# Probability II.

## Chapter I. Martingales

Probability space

σ-field　　probability measure

### Section 1.1 Conditional Expectation

Given a probability space $(\Omega, \mathcal{F}_0, \mathbb{P})$, a sub-σ-field $\mathcal{F} \subseteq \mathcal{F}_0$, and a random variable $X \in \mathcal{F}_0$ with $\mathbb{E}|X| < +\infty$. We define the conditional expectation of $X$ given $\mathcal{F}$, $\mathbb{E}(X \mid \mathcal{F})$, to be any random variable $Y$ satisfying that:

(i) $Y \in \mathcal{F}$, i.e., $Y$ is $\mathcal{F}$-measurable.

(ii) $\forall A \in \mathcal{F}$, $\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$. (i.e. $\underset{\mathbb{E}(X;A)}{\mathbb{E}(X \mathbb{1}_A)} = \underset{\mathbb{E}(Y;A)}{\mathbb{E}(Y \mathbb{1}_A)}$)

Then, we say $Y$ is a version of $\mathbb{E}(X \mid \mathcal{F})$.

**Theorem:** The conditionl expectation defined above exists and is unique in the following sense: if $Y'$ is another random variable satisfying properties (i)-(ii), then $Y = Y'$ a.s.

$\boxed{\text{Proof } \cancel{\text{of}} \text{ uniqueness}}$:

Lemma 1. If $Y$ satisfies (i) and (ii), then it is integrable and there is $\mathbb{E}|Y| \leq \mathbb{E}|X|$.

$\forall A \in \mathcal{F}$, we have $\mathbb{E}[(Y-Y')\mathbb{1}_A] = \int_A (Y - Y') \, d\mathbb{P} = 0 \Rightarrow Y = Y'$ $\mathcal{F}$-a.s.

Take $A = \{Y - Y' > 0\}$ and $A = \{Y - Y' < 0\}$ respectively. □

Technically, $Y = \mathbb{E}(X \mid \mathcal{F})$ should be written as $Y = \mathbb{E}(X \mid \mathcal{F})$ a.s. but we will ignore this notational complexity.

As an extension of the above result, we claim that:

Claim: If $X_1 = X_2$ on $B \in \mathcal{F}$, then $\mathbb{E}(X_1 \mid \mathcal{F}) = \mathbb{E}(X_2 \mid \mathcal{F})$ a.s. on $B$.

$\nu = \nu^+ - \nu^-$ , $|\nu| = \nu^+ + \nu^-$

Two positive measures called the positive and negative variations of $\nu$.

Pf: $Y_1 = \mathbb{E}(X_1 | \mathcal{F})$, $Y_2 = \mathbb{E}(X_2|\mathcal{F})$. For any $A \in \mathcal{F}$, we have that

$$\int_{A\cap B} X_1 \, d\mathbb{P} = \int_{A \cap B} X_2 \, d\mathbb{P} \Rightarrow \int_{A\cap B} (Y_1 - Y_2)\, d\mathbb{P} = 0 . \Rightarrow Y_1 = Y_2 \text{ a.s.}$$

Take $A = \{ Y_1 - Y_2 \overset{>0}{\cancel{\geq 0}} \}$ and $A = \{ Y_1 - Y_2 < 0\}$, respectively. □

$\boxed{\text{Pf of Existence}}$ We recall that a (complex/signed) ~~positive~~ measure $\nu$ is said to be absolutely continuous with respect to $\mu$ ($\nu \ll \mu$) if $|\nu|(A) = 0$ whenever $\mu(A) = 0$.

a positive measure

Radon-Nikodym Theorem: Let $\mu$ and $\nu$ be $\sigma$-finite measures on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, then there exists a function $f \in \mathcal{F}$ such that for all $A \in \mathcal{F}$,

$$\int_A f \, d\mu = \nu(A) .$$

We denote $f$ by $f = d\nu / d\mu$, which is also called the Randon-Nikodym derivative.

We use this theorem to show the existence of conditional expectation.

Let $\mu = \mathbb{P}$ and define $\nu$ on $\mathcal{F}$ as

$$\nu(A) = \int_A X \, d\mathbb{P} , \quad |\nu|(A) = \int_A |X| \, d\mathbb{P} .$$

By definition, we easily see that $\nu$ is a signed measure with $\nu \ll \mathbb{P}$.

Let $f = d\nu / d\mathbb{P}$ be the Randon-Nikodym derivative. Then, we know $f \in \mathcal{F}$ and $\forall A \in \mathcal{F}$,

$$\int_A f \, d\mathbb{P} = \int_A \frac{d\nu}{d\mathbb{P}} \, d\mathbb{P} = \nu(A) = \int_A X \, d\mathbb{P} .$$

Hence, $f = \frac{d\nu}{d\mathbb{P}}$ is a version of $\mathbb{E}(X|\mathcal{F})$. □

Intuitively, $\mathcal{F}$ describes the information we have at our disposal: $\forall A \in \mathcal{F}$, we know whether $A$ has occured or not. Then, we think $\mathbb{E}(X|\mathcal{F})$ as our "best guess" of $X$ given the information we have.



Claim: (1) (Perfect information) If $X$ is $\mathcal{F}$-measurable, then $\mathbb{E}(X|\mathcal{F}) = X$.

(2) (No information) If $\mathcal{F} = \{\Omega, \phi\}$, then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}X$.

Pf: Check through definition. □

Example: (Conditional ~~Pro~~ Probability) Let $B$ be a probability event with $0 < \mathbb{P}(B) < 1$. Let $\mathcal{F}$ be the $\sigma$-field generated by $B$, i.e., $\mathcal{F} = \{\phi, \Omega, B, B^c\}$. Then, we ~~define~~ claim that

$$\mathbb{E}(X|\mathcal{F}) = \frac{\mathbb{E}(X \mathbb{1}_B)}{\mathbb{P}(B)} \mathbb{1}_B + \frac{\mathbb{E}(X \mathbb{1}_{B^c})}{\mathbb{P}(B^c)} \mathbb{1}_{B^c} . \quad (*)$$

As a convention, we define the notations

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)} , \quad \mathbb{E}(X|B^c) = \frac{\mathbb{E}(X \mathbb{1}_{B^c})}{\mathbb{P}(B^c)} .$$

To see (i), we first observe that the proposed formula is constant on $B$ & $B^c$, so it is measurable w.r.t. $\mathcal{F}$. To see (ii), we have

$$\int_B \mathbb{E}(X|\mathcal{F}) \, d\mathbb{P} = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)} \mathbb{P}(B) = \mathbb{E}(X \mathbb{1}_B) .$$

To establish the connection with undergraduate notations, we denote

$$\mathbb{P}(A|\mathcal{F}) := \mathbb{E}(\mathbb{1}_A | \mathcal{F}) \text{ and } \mathbb{P}(A|B) = \mathbb{E}(\mathbb{1}_A | B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} .$$

Example: (Conditional Probability density) As a special case of the conditional expectation, we define $\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$, $\sigma(Y)$ – the $\sigma$-field generated by the random variable $Y$.

Consider two random variables $X$ & $Y$ with joint probability density $f(x,y)$. Suppose the marginal density $f_Y(y) = \int f(x,y)\,dx > 0$ for all $y$.

Let $g$ be a function s.t. $\mathbb{E}|g(X)| < +\infty$. We claim that

$$(**) \quad \mathbb{E}[g(X)|Y] = h(Y) \text{ with } h(y) = \int g(x) f_{X|Y}(x,y)\, dx$$

where $f_{X|Y}(x,y)$ is the conditional density $f_{X|Y=y}(x,y) = \frac{f(x,y)}{f_Y(y)}$   $f_{X|Y=y}(x)$

$$\left[\ \mathbb{P}(X=x \mid Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} = \frac{f(x,y)}{\int f(x,y)\,dx}\ \right]$$

To prove (**), we first observe that $h(Y) \in \sigma(Y)$. Next, pick any $A \in \sigma(Y)$. we have $A = Y^{-1}(B)$ for some Borel measurable $B$. Then,

$$\mathbb{E}(h(Y)\mathbb{1}_A) = \int_B h(y) f(x,y)\,dx\,dy$$

$A = \{\omega : Y(\omega) \in B\}$

Fubini

$$= \int_B \frac{\int g(x')f(x',y)\,dx'}{\int f(x,y)\,dx}\left(\int f(x,y)\,dx\right)dy$$

$$= \int_B \left(\int g(x) f(x,y)\,dx\right)dy$$

$$= \mathbb{E}[g(X)\mathbb{1}_B(Y)] = \mathbb{E}[g(X)\mathbb{1}_A].$$

Hence, $h(Y)$ satisfies (ii).

If $f_Y(y)=0$ for some $y$, we can let $h(y)$ be anything where $f_Y(y)=0$.

I.e., we let $h(y)\int f(x,y)\,dx = \int g(x) f(x,y)\,dx$.

## Properties of Conditional Expectation

**Lemma** (Some almost trivial properties) $X, Y \in L^1(\mathbb{P})$, $\mathcal{F}$ is a sub-$\sigma$-field.

(1) (Linearity) $\mathbb{E}(aX+bY \mid \mathcal{F}) = a\mathbb{E}(X\mid\mathcal{F}) + b\mathbb{E}(Y\mid\mathcal{F})$, $a, b \in \mathbb{C}$ are fixed.

(2) (Monotonicity) If $X \le Y$ a.s., then $\mathbb{E}(X\mid\mathcal{F}) \le \mathbb{E}(Y\mid\mathcal{F})$ a.s.

As a special case, if $X \ge 0$ a.s., then $\mathbb{E}(X\mid\mathcal{F}) \ge 0$ a.s.

(3) $\mathbb{E}(X\mid\mathcal{F}) \le \mathbb{E}(Y\mid\mathcal{F})$ a.s. if and only if $\mathbb{E}(X\mathbb{1}_A) \le \mathbb{E}(Y\mathbb{1}_A)$ $\forall A \in \mathcal{F}$.

(4) $|\mathbb{E}(X\mid\mathcal{F})| \le \mathbb{E}(|X| \mid \mathcal{F})$ a.s., and so $\mathbb{E}(|\mathbb{E}(X\mid\mathcal{F})|) \le \mathbb{E}|X|$.

Pf: (1) Check with definition.

(2) ~~We check that~~ $\cancel{\mathbb{E}(X-Y\mid\mathcal{F}) \le 0}$ follows from (3).

(3) "$\Rightarrow$" $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(\mathbb{E}(X\mid\mathcal{F})\mathbb{1}_A) \le \mathbb{E}(\mathbb{E}(Y\mid\mathcal{F})\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)$.

"$\Leftarrow$" Define $A_n := \{\mathbb{E}(X\mid\mathcal{F}) - \mathbb{E}(Y\mid\mathcal{F}) \ge \frac{1}{n}\}$. $A_n \in \mathcal{F}$.

Then, $\frac{1}{n}\mathbb{P}(A_n) \le \mathbb{E}[(\mathbb{E}(X\mid\mathcal{F}) - \mathbb{E}(Y\mid\mathcal{F}))\mathbb{1}_{A_n}]$

$= \mathbb{E}(X\mathbb{1}_{A_n}) - \mathbb{E}(Y\mathbb{1}_{A_n}) \le 0 \Rightarrow \mathbb{P}(A_n) = 0$.

Then, $\mathbb{P}(\mathbb{E}(X\mid\mathcal{F}) > \mathbb{E}(Y\mid\mathcal{F})) = \mathbb{P}(\bigcup_n A_n) = \lim_{n\to\infty}\mathbb{P}(A_n) = 0$.

