be a Gaussian vector in $\mathbb{R}^2$. We want to compute $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$. Let $X' = \mathbb{E}[X|Y]$. Since X' is $\sigma(Y)$-measurable it follows X' is a measurable function of Y. So are looking for f Borel such that $\mathbb{E}[X|Y] = f(Y)$ almost surely. Let $f(y) = ay + b$ for some $a, b \in \mathbb{R}$ to be determined.

Since $\mathbb{E}[X'] = \mathbb{E}[X]$ we have $a\mathbb{E}Y + b = \mathbb{E}X$. Also

$$\begin{aligned}\mathbb{E}[XY] = \mathbb{E}[X'Y] &\implies \mathbb{E}[(X - X')Y] = 0\\ &\implies \mathrm{Cov}(X - X', Y) = 0\\ &\implies \mathrm{Cov}(X, Y) = a\mathrm{Var}(Y)\end{aligned}$$

so we have determined a, b. We need to check that for any Z bounded and $\sigma(Y)$-measurable we have $\mathbb{E}[(X - X')Z] = 0$. Write $Z = g(Y)$ and note $\mathrm{Cov}(X - X', Y) = 0$, implying $X - X'$ is independent of Y. Therefore $\mathbb{E}[(X - X')g(Y)] = \mathbb{E}[X - X']\mathbb{E}[g(Y)] = 0$.

Example. Let (X, Y) be a random vector in $\mathbb{R}^2$ with joint density function $f_{X,Y}(x, y)$. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function such that $h(X)$ is integrable. We want to compute $\mathbb{E}[h(X)|Y]$. Note

$$\mathbb{E}[h(X)g(Y)] = \int_{\mathbb{R}^2} h(x)g(x)f_{X,Y}(x, y)\mathrm{d}x\mathrm{d}y$$

and write

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\mathrm{d}x$$

for the density of Y. So (using the convention $0/0 = 0$)

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)\frac{f_{X,Y}(x, y)}{f_Y(y)}\mathrm{d}x \right) g(y)f_Y(y)\mathrm{d}y$$

define

$$\varphi(y) = \begin{cases} \int_{\mathbb{R}} h(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}\mathrm{d}x & \text{if } f_Y(y) > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Then $\mathbb{E}[h(X)|Y] = \varphi(Y)$ almost surely.

2 Martingales

2.1 Discrete-time Martingales

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* is a sequence of increasing sub σ-algebras of $\mathcal{F}$, $(\mathcal{F}_n)_{n\geq 0}$, $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. We call $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ a *filtered probability space*.

If $X = (X_n)_{n\geq 0}$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, define $\mathcal{F}_n^X = \sigma(X_k : k \leq n)$, the *natural filtration* associated with X. We say X is *adapted* to a filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable for all n. X is *integrable* if X_n is integrable for all n.

Definition. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. We say an integrable adapted process $X = (X_n)_{n\geq 0}$ is called a

- *martingale* if
$$\mathbb{E}[X_n|\mathcal{F}_m] = X_m \text{ almost surely } \forall n \geq m.$$
- *super-martingale* if
$$\mathbb{E}[X_n|\mathcal{F}_m] \leq X_m \text{ almost surely } \forall n \geq m.$$
- *sub-martingale* if
$$\mathbb{E}[X_n|\mathcal{F}_m] \geq X_m \text{ almost surely } \forall n \geq m.$$

Remark. If X is a martingale with respect to $(\mathcal{F}_n)$, then it is also a martingale with respect to the natural filtration $(\mathcal{F}_n^X)$.

Example. Let (ξ_i) be a sequence of iid random variables with $\mathbb{E}[\xi_1] = 0$. Let $X_n = \xi_1 + \ldots + \xi_n$, $X_0 = 0$. This is a martingale. We have
$$\mathbb{E}[X_n|\mathcal{F}_{n-1}] = \xi_1 + \ldots + \xi_{n-1} + \mathbb{E}[\xi_n|\mathcal{F}_{n-1}] = \xi_1 + \ldots + \xi_{n-1}$$
by independence.

Example. Let (ξ_i) be a sequence of iid random variables with $\mathbb{E}[\xi_1] = 1$. Let $X_n = \prod_{i=1}^n \xi_i$, $X_0 = 1$. This is a martingale.

Definition. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. A *stopping time* T is a random variable $T : \Omega \to \mathbb{Z}_+ \cup \{\infty\}$ such that $\{T \leq n\} \in \mathcal{F}_n$ for all n.

Note. T being a stopping time is equivalent to $\{T = n\} \in \mathcal{F}_n$ for all n.

Examples.

- Constant times are trivial stopping times;
- Suppose $(X_n)_{n\geq 0}$ is an adapted process taking values in $\mathbb{R}$. For $A \in \mathcal{B}$ define $T_A = \inf\{n \geq 0 : X_n \in A\}$ (with the convention that $\inf \emptyset = \infty$). Then $\{T_A \leq n\} = \bigcup_{k\leq n}\{X_k \in A\} \in \mathcal{F}_n$, so T_A is a stopping time;
- In the setting above, let $L_A = \sup\{n \geq 0 : X_n \in A\}$. This is in general not a stopping time.

Proposition. Let $S, T, (T_n)$ be stopping times. Then $S \wedge T$, $S \vee T$, $\inf T_n$, $\sup T_n$, $\liminf T_n$ and $\limsup T_n$ are also stopping times.

Proof. Follows directly from the definition. □

Definition. If T is a stopping time, we define

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t,\ \forall t\}.$$

If $(X_n)_{n\geq 0}$ is a process, write $X_T(\omega) = X_{T(\omega)}(\omega)$ whenever $T(\omega) < \infty$. We define the *stopped process* $X_t^T = X_{T\wedge t}$.