(4) By (3), it suffices to show that $\forall A \in \mathcal{F}$,

$$\mathbb{E}[|\mathbb{E}(X\mid\mathcal{F})|\mathbb{1}_A] \le \mathbb{E}(|X|\mathbb{1}_A).$$

Let $A_+ = \{\mathbb{E}(X\mid\mathcal{F}) \ge 0\}$, $A_- = \{\mathbb{E}(X\mid\mathcal{F}) < 0\}$. Then

$$\mathbb{E}[|\mathbb{E}(X\mid\mathcal{F})|\mathbb{1}_{A_+}\mathbb{1}_A] = \mathbb{E}[\mathbb{E}(X\mid\mathcal{F})\mathbb{1}_{A_+\cap A}] = \mathbb{E}(X\mathbb{1}_{A_+\cap A}) \le \mathbb{E}[|X|\mathbb{1}_{A_+\cap A}].$$

$$\mathbb{E}[|\mathbb{E}(X\mid\mathcal{F})|\mathbb{1}_{A_-}\mathbb{1}_A] = -\mathbb{E}(X\mathbb{1}_{A_-\cap A}) \le \mathbb{E}[|X|\mathbb{1}_{A_-\cap A}].$$ □

**Theorem** (Convergence theorems for conditional expectation) Suppose $X$ and $\{X_n\}$ are random variables in $L^1(\Omega, \mathcal{F}_0, \mathbb{P})$ and $\mathcal{F}$ is a sub-$\sigma$-field.

(1) (MCT) If $X_n \ge 0$ and $X_n \uparrow X$, then $\mathbb{E}(X_n\mid\mathcal{F}) \uparrow \mathbb{E}(X\mid\mathcal{F})$ a.s.

(2) (Fatou's Lemma) If $X_n \ge 0$ a.s., then $\mathbb{E}(\liminf_n X_n \mid \mathcal{F}) \le \liminf_n \mathbb{E}(X_n\mid\mathcal{F})$ a.s.

(3) (DCT) If $X_n \to X$ a.s. and $\exists Z$ s.t. $\mathbb{E}|Z| < +\infty$ and $|X_n| \le Z$ for all $n$, then $\mathbb{E}(X_n\mid\mathcal{F}) \to \mathbb{E}(X\mid\mathcal{F})$ a.s.

Pf: (1) Due to montonicity of CE, we know $\mathbb{E}(X_n\mid\mathcal{F})\uparrow$. Define $Y = \lim_n \mathbb{E}(X_n\mid\mathcal{F})$. For any $A\in\mathcal{F}$, we have

$$\mathbb{E}(Y\mathbb{1}_A) = \int_A \lim_n \mathbb{E}(X_n\mid\mathcal{F})\,d\mathbb{P} = \lim_{n\to\infty}\int_A \mathbb{E}(X_n\mid\mathcal{F})\,d\mathbb{P}$$

$$= \lim_{n\to\infty}\int_A X_n\,d\mathbb{P} = \int_A X\,d\mathbb{P} = \mathbb{E}(X\mathbb{1}_A)$$

By original MCT

$\Rightarrow Y = \mathbb{E}(X\mid\mathcal{F})$.

(2) Let $Z_n = \inf_{m\ge n} X_m$. Note $\liminf_n X_n = \lim_n Z_n$, and $Z_n \uparrow$.

By MCT, we have $\mathbb{E}(\lim_n Z_n \mid \mathcal{F}) = \lim_n \mathbb{E}(Z_n \mid \mathcal{F}) \overset{a.s.}{\leq} \lim_n [\inf_{m\geq n} \mathbb{E}(X_m \mid \mathcal{F})]$

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

(3) Note $Z + X_n \geq 0$. By (2), we have

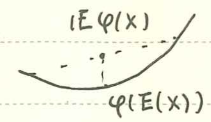
$$\mathbb{E}(Z + X \mid \mathcal{F}) \leq \liminf_n \mathbb{E}(Z + X_n \mid \mathcal{F}) \Rightarrow \mathbb{E}(X \mid \mathcal{F}) \overset{a.s.}{\leq} \liminf_n \mathbb{E}(X_n \mid \mathcal{F}).$$

Similarly, with $Z - X_n \geq 0$, we can derive $\mathbb{E}(X \mid \mathcal{F}) \geq \limsup_n \mathbb{E}(X_n \mid \mathcal{F})$. □

Theorem (Jensen's inequality) If $\varphi$ is convex and $\mathbb{E}|X|, \mathbb{E}|\varphi(X)| < +\infty$, then

$$\varphi(\mathbb{E}(X \mid \mathcal{F})) \leq \mathbb{E}(\varphi(X) \mid \mathcal{F}) \quad a.s.$$

Pf: Let $S = \{(a,b) : a, b \in \mathbb{Q},\ ax + b \leq \varphi(x)\ \forall x \in \mathbb{R}\}$.




Then, $\varphi(x) = \sup\{ax + b : (a,b) \in S\}$.

If $\varphi(x) \geq ax + b$, then $\mathbb{E}(\varphi(X) \mid \mathcal{F}) \geq a\,\mathbb{E}(X \mid \mathcal{F}) + b$, a.s.

Taking sup over $(a,b)$ gives $\mathbb{E}(\varphi(X) \mid \mathcal{F}) \geq \varphi(\mathbb{E}(X \mid \mathcal{F}))$ a.s. □

~~Corollary~~ Theorem (Tower property ~~Some key properties~~) $X \in L^1(\Omega, \mathcal{F}_0, \mathbb{P})$, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_0$.

Then, we have (1) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_1) \mid \mathcal{F}_2) = \mathbb{E}(X \mid \mathcal{F}_1)$ a.s.;

(2) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_2) \mid \mathcal{F}_1) = \mathbb{E}(X \mid \mathcal{F}_1)$ a.s.

In other words, "Smaller $\sigma$-field always wins". As ~~a corollary~~ special cases, we have:

(3) (Law of Total expectation) When $\mathcal{F}_1 = \{\phi, \Omega\}$, we have $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_2)) = \mathbb{E}X$.

[This also follows immediately from the definition of CE.]

(4) If $\mathbb{E}(X \mid \mathcal{F}_2) \in \mathcal{F}_1$, then $\mathbb{E}(X \mid \mathcal{F}_1) = \mathbb{E}(X \mid \mathcal{F}_2)$ a.s.

[ $\mathbb{E}(X \mid \mathcal{F}_1) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_2) \mid \mathcal{F}_1) = \mathbb{E}(X \mid \mathcal{F}_2)$ ]

Pf: (1) is trivial by definition since $\mathbb{E}(X \mid \mathcal{F}_1) \in \mathcal{F}_2$. For (2), we notice for any $A \in \mathcal{F}_1 \subset \mathcal{F}_2$, we have

$$\int_A \mathbb{E}(X \mid \mathcal{F}_1)\, d\mathbb{P} = \int_A X\, d\mathbb{P} = \int_A \mathbb{E}(X \mid \mathcal{F}_2)\, d\mathbb{P}.$$ □

Theorem (Fix what is known) $X \in \mathcal{F}$ and $\mathbb{E}|Y|, \mathbb{E}|XY| < +\infty$. Then,

$$\mathbb{E}(XY \mid \mathcal{F}) = X\,\mathbb{E}(Y \mid \mathcal{F}) \quad a.s.$$

Pf: Note $X\,\mathbb{E}(Y \mid \mathcal{F}) \in \mathcal{F}$. Then we adopt the standard four-step procedure.

① If $X = \mathbb{1}_B$ for some $B \in \mathcal{F}$, then $\forall A \in \mathcal{F}$,

$$~~\mathbb{E}(XY \mid \mathcal{F}) =~~ \mathbb{E}[X\,\mathbb{E}(Y \mid \mathcal{F})\mathbb{1}_A] = \mathbb{E}[\mathbb{E}(Y \mid \mathcal{F})\mathbb{1}_{A\cap B}] = \mathbb{E}(Y\mathbb{1}_{A\cap B})$$

$$= \mathbb{E}(XY\mathbb{1}_A) \Rightarrow \mathbb{E}(XY \mid \mathcal{F}) = X\,\mathbb{E}(Y \mid \mathcal{F}) \ a.s.$$

② By linearity, the conclusion holds for any simple $X$.

③ If $X \geq 0$, $Y \geq 0$, let $X_n$ be simple functions s.t. $X_n \uparrow X$. By MCT,

$$\mathbb{E}(XY \mid \mathcal{F}) = \lim_{n\to\infty} \mathbb{E}(X_n Y \mid \mathcal{F}) = \lim_{n\to\infty} X_n\,\mathbb{E}(Y \mid \mathcal{F}) = X\,\mathbb{E}(Y \mid \mathcal{F}).$$

④ For general $X, Y$, split them into + and − parts. □

Theorem (Independence) Suppose $\tilde{\mathcal{F}}$ is a $\sigma$-field independent of $\sigma(X, \mathcal{F})$, then

$$\mathbb{E}[X \mid \sigma(\mathcal{F}, \tilde{\mathcal{F}})] = \mathbb{E}(X \mid \mathcal{F}) \quad a.s.$$

In ~~particular,~~ particular, if $X$ is independent of $\tilde{\mathcal{F}}$, then $\mathbb{E}(X \mid \tilde{\mathcal{F}}) = \mathbb{E}(X)$.

[ In other words, an independent $\sigma$-field does not provide any information about $X$. ]

Proof: We need to show that $\forall\ C \in \sigma(\mathcal{F}, \tilde{\mathcal{F}})$,

$$(+)\quad \mathbb{E}(X\mathbb{1}_C) = \mathbb{E}(Y\mathbb{1}_C), \quad \text{for } Y := \mathbb{E}(X \mid \mathcal{F}).$$

We first show that (+) holds for $C = A \cap B$, $\forall A \in \mathcal{F}, B \in \tilde{\mathcal{F}}$.

$$\text{We have} \begin{cases} \mathbb{E}(X\mathbb{1}_C) = \mathbb{E}(X\mathbb{1}_A\mathbb{1}_B) = \mathbb{E}(X\mathbb{1}_A)\,\mathbb{P}(B) \\ \mathbb{E}(Y\mathbb{1}_C) = \mathbb{E}(Y\mathbb{1}_A)\,\mathbb{P}(B) = \mathbb{E}(X\mathbb{1}_A)\,\mathbb{P}(B) \end{cases}$$

Now, let $\mathcal{C}$ be the collection of sets such that (+) holds. Let $\mathcal{A}$ be the collection of all sets that can be written as $A \cap B$ for $A \in \mathcal{F}$ and $B \in \tilde{\mathcal{F}}$.

Then, obviously $\mathcal{A}$ is a $\pi$-system. We can check that $\mathcal{C}$ is a $\lambda$-system.

① $\Omega \in \mathcal{C}$. ② If $C_1, C_2 \in \mathcal{C}$, and $C_1 \subset C_2$, then $C_2 - C_1 \in \mathcal{C}$. ③ If $C_n \in \mathcal{C}$ and $C_n \uparrow C$, then $C \in \mathcal{C}$ (by MCT).

By the $\pi$-$\lambda$ theorem, we have $\sigma(\mathcal{A}) = \sigma(\mathcal{F}, \tilde{\mathcal{F}}) \subseteq \mathcal{G}$. □

$\boxed{\text{A Metatheorem: } \pi\text{-}\lambda \text{ Theorem}}$ $\mathcal{P}$ is a $\pi$-system if it is closed under intersection, i.e., $A, B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$. $\mathcal{L}$ is a $\lambda$-system if it satisfies: (i) $\Omega \in \mathcal{L}$. (ii) If $A, B \in \mathcal{L}$ and $A \subset B$, then $B - A \in \mathcal{L}$. (iii) If $A_n \in \mathcal{L}$ and $A_n \uparrow A$, then $A \in \mathcal{L}$.