Proposition. Let S and T be stopping times and let X be an adapted process. Then

1. If $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$;
2. $X_T \mathbb{1}(T < \infty)$ is $\mathcal{F}_T$-measurable;
3. X^T is adapted;
4. If X is integrable, then X^T is also integrable.

Proof.

1. Immediate from the definition;
2. Let $A \in \mathcal{B}(\mathbb{R})$. We need to show $\{X_T \mathbb{1}(T < \infty) \in \mathcal{A}\} \in \mathcal{F}_T$. Note that

$$\{X_T \mathbb{1}(T < \infty) \in A\} \cap \{T \leq t\} = \bigcup_{s=0}^{t} \underbrace{\{X_s \in A\}}_{\in \mathcal{F}_s \subseteq \mathcal{F}_t} \cap \underbrace{\{T = s\}}_{\in \mathcal{F}_s} \in \mathcal{F}_t.$$

3. $X_t^T = X_{T\wedge t}$ is $\mathcal{F}_{T\wedge t}$-measurable so $\mathcal{F}_t$-measurable by (1).

4. We have

$$\begin{aligned}\mathbb{E}[|X_t^T|] = \mathbb{E}[|X_{T\wedge t}|] &= \sum_{s=0}^{t-1} \mathbb{E}[|X_s|\mathbb{1}(T=s)] + \mathbb{E}[|X_t|\mathbb{1}(T \geq t)] \\ &\leq \sum_{s=0}^{t} \mathbb{E}[|X_s|] < \infty.\end{aligned}$$

□

Theorem (Optional Stopping Theorem)**.** *Let (X_n) be a martingale.*

1. *If T is a stopping time, then X^T is also a martingale. In particular $\mathbb{E}[X_{T\wedge t}] = \mathbb{E}[X_0]$ for all t;*
2. *If $S \leq T$ are bounded stopping times then $\mathbb{E}[X_T|\mathcal{F}_S] = X_S$ almost surely, and $\mathbb{E}[X_T] = \mathbb{E}[X_S]$;*
3. *If there exists an integrable random variable Y such that $|X_n| \leq Y$ for all n, and T is finite almost surely then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$;*
4. *If there exists $M > 0$ such that $|X_{n+1} - X_n| \leq M$ for all n, and T is a stopping time with $\mathbb{E}T < \infty$, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.*

Proof.

1. We need to show that for all t we have

$$\mathbb{E}[X_{T\wedge t}|\mathcal{F}_{t-1}] = X_{T\wedge(t-1)}$$

almost surely. Indeed

$$\begin{aligned}\mathbb{E}[X_{T\wedge t}|\mathcal{F}_{t-1}] &= \mathbb{E}\left[\sum_{s=0}^{t-1} X_s \mathbb{1}(T=s)|\mathcal{F}_{t-1}\right] + \mathbb{E}[X_t \mathbb{1}(T \geq t)|\mathcal{F}_{t-1}] \\ &= \sum_{s=0}^{t-1} X_s \mathbb{1}(T=s) + \mathbb{1}(T \geq t) X_{t-1} \\ &= X_{T\wedge(t-1)}\end{aligned}$$

using the fact that $\mathbb{1}(T \geq t)$ is $\mathcal{F}_{t-1}$-measurable;

2. Suppose $S \leq T \leq n$ and let $A \in \mathcal{F}_S$. We need to show $\mathbb{E}[X_T \mathbb{1}(A)] = \mathbb{E}[X_S \mathbb{1}(A)]$. Note

$$\begin{aligned}X_T - X_S &= (X_T - X_{T-1}) + \ldots + (X_{S+1} - X_S) \\ &= \sum_{k\geq 0} (X_{k+1} - X_k)\mathbb{1}(S \leq k < T) \\ &= \sum_{k=0}^{n} (X_{k+1} - X_k)\mathbb{1}(S \leq k < T). \qquad (T \leq n)\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}[X_T \mathbb{1}(A)] &= \mathbb{E}[X_S \mathbb{1}(A)] + \sum_{k=0}^{n} \mathbb{E}[(X_{k+1} - X_k) \underbrace{\mathbb{1}(S \le k < T)\mathbb{1}(A)}_{\in \mathcal{F}_k}] \\ &= \mathbb{E}[X_S \mathbb{1}(A)]\end{aligned}$$

since $\mathbb{E}[X_{k+1}|\mathcal{F}_k] = X_k$ almost surely. Taking expectations gives $\mathbb{E}[X_T] = \mathbb{E}[X_S]$;

3. Example Sheet;

4. Example Sheet.

□

Note. Analogous results follow if (X_n) is instead a sub/super-martingale.

Corollary. *If X is a positive super-martingale, T is a stopping time, $T < \infty$ almost surely, then $\mathbb{E}[X_T] \le \mathbb{E}[X_0]$.*

Proof. Fatou's lemma gives $\mathbb{E}[\liminf_t X_{T\wedge t}] \le \liminf_t \mathbb{E}[X_{T\wedge t}] \le \mathbb{E}[X_0]$. □

Example. Let $(\xi_i)_{i\ge 0}$ be iid with $\mathbb{P}(\xi_0 = 1) = \mathbb{P}(\xi_0 = -1) = 1/2$. Define $X_0 = 0$ and $X_n = \sum_{i=1}^{n} \xi_i$ for $n \ge 1$. Then $(X_n)_{n\ge 0}$ is a martingale. Define $T = \inf\{n \ge 0 : X_n = 1\}$. Then $\mathbb{P}(T < \infty) = 1$ and for all t we have $\mathbb{E}[X_{T\wedge t}] = 0$, while $\mathbb{E}[X_T] = 1$. Hence (4) from the previous theorem tells us $\mathbb{E}T = \infty$.