Thm ($\pi$-$\lambda$ theorem): If $\mathcal{P}$ is a $\pi$-system and $\mathcal{L}$ is a $\lambda$-system that contains $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Theorem (Condition on an independent RV) $X, Y$ are independent. Let $\varphi$ be a ~~function~~ measurable function s.t. $\mathbb{E}|\varphi(X,Y)| < +\infty$ and let $g(x) := \mathbb{E}[\varphi(x, Y)]$. Then, we have

$$\mathbb{E}[\varphi(X,Y) \mid X] := \mathbb{E}[\varphi(X,Y) \mid \sigma(X)] = g(X) \text{ a.s.}$$

Pf: It is clear that $g(X) \in \sigma(X)$. $\forall A \in \sigma(X)$, we ~~have~~ need to show

$$\mathbb{E}[\varphi(X,Y)\mathbb{1}_A] = \mathbb{E}[g(X)\mathbb{1}_A]. \quad \text{Let } \mu = \mathcal{L}(X),\ \nu = \mathcal{L}(Y).$$

$A$ can be written as $A = \{X \in B\} = X^{-1}(B)$ for some Borel measurable $B$.

Then, $\mathbb{E}[\varphi(X,Y)\mathbb{1}_{X \in B}] = \iint \varphi(x,y)\mathbb{1}(x \in B)\,\mu(dx)\,\nu(dy)$ (by independence)

$$= \int \mathbb{1}(x \in B)\left(\int \varphi(x,y)\,\nu(dy)\right)\mu(dx) \quad \text{(Fubini)}$$

$$= \int \mathbb{1}(x \in B)\,g(x)\,\mu(dx) = \mathbb{E}[g(X)\mathbb{1}_{X \in B}]. \quad □$$

Theorem: Conditional expectation is a contraction in $L^p$, $p \geq 1$.

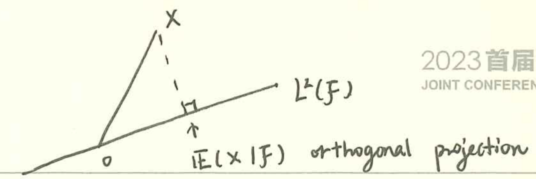
Pf: By Jensen's ineq., we get $|\mathbb{E}(X \mid \mathcal{F})|^p \leq \mathbb{E}[|X|^p \mid \mathcal{F}]$

$$\Rightarrow \mathbb{E}|\mathbb{E}(X \mid \mathcal{F})|^p \leq \mathbb{E}[|X|^p]. \quad □$$

Theorem: (Geometric meaning) Suppose $X \in L^2(\mathbb{P})$. $\mathbb{E}(X \mid \mathcal{F})$ is the RV $Y \in \mathcal{F}$ that minimizes the mean square error $\mathbb{E}(X-Y)^2$.



$L^2(\mathcal{F}) \subset L^2(\mathcal{F}_0)$




Pf: Let $Y = \mathbb{E}(X \mid \mathcal{F})$ and $Z$ be any RV s.t. $Z \in \mathcal{F}$. Then,

$$\mathbb{E}(X-Z)^2 = \mathbb{E}(X - Y + Y - Z)^2 = \mathbb{E}(X-Y)^2 + \mathbb{E}(Y-Z)^2 + 2\mathbb{E}[\overset{0}{(X-Y)}(Y-Z)]$$

$$\mathbb{E}[(X-Y)(Y-Z)] = \mathbb{E}[\mathbb{E}((X-Y)(Y-Z) \mid \mathcal{F})] = \mathbb{E}[(Y-Z)\,\mathbb{E}(X-Y \mid \mathcal{F})]$$

$$= \mathbb{E}[(Y-Z)(\mathbb{E}(X \mid \mathcal{F}) - Y)] = 0.$$

Thus, $\mathbb{E}(X-Z)^2$ is minimized when $Z = \mathbb{E}(X \mid Y)$ a.s. □

Example: $Y_1, Y_2, \cdots$ i.i.d. with mean $\mu$ and variance $\sigma^2$, $N$ an independent integer valued RV with $\mathbb{E}N^2 < +\infty$, $X = Y_1 + \cdots + Y_N$. Find $\mathbb{E}X$ and $Var(X)$.

Sol: $\mathbb{E}X = \mathbb{E}[\sum_{i=1}^N Y_i] = \mathbb{E}\{\mathbb{E}[\sum_{i=1}^N Y_i \mid N]\} = \mathbb{E}[\mu N] = \mu\mathbb{E}N$.

$$\mathbb{E}(X^2) = \mathbb{E}[\mathbb{E}(X^2 \mid N)] = \mathbb{E}\{\mathbb{E}[(\sum_{i=1}^N Y_i)^2 \mid N]\} = \mathbb{E}[\sigma^2 N + (\mu N)^2]$$

$$= \sigma^2\mathbb{E}N + \mu^2\mathbb{E}[N^2].$$

Then, $Var(X^2) = \sigma^2\mathbb{E}N + \mu^2\{\mathbb{E}[N^2] - (\mathbb{E}N)^2\} = \sigma^2\mathbb{E}N + \mu^2 Var(N)$. □

Example: $X_1, \cdots, X_n, \cdots$ i.i.d with ~~expectation~~ $\mathbb{E}|X_1| < +\infty$. Let $S_n = \sum_{i=1}^n X_i$.

Let $\mathcal{F}_n = \sigma(S_n, S_{n+1}, \cdots)$. Calculate $\mathbb{E}(X_1 \mid \mathcal{F}_n)$.

Sol: $\mathcal{F}_n = \sigma(S_n, X_{n+1}, \cdots)$, so by independence, $\mathbb{E}(X_1 \mid \mathcal{F}_n) = \mathbb{E}(X_1 \mid S_n)$.

We ~~argue~~ argue that $\mathbb{E}(X_i \mid S_n) = \mathbb{E}(X_1 \mid S_n)$ $\forall i = 1, 2, \cdots, n$.

Let $A \in \sigma(S_n)$. We want to show $\mathbb{E}(X_i\mathbb{1}_A) = \mathbb{E}(X_1\mathbb{1}_A)$.

$A$ can be written as $A = \{S_n \in B\}$ for some measurable $B$:

$\mathbb{E}(X_i\mathbb{1}_{\{S_n \in B\}}) = \mathbb{E}(X_1\mathbb{1}_{\{S_n \in B\}})$, which holds due to exchangability. Hence, we get

$$\mathbb{E}(X_1 \mid S_n) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i \mid S_n) = \frac{1}{n}\mathbb{E}(S_n \mid S_n) = \frac{S_n}{n}. \quad □$$

## Section 1.2 Martingales

$(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. A filtration $\{\mathcal{F}_n\}$ is an increasing sequence of sub-$\sigma$-fields: $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots$

A sequence of random variables $\{X_n\}$ is said to be adapted to $\{\mathcal{F}_n\}$ if $X_n \in \mathcal{F}_n$ for all $n$. Given a sequence $\{X_n\}$, there is a natural filtration associated with it: $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$.

**Definition** (Martingale) Let $\{X_n\}$ be a sequence s.t.

(i) $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$;

(ii) $\mathbb{E}|X_n| < +\infty \quad \forall n$;

(iii) $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n$ a.s. $\forall n$.

Then, we say $X = \{X_n\}$ is a martingale (w.r.t. $\{\mathcal{F}_n\}$). If we change (iii) to

(sup) $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n$ a.s. $\forall n$, then $X$ is called a supermartingale;

(sub) $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$ a.s. $\forall n$, then $X$ is called a submartingale.

[If we do not give $\{\mathcal{F}_n\}$ explicitly, it is usually the natural filtration.]

Proposition ~~Rmk~~: If (iii) holds, ~~by induction~~, it is equivalent to that

$$\mathbb{E}(X_n | \mathcal{F}_m) = X_m \quad \forall\ n > m \geq 0.$$

Similarly, (sup) $\Leftrightarrow$ $\mathbb{E}(X_n|\mathcal{F}_m) \leq X_m$ a.s.

(sub) $\Leftrightarrow$ $\mathbb{E}(X_n|\mathcal{F}_m) \geq X_m$ a.s.

Pf: For (sup), let $n = m+k$ with $k=2$. Then,

$$\mathbb{E}(X_n | \mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_n|\mathcal{F}_{n-1}) | \mathcal{F}_m) \leq \mathbb{E}(X_{n-1} | \mathcal{F}_m).$$

The desired result then follows by induction. □

**Examples**: (i) Let $\xi_1, \xi_2, \cdots$ be a sequence of i.i.d. random variables in $L^1$. Let $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$. Define $X_n = \sum_{i=1}^n \xi_i$.

$\mathcal{F}_0 = \{\phi, \Omega\}$.

(i) If $\mu = \mathbb{E}\xi_1 = 0$, then $X_n$ is a martingale w.r.t. $\{\mathcal{F}_n\}$.

$\mu \leq 0$ $(\geq 0)$, —— " —— sup (sub)-martingale.

Pf: $X_n \in L^1$ because $\mathbb{E}|X_n| \leq \sum_{i=1}^n \mathbb{E}|\xi_i| < +\infty$.

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_n + \xi_{n+1} | \mathcal{F}_n] = X_n + \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) \overset{\text{by independence}}{=} X_n + \overset{\mu}{\overset{\shortparallel}{\mathbb{E}\xi_n}}$$

$$\begin{cases} = X_n & \text{if } \mu = 0 \\ \leq X_n & \text{if } \mu \leq 0 \\ \geq X_n & \text{if } \mu \geq 0. \end{cases}$$ □

(ii) If $\mu = 0$ and $\sigma^2 = \mathrm{Var}(\xi_1) < +\infty$, then $X_n^2 - n\sigma^2$ is a martingale.

Pf: $\mathbb{E} X_n^2 = n\sigma^2 < +\infty$, so $X_n^2 - n\sigma^2 \in L^1$.

$$\mathbb{E}(X_{n+1}^2 | \mathcal{F}_n) = \mathbb{E}[(X_n + \xi_{n+1})^2 | \mathcal{F}_n] = \mathbb{E}[X_n^2 + 2X_n\xi_{n+1} + \xi_{n+1}^2 | \mathcal{F}_n]$$

$$= X_n^2 + 2X_n \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) + \mathbb{E}(\xi_{n+1}^2 | \mathcal{F}_n)$$

$$= X_n^2 + \sigma^2.$$

Hence, $\mathbb{E}(X_{n+1}^2 - (n+1)\sigma^2) = X_n^2 + \sigma^2 - (n+1)\sigma^2 = X_n^2 - n\sigma^2$. □

(iii) If $\mathbb{E}\xi_1 = 1$, then $Y_n = \prod_{i=1}^n \xi_i$ is a martingale.

Pf: $\mathbb{E}|Y_n| \leq \mathbb{E}\prod_{i=1}^n |\xi_i| = \prod_{i=1}^n \mathbb{E}|\xi_i| < +\infty \Rightarrow Y_n \in L^1$

$$\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = \mathbb{E}(Y_n \xi_{n+1} | \mathcal{F}_n) = Y_n \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = Y_n \mathbb{E}(\xi_{n+1}) = Y_n.$$ □

As a special case of (iii), let $X_i = e^{\theta\xi_i}$ with i.i.d. $\xi_1, \cdots, \xi_n, \cdots$ and $\phi(\theta) = \mathbb{E}e^{\theta\xi_i} < +\infty$. Then, $M_n = \prod_{i=1}^n X_i / \phi(\theta)^n = \exp(\theta S_n)/\phi(\theta)^n$ is a ~~mang~~ martingale with $S_n = \sum_{i=1}^n \xi_i$. In particular, if $\xi_i$ are i.i.d. standard normal random variables, we have $\phi(\theta) = e^{\frac{1}{2}\theta^2}$. Hence

$$M_n = \exp\left(\theta \sum_{i=1}^n \xi_i - \tfrac{1}{2} n\theta^2\right) \text{ is an important martingale.}$$

Rmk: There is nothing "super" about a supermartingale. Its name come from superharmonic functions: $f(x) \geq \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\,dy$.

Let $f$ be a superharmonic function and $X_n$ a $\mathbb{R}^d$-valued process s.t. $X_{n+1}$ is uniformly distributed on a ball $B(X_n, r_n)$ centered at $X_n$ with radius $r_n$. Then, $\mathbb{E} f(X_{n+1}) \le f(X_n)$.

Theorem: $X_n$ is a martingale w.r.t. $\{\mathcal{F}_n\}$ and $\varphi$ is a convex function with $\mathbb{E}|\varphi(X_n)| < +\infty$ $\forall n$. Then, $\varphi(X_n)$ is a submartingale. In particular, $\forall p \ge 1$, if $X_n \in L^p$ for all $n$, then $|X_n|^p$ is a submartingale.

Pf: By Jensen, $\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] \ge \varphi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) = \varphi(X_n)$. □

This result does not hold in general if $X_n$ is only a sub-mg. But, if we assume that $\varphi(X_n)$ is increasing additionally, then it still holds.

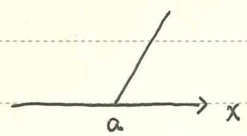
Theorem. If $X_n$ is a sub-mg and $\varphi$ is an increasing convex function, with $\mathbb{E}|\varphi(X_n)| < +\infty$ $\forall n$, then $\varphi(X_n)$ is a sub-mg. Consequently,

(i) $X_n$ is a sub-mg $\Rightarrow$ $(X_n - a)_+$ is a sub-mg.

(ii) $X_n$ is a super-mg $\Rightarrow$ $X_n \wedge a$ is a super-mg.

Pf: $\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] \ge \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) \ge \varphi(X_n)$ a.s.

(i) follows because $\varphi(x) = (x-a)_+ = \begin{cases} x-a & \text{if } x>a \\ 0 & \text{if } x\le 0 \end{cases}$




is increasing convex.

(ii) $X_n$ is a super-mg $\Leftrightarrow$ $-X_n$ is a sub-mg. $X_n \wedge a = -(-X_n)\vee(-a)$

Hence, only need to show that $(-X_n)\vee(-a)$ is a sub-mg.

Notice that $(-X_n)\vee(-a) = (-X_n+a)_+ - a$. □

$\boxed{\text{A Gambling System}}$ Given a filtration $\{\mathcal{F}_n\}$, $\{H_n\}$ is said to be a predictable system if $H_n \in \mathcal{F}_{n-1}$ $\forall n \ge 1$, i.e., $H_n$ can be predicted with certainty from the information at time $n-1$.

We now think of $H_n$ as the amount of money a gambler will bet at time $n$. This must be based on the outcomes at times (rounds) $1, 2, \cdots, n-1$.

Let $X_n$ be the ~~# of gaming chips~~ ~~the~~ net number of times of winning, i.e., $X_n - X_{n-1} = 1$ if wins for the $n$-th round, $=-1$ if loses.

If the Gambler bets according to a gambling system $H$, then the winning at time $n$ is $(H\cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1})$.

Casino always design unfavorable games for gamblers: $X_n$ is a super-mg.

I.e., if we bet 1¥ each time, our fortune always decrease

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) \le X_n.$$

But, there seems to be a gambling system that can ~~surely ensure~~ beat casino, called "martingale": $H_1 = 1$, and for $n \ge 2$,

$$H_n = \begin{cases} 2H_{n-1}, & \text{if } X_n - X_{n-1} = -1, \\ 1, & \text{if } X_n - X_{n-1} = 1. \end{cases}$$

I.e., we doubles our bet when we lose and then win, our net winning will be 1.

But, the next thm says that no ~~favorable game~~ gambling system can beat a ~~unfav~~ unfavorable game.

Theorem: Let $X_n$ be a super-mg. If $H_n \ge 0$ is predictable and each $H_n$ is bounded, then $(H\cdot X)_n$ is also a super-mg.

Proof: $(H\cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1}) \in \mathcal{F}_n$. $\mathbb{E}|(H\cdot X)_n| \le \sum_{m=1}^{n} \mathbb{E}|H_m(X_m - X_{m-1})|$

$$\le \sum_{m=1}^{n} \mathbb{E}|H_m|(|X_m| + |X_{m-1}|) < +\infty.$$

$$\mathbb{E}[(H\cdot X)_{n+1}|\mathcal{F}_n] = (H\cdot X)_n + \mathbb{E}(H_{n+1}(X_{n+1}-X_n)|\mathcal{F}_n) = (H\cdot X)_n + H_{n+1}\cdot\mathbb{E}(X_{n+1}-X_n|\mathcal{F}_n) \le (H\cdot X)_n.$$ □

Proposition: Same result also holds for submartingales and martingales. In particular, for martingales, we do not require that $H_n \ge 0$.

[You will encounter this again in Probability III, when ~~to~~ you learn stochastic integrations.]

$\boxed{\text{Stopping time}}$ An integer-valued RV $N$ is said to be a <u>stopping time</u> if $\{N=n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.

Intuitively, this says that the decision to stop at time $n$ must be measurable w.r.t. the information $\mathcal{F}_n$ known at that time.

Note that, if $N$ is a stopping time, then
- $\{N \le n\} = \bigcup_{k=1}^{n} \{N \le k\} \in \mathcal{F}_n$
- $\{N > n\} = \{N \le n\}^c \in \mathcal{F}_n$
- $\{N \ge n\} = \{N \le n-1\}^c \in \mathcal{F}_{n-1}$.

In particular, $H_n = \mathbb{1}_{\{N \ge n\}}$ is predictable, so $(H \cdot X)_n$ is a super/sub-martingale is $X$ is a ~~sub~~ super/sub-martingale. We can check by definition that

$$(H\cdot X)_n = \sum_{m=1}^{n} H_m (X_m - X_{m-1}) = \sum_{m=1}^{n} \mathbb{1}_{\{N \ge m\}} (X_m - X_{m-1}) = \sum_{m=1}^{N\wedge n} (X_m - X_{m-1})$$
$$= X_{N\wedge n} - X_0 .$$

Note the constant sequence $Y_n = X_0$ is a super/sub-martingale. Hence:

<u>Theorem</u>: If $N$ is a stopping time and $X_n$ is a supermartingale, then $X_{N\wedge n}$ is also a super-martingale.

<u>Rmk</u>: More on this when we talk about optional stopping theorem.

☆ $\boxed{\text{Martingale Convergence Theorem}}$ $\underline{X_n^+ = X_n \vee 0}$.

<u>Theorem</u>: If $X_n$ is a submartingale with $\underline{\sup \mathbb{E} X_n^+ < +\infty}$, then as $n \to +\infty$, $X_n$ converges a.s. to a limit $X$ with $\mathbb{E}|X| < +\infty$.

<u>Rmk</u>: Intuitively, submartingale is a process that can only "goes up". This theorem says that if it is "upper bounded", then the process finall converges a.s.

<u>Corollary</u>: If $X_n$ is a martingale with $\sup_n \mathbb{E}|X_n| < +\infty$, then ~~lim~~ as $n \to +\infty$, $X_n$ converges a.s. to a limit $X$ with ~~$\mathbb{E}X$~~ $\mathbb{E}|X| < +\infty$.

<u>Corollary</u>: If $X_n \ge 0$ is a supermartingale, then as $n \to +\infty$, $X_n \to X$ a.s. and $\mathbb{E}X \le \mathbb{E}X_0$.

<u>Pf</u>: $Y_n := -X_n$ is a submartingale with $\mathbb{E}Y_n^+ = 0$. $\overbrace{\phantom{xxxx}}^{\text{So } X_n \to X \text{ a.s.}}$ Note $\mathbb{E}X_n \le \mathbb{E}X_0$ $\forall n$. Applying Fatou's lemma, we get $\mathbb{E}X \le \mathbb{E}X_0$. □

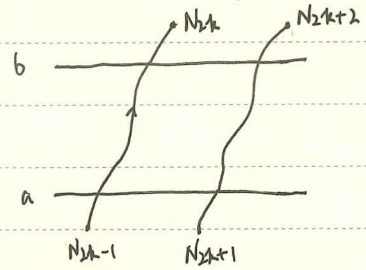
<u>Proof of the Theorem</u>: $X_n$ is a sub-martingale. Pick any $a<b$. Let $N_0 = -1$, and for $k \ge 1$, let

$$N_{2k-1} = \inf\{m \ge N_{2k-2} : X_m \le a\}, \quad N_{2k} = \inf\{m \ge N_{2k-1} : X_m \ge b\}.$$

As a convention, $\inf \phi = \infty$. Then, these $N_j$'s are stopping times and $\{N_{2k-1} < m \le N_{2k}\} = \{N_{2k-1} \le m-1\} \cap \{N_{2k} \le m-1\}^c \in \mathcal{F}_{m-1}$. Hence,

$$H_m = \begin{cases} 1, & \text{if } N_{2k-1} < m \le N_{2k} \text{ for some } k \\ 0, & \text{otherwise,} \end{cases}$$

is predictable. Note between $N_{2k-1}$ and $N_{2k}$, $X_m$ crosses from below $a$ to above $b$.




Hence, $H_m$ is a gambling system that takes advantage of these "upcrossings". In stock markets, we buy when $X_m \le a$ and sell when $X_m \ge b$, making a profit of $\ge (b-a)$ over every upcrossing.

Let $U_n := \sup\{k : N_{2k} \le n\}$ be the # of upcrossings by time $n$.

<u>Lemma</u> (Doob's upcrossing inequality) If $X$ is a sub-martingale, then

$$(b-a)\,\mathbb{E}U_n \le \mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+ .$$

<u>Pf</u>: Let $Y_n = a + (X_n - a)^+$. Then, $Y_n$ is a ~~martingale~~ sub-mg. It upcrosses $[a,b]$ the same # of times as $X_n$. We claim: for $N_{2k+1} \le n < N_{2k+2}$, $N_{2k} \le n < N_{2k+1}$,

$$(b-a)\,U_n \le (H\cdot Y)_n = \sum_{j=1}^{k+1} (H\cdot Y)_{N_{2k}} + \underline{(Y_n - a)} \to \underline{\text{why use } Y}$$
$$\text{or} = (H\cdot Y)_{N_{2k}} \quad \text{if } n = N_{2k}$$

Let $K_m = 1 - H_m$, it is also predictable. Then, we know $(K\cdot Y)_n$ is a sub-martingale. Hence, $\mathbb{E}(Y_n - Y_0) = \mathbb{E}(H\cdot Y)_n + \mathbb{E}(K\cdot Y)_n$

$$\geq \mathbb{E}(H\cdot Y)_n + \mathbb{E}(K\cdot Y)_0 = \mathbb{E}(H\cdot Y)_n$$
$$\geq (b-a)\mathbb{E}U_n. \qquad \square$$

With the up-crossing inequality, we obtain that ( $(X-a)^+ \leq X^+ + |a|$ )

$$\mathbb{E}U_n \leq \mathbb{E}(X_n-a)^+/(b-a) \leq \frac{|a| + \mathbb{E}X_n^+}{b-a} \quad \forall a<b \in \mathbb{R}.$$

By MCT, we have $\mathbb{E}U \leq \frac{\sup \mathbb{E}X_n^+ + |a|}{b-a} < +\infty$ by our assumption, where $U$ is the total # of upcrossings of $[a,b]$ by the whole sequence. Thus, we have $U<\infty$ a.s., ~~Taking union on which~~ so

$$\mathbb{P}(\liminf X_n < a < b < \limsup X_n) = 0.$$

Taking union over all $a<b \in \mathbb{Q}$, we get

$$\mathbb{P}(\liminf X_n < a < b < \limsup X_n \text{ for some } a<b\in\mathbb{Q}) = 0.$$

This implies $\liminf X_n = \limsup X_n$ a.s., i.e., $\lim_n X_n$ exists a.s. (Note this limit can be in $\{\pm\infty\}$.)

By Fatou's lemma, we have

$$\mathbb{E}X^+ \leq \liminf \mathbb{E}X_n^+ < +\infty, \text{ so } X < +\infty \text{ a.s.}$$

On the other hand, $\mathbb{E}X_n^- = \mathbb{E}X_n^+ - \mathbb{E}X_n \leq \mathbb{E}X_n^+ - \mathbb{E}X_0$, so applying Fatou gives

$$\mathbb{E}X^- \leq \liminf \mathbb{E}X_n^- \leq \sup \mathbb{E}X_n^+ - \mathbb{E}X_0 < +\infty.$$

Thus, we have $\mathbb{E}|X| = \mathbb{E}X^+ + \mathbb{E}X^- < +\infty$. $\square$

Example: Let $\{\xi_i\}$ be a sequence of non-neg. independent random variables in $L^1$ with $\mathbb{E}\xi_i = 1$. Then, $X_n = \prod_{i=1}^n \xi_i$ is a non-negative martingale and $X_n$ converges a.s. to some limit $X_\infty \in L^1$.

Note: even if we have almost sure convergence of a martingale $X_n$ to some $X$, we do not necessarily convergence to $L^1$. In particular, we ~~may~~ can have

$$\mathbb{E}X \neq \lim_{n\to\infty}\mathbb{E}X_n = \mathbb{E}X_0.$$

Example: $S_n$ is a symmetric simple random walk with $S_0 = 1$, $S_n = S_{n-1} + \xi_n$ where $\xi_1, \dots, \xi_n, \dots$ are i.i.d. with $\mathbb{P}(\xi_i = \pm 1) = 1/2$. Let $N = \inf\{n : S_n = 0\}$. Define $X_n = S_{N\wedge n}$. Since $N$ is a stopping time, we know $X_n$ is a martingale, and it is non-negative. Then, the martingale convergence theorem tells that $X_n$ converges to a limit $X_\infty < +\infty$. This limit must be zero, because convergence to other values are impossible. But, we know $\mathbb{E}X_n = \mathbb{E}X_0 = 1 \neq \mathbb{E}X_\infty$, so convergence does not occur in $L^1$.

☆ $\boxed{\text{Martingale convergence in } L^p,\ p>1}$ we look at the slightly stronger $L^p$ convergence, which implies the $L^1$-convergence.

Lemma: If $X_n$ is a sub-mg and $N$ is a stopping time with $\mathbb{P}(N\leq k) = 1$, then

$$\mathbb{E}X_0 \leq \mathbb{E}X_N \leq \mathbb{E}X_k.$$

Rmk: In the above example, we have $\mathbb{E}X_0 = 1 > 0 = \mathbb{E}X_N$. The reason is that $N$ is an unbounded stopping time.

Pf: We know that $X_{N\wedge n}$ is a sub-mg, so

$$\mathbb{E}X_0 = \mathbb{E}X_{N\wedge 0} \leq \mathbb{E}X_{N\wedge k} = \mathbb{E}X_N. \qquad \square$$

Theorem (Doob's maximal inequality) Let $X_n$ be a sub-mg. Define

$$\bar{X}_n = \max_{0\leq m\leq n} X_m^+.$$

Then, $\forall \lambda > 0$,

$$\lambda\mathbb{P}(\bar{X}_n \geq \lambda) \leq \mathbb{E}[X_n \mathbb{1}_{\{\bar{X}_n\geq\lambda\}}] \leq \mathbb{E}[X_n^+ \mathbb{1}_{\{\bar{X}_n\geq\lambda\}}] \leq \mathbb{E}X_n^+.$$

Proof: Define $N = \inf\{m \ge 0: X_m \ge \lambda \text{ or } m = n\}$. $N$ is a stopping time.

Note $X_N \ge \lambda$ on $\{\bar{X}_n \ge \lambda\}$, so

$$\lambda \mathbb{P}(\bar{X}_n \ge \lambda) \le \mathbb{E} X_N \mathbb{1}\{\bar{X}_n \ge \lambda\} \le \mathbb{E} X_n \mathbb{1}\{\bar{X}_n \ge \lambda\} \text{ by previous lemma.}$$

[ $\mathbb{E}X_N \le \mathbb{E}X_n$ and $X_N = X_n$ on $A^c$. ] □

Rmk: In essence, this is ~~similar to~~ an extension of Kolmogorov's maximal inequality.

$S_n = \sum_{i=1}^{n} \xi_i$, where $\xi_1, \xi_2, \cdots$ are independent with $\mathbb{E}\xi_i = 0$ and $\sigma_i^2 = \mathbb{E}X_i^2 < +\infty$.

Then, $S_n$ is a martingale and $X_n = S_n^2$ is a sub-mg. Let $\lambda = x^2$ and applying Doob's inequality, we get

$$\mathbb{P}\left(\max_{1\le m\le n} |S_m| \ge x\right) \le x^{-2}\, \mathbb{E}(S_n^2) = x^{-2}\, \mathrm{Var}(S_n).$$

Recall the formula: If $Y \ge 0$ and $p > 0$, then $\mathbb{E}(Y^p) = \int_0^\infty p y^{p-1} \mathbb{P}(Y \ge y)\, dy$.

Combining it with Doob's inequality, we obtain that:

Theorem ($L^p$ maximal inequality) If $X_n$ is a sub-mg, then for $1 < p < +\infty$,

$$\mathbb{E}(\bar{X}_n^p) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}(X_n^+)^p.$$

Consequently, if $Y_n$ is a martingale and $Y_n^* = \max_{0 \le m \le n} |Y_m|$, then

$$\mathbb{E}|Y_n^*|^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|Y_n|^p, \text{ i.e., } \|Y_n^*\|_p \le \frac{p}{p-1}\|Y_n\|_p.$$

Proof: $\mathbb{E}(\bar{X}_n^p) = \int_0^\infty p y^{p-1} \mathbb{P}(\bar{X}_n \ge y) = \int_0^\infty p y^{p-2} \cdot y \mathbb{P}(\bar{X}_n \ge y)\, dy$

Doob's maximal inequality $\Leftarrow$ $\le \int_0^\infty p y^{p-2} \mathbb{E}(X_n^+ \mathbb{1}\{\bar{X}_n \ge y\})\, dy \underset{\text{Fubini}}{=} \mathbb{E}\left[\int_0^\infty p y^{p-2} X_n^+ \mathbb{1}\{\bar{X}_n \ge y\}\, dy\right]$

$= \mathbb{E}\left[X_n^+ \int_0^{\bar{X}_n} p y^{p-2}\, dy\right] = \mathbb{E}[X_n^+ (\bar{X}_n^{p-1})] \cdot \frac{p}{p-1}$

$\le \|X_n^+\|_p (\mathbb{E}\bar{X}_n^p)^{\frac{p-1}{p}}$ $\boxed{\text{Hölder}}$ $\overset{(\star)}{\Rightarrow} (\mathbb{E}\bar{X}_n^p)^{\frac{1}{p}} \le \frac{p}{p-1}\|X_n^+\|_p$

$\Rightarrow \|\bar{X}_n\|_p \le \frac{p}{p-1}\|X_n^+\|_p$.

Here, in step ($\star$), we need that $\mathbb{E}\bar{X}_n^p < +\infty$. We can take any $M > 0$, and repeat the same calculation to get

$$\mathbb{E}(\bar{X}_n \wedge M)^p \le \|X_n^+\|_p \left[\mathbb{E}(\bar{X}_n \wedge M)^p\right]^{\frac{p-1}{p}} \Rightarrow \|\bar{X}_n \wedge M\|_p \le \frac{p}{p-1}\|X_n^+\|_p.$$

Then, take $M \to \infty$ and use the monotone convergence theorem. □

Example: ($L^1$ maximal inequality does not hold) Look at the simple RW example again.

$X_n = S_{N \wedge n}$. We will see that $\mathbb{P}(\max_m X_m \ge M) = \frac{1}{M}$. Hence,

$\mathbb{E}(\max_m X_m) = \sum_{M=1}^{\infty} 1/M = \infty$. Thus, by MCT, we have

$\mathbb{E}\max_{m \le n} X_m \to \infty$ as $n \to +\infty$, and $L^1$ maximal inequality cannot hold.

☆ Theorem: (Martingale convergence in $L^p$) $p > 1$, $X_n$ is a martingale. Then, the followings are equivalent.

(i) $\{X_n\}$ is $L^p$-bounded, i.e., $\sup \mathbb{E}(|X_n|^p) < +\infty$.

(ii) $X_n$ converges a.s. and in $L^p$ to some RV $X_\infty \in L^p$.

(iii) There exists a RV $Z \in L^p$ s.t. $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ a.s. $\forall n$.

Proof: (i) $\Rightarrow$ (ii). $\mathbb{E}X_n^+ \le \mathbb{E}|X_n| \le (\mathbb{E}|X_n|^p)^{\frac{1}{p}}$. Hence, by martingale convergence theorem, $X_n \to X_\infty$ a.s. for some RV $X_\infty$. Then, by $L^p$ maximal inequality,

$$\mathbb{E}\left(\sup_{0 \le m \le n} |X_m|^p\right) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|X_n|^p$$

Taking $n \to \infty$, by MCT, we have $\mathbb{E}(\sup_m |X_m|^p) \le \sup_n \mathbb{E}|X_n|^p < +\infty$.

Hence, $\sup_n |X_n| \in L^p$. Note that $|X_n - X_\infty|^p \le (2\sup_n |X_n|)^p$. Then, by DCT, we have $\mathbb{E}|X_n - X_\infty|^p \to 0$, i.e., $X_n \to X_\infty$ in $L^p$.

(ii) $\Rightarrow$ (iii). Take $Z = X_\infty$. Then, for ~~$m \le n$~~ $n \le m$,

$$|X_n - \mathbb{E}[Z \mid \mathcal{F}_n]| = |\mathbb{E}[X_m - Z \mid \mathcal{F}_n]| \le \mathbb{E}[|X_m - Z| \mid \mathcal{F}_n]$$

$\Rightarrow \mathbb{E}|X_n - \mathbb{E}[Z \mid \mathcal{F}_n]| \le \mathbb{E}|X_m - Z| \to 0$ as $m \to \infty$. Hence, $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ a.s.

(iii) ⇒ (i). By Jensen's ineq., $|X_n|^p \le \mathbb{E}[|Z|^p | \mathcal{F}_n]$. □

Corollary: Let $Z \in L^p$ for some $p>1$. Then, $\mathbb{E}[Z|\mathcal{F}_n] \to \mathbb{E}[Z|\mathcal{F}_\infty]$ a.s. and in $L^p$.

Pf: Denote $X_n = \mathbb{E}[Z|\mathcal{F}_n]$. Then, $X_n$ is a martingale, and the previous theorem implies that $X_n$ converges a.s. and in $L^p$ to some variable $Y \in L^p$. Moreover, $X_n = \mathbb{E}[Y|\mathcal{F}_n]$. Hence, $Y$ is a RV measurable w.r.t. to $\mathcal{F}_\infty$ and satisfying

$$\mathbb{E}[Y|\mathcal{F}_n] = \mathbb{E}[Z|\mathcal{F}_n] \quad \forall n \ge 0.$$

It remains to show $Y = \mathbb{E}[Z|\mathcal{F}_\infty]$. Set $\mathcal{L} := \{A \in \mathcal{F} : \mathbb{E}[Y \mathbb{1}_A] = \mathbb{E}[Z \mathbb{1}_A]\}$.
We know $\mathcal{F}_n \subset \mathcal{L}$ for all $n$. Moreover, we can show that $\mathcal{L}$ is a $\lambda$-system.
By $\pi$-$\lambda$ theorem, we know $\mathcal{F}_\infty \subseteq \mathcal{L}$. □

**Martingale convergence in $L^2$**

Theorem (Doob's decomposition) Any sub-mg $X_n$, $n \ge 0$, can be written ~~as~~ in a unique way as $X_n = M_n + A_n$, where $M_n$ is a martingale and $A_n$ is a predictable increasing sequence with $A_0 = 0$.

Proof: We want $X_n = M_n + A_n$, $\mathbb{E}[M_n|\mathcal{F}_{n-1}] = M_{n-1}$ and $A_n \in \mathcal{F}_{n-1}$. So,

$$\mathbb{E}[X_n|\mathcal{F}_{n-1}] = \mathbb{E}[M_n|\mathcal{F}_{n-1}] + \mathbb{E}[A_n|\mathcal{F}_{n-1}]$$
$$= M_{n-1} + A_n = M_{n-1} + A_{n-1} + (A_n - A_{n-1}) = X_{n-1} + (A_n - A_{n-1})$$

$\Rightarrow$ $A_n - A_{n-1} = \mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}$. Since $A_0 = 0$, we must have

$$A_n = \sum_{m=1}^{n} (\mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}), \text{ and let } M_n = X_n - A_n.$$

We now check that this choice works. First, $A_n - A_{n-1} \ge 0$ since $X_n$ is sub-mg, and $A_n \in \mathcal{F}_{n-1}$. Now, we show that $M_n$ is a martingale.

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] = \mathbb{E}[X_n - A_n|\mathcal{F}_{n-1}] = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - A_n$$
$$= X_{n-1} + (A_n - A_{n-1}) - A_n = X_{n-1} - A_{n-1} = M_{n-1}. \quad \square$$

Let $X_n$ be a martingale in $L^2$ with $\mathbb{E}X_0 = 0$. Then, $X_n^2$ is a positive sub-mg.
By Doob's decomposition, $X_n^2 = \underbrace{M_n}_{\text{martingale part}} + A_n$ → a predictable increasing sequence

$$A_n = \sum_{m=1}^{n} [\mathbb{E}(X_m^2 | \mathcal{F}_{n-1}) - X_{m-1}^2] = \sum_{m=1}^{\infty} \mathbb{E}[(X_m - X_{m-1})^2 | \mathcal{F}_{m-1}].$$

↓ Quadratic variation: a pathwise measurement of the variance at time $n$.

$$\mathbb{E}|X_n|^2 = \mathbb{E}\Big[\sum_{m=1}^{n} (X_m - X_{m-1})\Big]^2 = \cancel{\mathbb{E}} \sum_{m=1}^{n} \mathbb{E}[(X_m - X_{m-1})^2] = \mathbb{E}A_n.$$
$$= \mathbb{E}(M_n + A_n) = \mathbb{E}M_0 + \mathbb{E}A_n = \mathbb{E}A_n.$$

Corollary: $X_n$ converges a.s. and in $L^2$ if $\mathbb{E}A_\infty < +\infty$.

Optional

But, we can extend this result a little bit.

Theorem: $\lim_{n\to\infty} X_n$ exists and is finite a.s. on $\{A_\infty < \infty\}$.

Pf: By the $L^2$-maximal ineq., we have

$$\mathbb{E} \sup_{0 \le m \le n} |X_m|^2 \le 4\,\mathbb{E}|X_n|^2 = 4\,\mathbb{E}A_n. \quad (*)$$

~~Taking $n\to\infty$, by MCT, we have $\mathbb{E}\sup_n |X_n|^2 \le 4\mathbb{E}$~~

Let $a > 0$ and $N = \inf\{n : A_{n+1} > a^2\}$. $N$ is a stopping time since $A_{n+1} \in \mathcal{F}_n$.
By (*) for $X_{N\wedge n}$, we have

$$\mathbb{E} \sup_{0 \le m \le n} |X_{N\wedge m}|^2 \le 4\,\mathbb{E}A_{N\wedge n} \le 4a^2 \xRightarrow{n\to\infty} \mathbb{E}\sup_n |X_{N\wedge n}|^2 \le 4a^2.$$

By the $L^p$-convergence theorem, $\lim X_{N\wedge n}$ exists and is finite a.s. Take $a \to \infty$, the results follows. □

☆ **Martingale Convergence in $L^1$; uniform integrability**

The key to the convergence in $L^1$ is the following concept.

**Def:** A collection of random variables $X_i$, $i\in I$, is said to be <u>uniformly integrable</u> if $\lim_{M\to+\infty} \sup_{i\in I} \mathbb{E}(|X_i|;\ |X_i|>M)=0$.

Note that $\{X_i: i\in I\}$ is uniformly integrable implies that it is uniformly $L^1$-bounded.

In fact, there exists $M>0$ s.t. $\sup_{i\in I} \mathbb{E}(|X_i|;\ |X_i|>M)\le 1 \Rightarrow \sup_{i\in I}\mathbb{E}|X_i|\le M+1<+\infty$.

If $\{X_i\}$ are bounded by an integrable RV $Y\ge 0$ with $\mathbb{E}Y<+\infty$ (i.e. $|X_i|\le Y$), then $\{X_i\}$ is uniformly integrable since

$$\sup_{i\in I}\mathbb{E}(|X_i|;\ |X_i|\ge M)\le \mathbb{E}(Y;\ Y\ge M)\to 0 \text{ as } M\to\infty \text{ by DCT.}$$

The following is an extension of this result.

**Lemma:** $(\Omega,\mathcal{F}_0,\mathbb{P})$. If $X\in L^1$, then the family $\{\mathbb{E}(X|\mathcal{F}): \mathcal{F} \text{ is a sub-}\sigma\text{-field of } \mathcal{F}_0\}$ is UI.

**Proof:** First, we claim that $\forall\varepsilon>0$, there exists $\delta>0$ s.t. $\mathbb{P}(A)\le\delta$ implies that $\mathbb{E}(|X|;A)\le\varepsilon$. Suppose otherwise, there exists $A_n$ with $\mathbb{P}(A_n)\le 1/n$ such that $\mathbb{E}(|X|;A_n)>\varepsilon$. But we know $\mathbb{E}(|X|;A_n)\to 0$ by DCT, a contradiction.

To show the conclusion, we pick $M=\mathbb{E}|X|/\delta$. Then, since $\{|\mathbb{E}(X|\mathcal{F})|>M\}$ is $\mathcal{F}$-meas., by Jensen,

$$\mathbb{E}[|\mathbb{E}(X|\mathcal{F})|\mathbb{1}_{|\mathbb{E}(X|\mathcal{F})|>M}]$$
$$\le \mathbb{E}[\mathbb{E}(|X||\mathcal{F})\mathbb{1}_{|\mathbb{E}(X|\mathcal{F})|>M}] = \mathbb{E}[|X|\mathbb{1}_{\{|\mathbb{E}(X|\mathcal{F})|>M\}}].$$

By Markov's inequality, $\mathbb{P}(|\mathbb{E}(X|\mathcal{F})|>M)\le \mathbb{E}|\mathbb{E}(X|\mathcal{F})|/M\le \mathbb{E}|X|/M=\delta$.

Then, by the choice of $\delta$, we have $\mathbb{E}(|X|;\ |\mathbb{E}(X|\mathcal{F})|>M)\le\varepsilon\ \forall\mathcal{F}$. □

A common way to check UI is to use the following lemma.

**Lemma:** Let $\varphi\ge 0$ be any function with $\varphi(x)/x\to+\infty$ as $x\to+\infty$. (e.g., $\varphi(x)=x^p$ or $\varphi(x)=x\log^+x$)

If $\sup_{i\in I}\mathbb{E}\varphi(|X_i|)<+\infty$, then $\{X_i: i\in I\}$ is UI.

**Proof:** Let $\varepsilon_M=\sup\{\frac{x}{\varphi(x)}: x\ge M\}$. Then, $\varepsilon_M\to 0$ as $M\to+\infty$.

Moreover, for $i\in I$, we have

$$\mathbb{E}(|X_i|;\ |X_i|>M)=\mathbb{E}\Big(\frac{|X_i|}{\varphi(|X_i|)}\varphi(|X_i|);\ |X_i|>M\Big)\le(\mathbb{E}\varphi(|X_i|))\cdot\varepsilon_M. \quad □$$

The relevance of UI to convergence in $L^1$ is due to the following theorem.

**Theorem:** Suppose $\mathbb{E}|X_n|<+\infty\ \forall n$. If $X_n\to X$ in probability, then TFAE:

(i) $\{X_n: n\ge 0\}$ is UI.

(ii) $X_n\to X$ in $L^1$.

(iii) $\mathbb{E}|X_n|\to\mathbb{E}|X|<+\infty$.

**Proof:** (i)$\Rightarrow$(ii) Define $\varphi_M(x):=\begin{cases} M, & \text{if } x>M\\ x, & \text{if } |x|\le M\\ -M, & \text{if } x<-M\end{cases}$

Then, $|X_n-X|\le|X_n-\varphi_M(X_n)|+|\varphi_M(X_n)-\varphi_M(X)|+|\varphi_M(X)-X|$

Note $|\varphi_M(x)-x|=(|x|-M)^+\le|x|\mathbb{1}_{|x|>M}$. Hence, taking expectation gives

$$\mathbb{E}|X_n-X|\le\mathbb{E}|\varphi_M(X_n)-\varphi_M(X)|+\mathbb{E}(|X_n|;\ |X_n|>M)+\mathbb{E}(|X|;\ |X|>M)$$

Since $X_n\to X$ in probability, we have $|\varphi_M(X_n)-\varphi_M(X)|\le|X_n-X|\to 0$ in probability.

Then, we get $\mathbb{E}|\varphi_M(X_n)-\varphi_M(X)|\le\varepsilon+\mathbb{E}(|\varphi_M(X_n)-\varphi_M(X)|;\ |X_n-X|>\varepsilon)$

$$\le\varepsilon+2M\,\mathbb{P}(|X_n-X|>\varepsilon)\to\varepsilon \text{ as } n\to+\infty.$$

Since $\varepsilon$ is arbitrary, we get $\mathbb{E}|\varphi_M(X_n)-\varphi_M(X)|\to 0$ as $n\to+\infty$. Thus, we have

$$\limsup_{n\to\infty}\mathbb{E}|X_n-X|\le\sup_n\mathbb{E}(|X_n|;\ |X_n|>M)+\mathbb{E}(|X|;\ |X|>M)\quad\forall M>0.$$

By UI, we have $\sup_n\mathbb{E}(|X_n|;\ |X_n|>M)\to 0$ as $M\to+\infty$. For the last term, we have shown that $\mathbb{E}|\varphi_M(X)|=\lim_n\mathbb{E}|\varphi_M(X_n)|<\sup_n\mathbb{E}|X_n|<+\infty$. Moreover, by Fatou's lemma,

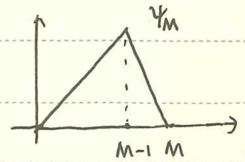
$$\mathbb{E}|X|\le\lim_{M\to\infty}\mathbb{E}|\varphi_M(X_n)|\le\sup_n\mathbb{E}|X_n|<+\infty.$$

Hence, $X\in L^1$ and we have $\mathbb{E}(|X|;|X|>M)\to 0$ as $M\to+\infty$.

In sum, we get that ~~$\lim \mathbb{E}|X_n-X|\le 2\varepsilon$~~ $\limsup_n \mathbb{E}|X_n-X|\le 2\varepsilon$ by choosing $M$ sufficiently large. Since $\varepsilon$ is arbitrary, we conclude (ii).

(ii) ⇒ (iii) By Jensen, $\big|\mathbb{E}|X_n| - \mathbb{E}|X|\big| \le \mathbb{E}\big||X_n|-|X|\big| \le \mathbb{E}|X_n-X| \to 0$.

(iii) ⇒ (i) Define $\psi_M(x) = \begin{cases} x, & \text{on } [0, M-1] \\ 0, & \text{on } [M,+\infty) \\ \text{linear}, & \text{on } [M-1, M] \end{cases}$




$\forall \varepsilon>0$, by DCT, we have that $|\mathbb{E}|X| - \mathbb{E}\psi_M(|X|)| \le \varepsilon/2$ for $M$ large enough.

As in part 1 of the proof, we have that $\mathbb{E}\psi_M(|X_n|) \to$ ~~$\mathbb{E}|X|$~~ $\mathbb{E}\psi(|X|)$ (by using that $X_n\to X$ in probability). On the other hand, by (iii), we have $\mathbb{E}|X_n|\to\mathbb{E}|X|$. ~~Hence~~ Choose $n_0$ large enough, s.t. $\mathbb{E}|X_n| \le \mathbb{E}|X| + \varepsilon/2$ $\forall n\ge n_0$.

So, $\forall n\ge n_0$,
$$\mathbb{E}(|X_n|;|X_n|>M) \le \mathbb{E}|X_n| - \mathbb{E}\psi_M(|X_n|) \le \mathbb{E}|X| + \varepsilon/2 - (\mathbb{E}|X| - \varepsilon/2) = \varepsilon.$$

Next, choosing $M$ sufficiently large, we can make $\mathbb{E}(|X_n|;|X_n|>M)\le\varepsilon$ for $0\le n<n_0$. □

Now, we are ready to state our main convergence theorem in $L^1$.

☆ Theorem (Martingale convergence in $L^1$) For a sub-martingale, TFAE.

(i) It is UI.

(ii) It converges a.s. and in $L^1$.

(iii) It converges in $L^1$.

Pf: (i) ⇒ (ii) UI implies $\sup_n \mathbb{E}|X_n| < +\infty$. So martingale convergence theorem implies $X_n\to X$ a.s. for some RV $X$. Then, the previous theorem gives $X_n\to X$ in $L^1$. (ii) ⇒ (iii): Trivial. (iii) ⇒ (i): Previous theorem. □



We have a stronger version for "martingales".

Lemma: If $X_n\to X$ in $L^1$, then $\mathbb{E}(X_n;A)\to\mathbb{E}(X;A)$ $\forall A\in\mathcal{F}_0$.

Proof: $|\mathbb{E}(X_n;A) - \mathbb{E}(X;A)| = |\mathbb{E}((X_n-X)\mathbb{1}_A)| \le \mathbb{E}|X_n-X|\to 0$. □

Lemma: If a martingale $X_n\to X$ in $L^1$, then $X_n = \mathbb{E}(X|\mathcal{F}_n)$.

Pf: $\forall m>n$, $\mathbb{E}(X_m|\mathcal{F}_n) = X_n$ a.s. So $\forall A\in\mathcal{F}_n$, $\mathbb{E}(X_n;A) = \mathbb{E}(X_m;A)$

Hence, we have $\mathbb{E}(X_n;A) = \mathbb{E}(X;A)$ $\forall A\in\mathcal{F}_n \Rightarrow X_n = \mathbb{E}(X|\mathcal{F}_n)$. $\longrightarrow \mathbb{E}(X;A)$.

☆ Theorem: (Martingale convergence in $L^1$) For a martingale, TFAE.

(i) It is UI.

(ii) It converges a.s. and in $L^1$.

(iii) It converges in $L^1$.

(iv) There exists an integrable $X^{\in L^1}$ s.t. $X_n = \mathbb{E}(X|\mathcal{F}_n)$.

Pf: (i), (ii), (iii) are equivalent since martingale is also sub-mg.

(iii) ⇒ (iv) by previous lemma. (iv) ⇒ (i) Have shown in previous lemma. □

We also have a reverse result.

Corollary: Let $Z\in L^1$. Then, $\mathbb{E}[Z|\mathcal{F}_n]\to\mathbb{E}[Z|\mathcal{F}_\infty]$ a.s. and in $L^1$.

Proof: $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ is a martingale. We know $X_n$ is UI, so $Y_n$ converges a.s. and in $L^1$ to a limit $X_\infty\in L^1$. Then, use the $\pi$-$\lambda$ theorem. □

Corollary: (Lévy's 0-1 law) If $\mathcal{F}_n\uparrow\mathcal{F}_\infty$ and $A\in\mathcal{F}_\infty$, then $\mathbb{E}(\mathbb{1}_A|\mathcal{F}_n)\to\mathbb{1}_A$ a.s.

Proof: This is an immediate consequence of the previous corollary. □

We can recover Kolmogorov's 0-1 law from Levy's 0-1 law.

Let $X_1, X_2, \cdots$ be independent, $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$, $\mathcal{T} = \bigcap_n \sigma(X_n, X_{n+1}, \cdots)$ the tail σ-algebra. For each $n$, $A$ is ind. of $\mathcal{F}_n$, so $\mathbb{E}(\mathbb{1}_A | \mathcal{F}_n) = \mathbb{P}(A)$. As $n \to +\infty$, we have $\mathbb{P}(A) = \mathbb{1}_A$ a.s., so $\mathbb{P}(A) \in \{0,1\}$.

**Theorem** (DCT for conditional expectations)

Suppose $Y_n \to Y$ a.s. and $|Y_n| \le Z$ for some $Z \in L^1$. If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, then

$$\mathbb{E}(Y_n | \mathcal{F}_n) \to \mathbb{E}(Y | \mathcal{F}_\infty) \text{ a.s.}$$

**Proof:** Let $W_N := \sup\{ |Y_n - Y_m| : n, m \ge N\}$. We know $W_N \le 2Z$, so $\mathbb{E}W_N < +\infty$. Applying the previous theorem to $W_N$ gives

$$\limsup_{n\to+\infty} \mathbb{E}\big(|Y_n - Y| \,\big|\, \mathcal{F}_n\big) \le \limsup_{n\to\infty} \mathbb{E}(W_N | \mathcal{F}_n) = \mathbb{E}(W_N | \mathcal{F}_\infty) \quad \forall N.$$

Taking $N \to +\infty$, by DCT, we have $\mathbb{E}(W_N | \mathcal{F}_\infty) \downarrow 0$ as $N \to +\infty$. Hence ~~Moreover~~,

$$|\mathbb{E}(Y_n|\mathcal{F}_n) - \mathbb{E}(Y|\mathcal{F}_n)| \le \mathbb{E}\big[|Y_n - Y| \,\big|\, \mathcal{F}_n\big] \to 0 \text{ as } n \to +\infty$$

Finally, by previous thm, $\mathbb{E}(Y|\mathcal{F}_n) \to \mathbb{E}(Y|\mathcal{F}_\infty)$ a.s. □

**Example** Suppose $X_1, X_2, \cdots$ are UI and $\to X$ a.s. Then, $X_n \to X$ in $L^1$ and $\mathbb{E}(X_n|\mathcal{F}) \to \mathbb{E}(X|\mathcal{F})$ in $L^1$. But, $\mathbb{E}(X_n|\mathcal{F})$ may not converge a.s. Let $Y_1, Y_2, \cdots$ and $Z_1, Z_2, \cdots$ be independent RV's with

$\mathbb{P}(Y_n = 1) = 1/n$, $\mathbb{P}(Y_n = 0) = 1 - 1/n$,

$\mathbb{P}(Z_n = n) = 1/n$, $\mathbb{P}(Z_n = 0) = 1 - 1/n$.

Let $X_n = Y_n Z_n$. Then, $\mathbb{P}(X_n > 0) = 1/n^2$, so by Borel-Cantelli, $X_n \to 0$ a.s. $\mathbb{E}(|X_n|; |X_n| \ge 1) = n/n^2 = 1/n$, so $X_n$ is UI. Let $\mathcal{F} = \sigma(Y_1, Y_2, \cdots)$. Then, $\mathbb{E}(X_n|\mathcal{F}) = Y_n \mathbb{E}(Z_n|\mathcal{F}) = Y_n \mathbb{E}Z_n = Y_n$. $Y_n \to 0$ in $L^1$ but not a.s. by B-C II.

## Section 1.3 Optional Stopping Theorems; Backwards Martingales

~~For a martingale,~~ If $X_n$ is a ~~martingale~~ sub-mg, then $\mathbb{E}X_0 \le \mathbb{E}X_n$ $\forall n \in \mathbb{N}$. But this may not hold if we replace $n$ by a stopping time $N$. In a previous lemma, we see that it holds if $N$ is bounded: $\mathbb{P}(N \le k) = 1$.

~~Hence, we will focus on the case of unbounded N.~~

**Def:** Let $\{\mathcal{F}_n\}$ be a filtration and $T$ be a stopping time. Define

$$\mathcal{F}_T := \{ A \in \mathcal{F} : A \cap \{T \le n\} \in \mathcal{F}_n \ \forall n \}.$$

We can check that $\mathcal{F}_T$ is a σ-algebra, which gives that information available by time $T$.

**Lemma:** (i) $\mathcal{F}_T$ is a σ-algebra. Moreover, if $T = n$, then $\mathcal{F}_T = \mathcal{F}_n$.

(ii) Suppose $S \le T$ are stopping times, then $\mathcal{F}_S \subset \mathcal{F}_T$.

(iii) Suppose $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$. Then, $X_T \mathbb{1}_{\{T < +\infty\}}$ is measurable w.r.t. $\mathcal{F}_T$.

(iv) —"— . Then, $\{X_{T\wedge n}\}$ is adapted to $\{\mathcal{F}_n\}$.

**Proof:** (i) Trivial. (iv) $X_{T\wedge n} = (H \cdot X)_n$ for $H_n = \mathbb{1}_{\{T \ge n\}}$.

(ii) $\forall A \in \mathcal{F}_S$, we need to show that $A \cap \{T \le n\} \in \mathcal{F}_n$.

We have $A \cap \{S \le n\} \in \mathcal{F}_n$. Moreover, $A \cap \{S \le n < T\} = (A \cap \{S \le n\}) \cap \{T > n\} \in \mathcal{F}_n$.

~~$A \cap \{T > S > n\} = A \cap \{S > n\} = A \cap \{S \le n\}^c \in \mathcal{F}_n$.~~

~~Hence, $A \cap \{T > n\} = (A \cap \{S > n\}) \cup (A \cap \{S \le n < T\}) \in \mathcal{F}_n$~~

Thus, $\mathcal{F}_n \ni A \cap \{S \le n\} - A \cap \{S \le n < T\} = A \cap \{S \le n\} \cap \{T \le n\} = A \cap \{T \le n\}$.

(iii) We need to show that $\{X_T \in B\} \cap \{T \le n\} \in \mathcal{F}_n$ $\forall$ Borel $B$ and $n$.

$$\{X_T \in B\} \cap \{T \le n\} = \bigcup_{j \le n} \big( \{X_j \in B\} \cap \{T = j\} \big) \in \bigcup_{j \le n} \mathcal{F}_j = \mathcal{F}_n.$$ □

**Thm:** (Optional Stopping Theorem 1) Let $\{X_n\}$ be a sub-mg.

(i) If $T$ is a stopping time, then $\{X_{T\wedge n}\}$ is a sub-mg, and in particular,

$$\mathbb{E}X_0 \le \mathbb{E}X_{T\wedge n} \quad \forall n.$$

(ii) If $T$ is a stopping time with $\mathbb{P}(T \le n) = 1$, then $X_T \le \mathbb{E}(X_n \mid \mathcal{F}_T)$ a.s. Furthermore, if $S \le T$ a.s. are stopping times, then $X_S \le \mathbb{E}(X_T \mid \mathcal{F}_S)$ a.s. In particular, we have $\mathbb{E}X_0 \le \mathbb{E}X_T$.

Pf: (i) Proved. (ii) $\forall A \in \mathcal{F}_T$, we have

$$\mathbb{E}(X_n \mathbb{1}_A) = \sum_{i=0}^n \mathbb{E}(X_n \mathbb{1}_A \mathbb{1}_{\{T=i\}}) = \sum_{i=0}^n \mathbb{E}(X_n \underbrace{\mathbb{1}_{A\cap\{T=i\}}}_{\in \mathcal{F}_i})$$

$$= \sum_{i=0}^n \mathbb{E}[\mathbb{E}[X_n \mid \mathcal{F}_i] \mathbb{1}_A \mathbb{1}_{\{T=i\}}] \ge \sum_{i=0}^n \mathbb{E}[X_i \mathbb{1}_A \mathbb{1}_{\{T=i\}}] = \mathbb{E}[X_T \mathbb{1}_A].$$

Hence, we have $\mathbb{E}[X_n \mid \mathcal{F}_T] \ge X_T$ a.s. To show the second state, we use that ~~tower property to get that~~ $\cancel{\mathbb{E} X_S \le \mathbb{E}[X_n \mid \mathcal{F}_S] = \mathbb{E}[\mathbb{E}[X_n \mid \mathcal{F}_T] \mid \mathcal{F}_S]}$ $\forall A \in \mathcal{F}_S$,

$$\cancel{\mathbb{E}[\mathbb{E}[X_T \mathbb{1}_A] = \sum_{i=1}^n \mathbb{E}[X_n \mathbb{1}_A \mathbb{1}_{\{S=i\}} \mathbb{1}_{\{T\ge i\}}]}$$

$$\cancel{\mathbb{E}[X_T \mathbb{1}_A] = \sum_{0\le i\le j\le n} \mathbb{E}[X_T \mathbb{1}_A \mathbb{1}_{\{S=i\}} \mathbb{1}_{\{T=j\}}]}$$

$$\cancel{= \sum_{0\le i\le j\le n} \mathbb{E}[X_j \mathbb{1}_{A\cap\{S=i\}} \mathbb{1}_{\{T=j\}}] - \sum_{0\le i\le j\le n} \mathbb{E}[\mathbb{E}[X_j \mathbb{1}_{\{T=j\}} \mid \mathcal{F}_i] \mathbb{1}_{A\cap\{S=i\}}]}$$

$$\cancel{= \sum_{i=0}^n \mathbb{E}[\mathbb{E}[X_T \mathbb{1}_{\{T\ge i\}} \mid \mathcal{F}_i] \mathbb{1}_A \mathbb{1}_{\{S=i\}}] - \sum_{i=0}^n \mathbb{E}[\mathbb{E}[X_T \mid \mathcal{F}_i] \mathbb{1}_A \mathbb{1}_{\{S=i\}}]}$$

$$\mathbb{E}[X_T \mathbb{1}_A] = \sum_{i=0}^n \mathbb{E}[X_T \mathbb{1}_A \mathbb{1}_{\{S=i\}} \mathbb{1}_{\{T\ge i\}}] = \sum_{i=0}^n \mathbb{E}[\mathbb{E}[X_T \mid \mathcal{F}_i] \mathbb{1}_{A\cap\{S=i\}} \mathbb{1}_{\{T\ge i\}}]$$

$$= \sum_{i=0}^n \sum_{j=i}^n \mathbb{E}[\mathbb{E}[X_j \mid \mathcal{F}_i] \mathbb{1}_A \mathbb{1}_{\{S=i\}} \mathbb{1}_{\{T=j\}}] \ge \sum_{i=0}^n \sum_{j=i}^n \mathbb{E}[X_i \mathbb{1}_A \mathbb{1}_{S=i} \mathbb{1}_{T=j}]$$

$$= \mathbb{E}[X_S \mathbb{1}_A]. \quad \Rightarrow \quad \mathbb{E}[X_T \mid \mathcal{F}_S] \ge X_S \text{ a.s.} \quad \square$$

When the stopping time is unbounded, we have the following result.

Theorem (Optional Stopping Theorem 2: UI sub-mg) If $\{X_n\}$ is a UI sub-mg, then for any stopping time $T \le \infty$, we have $\mathbb{E}X_0 \le \mathbb{E}X_T \le \mathbb{E}X_\infty$, where $X_\infty := \lim_{n\to\infty} X_n$.

Proof: We first show that $\{X_{T\wedge n}\}$ is also UI. By $X_n^+ = X_n \vee 0$ is a sub-mg. Then, noticing that $T\wedge n$ and $n$ are both bounded stopping times, the previous results gives $\mathbb{E}X_{T\wedge n}^+ \le \mathbb{E}X_n^+$. Since $X_n^+$ is UI, we have $\sup_n \mathbb{E}X_n^+ < +\infty \Rightarrow \sup_n \mathbb{E}X_{T\wedge n}^+ < +\infty$. Then, by the martingale convergence theorem, ~~$X_{N\wedge n} \to X_N$ a.s. (with $X_\infty$~~ $X_{T\wedge n} \to X_T$ a.s. (with $X_\infty = \lim_{n\to\infty} X_n$) with $\mathbb{E}|X_T| < +\infty$. Then, we have

$$\mathbb{E}[|X_{T\wedge n}|; |X_{T\wedge n}| > M) = \mathbb{E}[|X_T|; |X_T| > M, T \le n]$$
$$+ \mathbb{E}[|X_n|; |X_n| > M, T > n].$$

Since $\{X_n\}$ is UI and $X_T \in L^1$, the RHS is $\le \varepsilon$ if $M$ is large enough.

~~Since $X_{N\wedge n}$ is a sub-mg, we~~ By previous theorem, $\mathbb{E}X_0 \le \mathbb{E}X_{N\wedge n} \le \mathbb{E}X_n$. Letting $n\to\infty$, by the $L^1$-convergence theorem, $X_{N\wedge n} \to X_N$ in $L^1$ and $X_n \to X_\infty$ in $L^1$. Hence, $\mathbb{E}X_0 \le \mathbb{E}X_N \le \mathbb{E}X_\infty$. $\square$

The next result does not require UI if we have bounded increments.

Theorem (Optional Stopping Theorem 3: bounded increments) $\{X_n\}$ is a sub-mg and $\mathbb{E}(|X_{n+1} - X_n| \mid \mathcal{F}_n) \le C < +\infty$ a.s. If $T$ is a stopping time with $\mathbb{E}T < +\infty$, then $\mathbb{E}X_0 \le \mathbb{E}X_T$.

Proof: Again, we only need to show that $X_{T\wedge n}$ is UI. Note

$$|X_{T\wedge n}| \le |X_0| + \sum_{m=0}^\infty |X_{m+1} - X_m| \mathbb{1}_{\{T>m\}} =: Y.$$

It suffices to show that $\mathbb{E}|Y| < +\infty$. Since $\{T>m\} = \{T\le m\}^c \in \mathcal{F}_m$, we have

$$\mathbb{E}(|X_{m+1} - X_m|; T>m) = \mathbb{E}[\mathbb{E}(|X_{m+1} - X_m| \mid \mathcal{F}_m) \mathbb{1}_{\{T>m\}}) \le C\,\mathbb{P}(T>m).$$

$$\Rightarrow \mathbb{E}Y = \mathbb{E}|X_0| + \sum_{m=0}^\infty \mathbb{E}[|X_{m+1} - X_m|; T>m] \le \mathbb{E}|X_0| + C\sum_{m=0}^\infty \mathbb{P}(T>m)$$
$$= \mathbb{E}|X_0| + C\,\mathbb{E}T < +\infty. \quad \square$$

Rmk: The above theorems hold for super-mg and martingales with "$\le$" replaced by "$\ge$" and "$=$".

The next theorem holds for non-negative super-mg without UI and bounded increments.

<u>Theorem</u> (Optional Stopping 4: non-neg super-mg) If $\{X_n\}$ is a non-negative super-martingale and $T \le \infty$ is a stopping time, then $\mathbb{E}X_0 \ge \mathbb{E}X_T$ (with $X_\infty = \lim_n X_n$).

<u>Pf</u>: Appling Fatou to $\mathbb{E}X_0 \ge \mathbb{E}X_{T\wedge n} \Rightarrow \mathbb{E}X_0 \ge \mathbb{E}[\liminf X_{T\wedge n}] = \mathbb{E}X_T$. $\square$

Finally, we summarize the optional stopping theorems for martingales, which are most important to us.

<u>Theorem</u> (Optional Stopping Theorem for martingales) Let $S \le T$ be stopping times. $\{X_n\}$ is a martingale.

(i) $X_{T\wedge n}$ is a martingale and, in particular, $\mathbb{E}X_{T\wedge n} = \mathbb{E}X_0$.

(ii) If $S \le T \le n$ by a constant $n \in \mathbb{N}$, then we have $\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S$ a.s.

In particular, $\mathbb{E}X_T = \mathbb{E}X_0$.

(iii) If $\{X_n\}$ is UI, and $T \le \infty$, we have $\mathbb{E}X_0 = \mathbb{E}X_T = \mathbb{E}X_\infty$.

(iv) If $\mathbb{E}[|X_{n+1} - X_n| \mid \mathcal{F}_n] \le C < +\infty$, and $\mathbb{E}T < +\infty$, then $\mathbb{E}X_0 = \mathbb{E}X_T$.

(v) If $\{X_n\}$ is a non-neg. martingale, we have $\mathbb{E}X_T \le \mathbb{E}X_0$.

(vi) If $\{X_n\}$ is UI, then $\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S$ a.s. In particular, $\mathbb{E}X_T = \mathbb{E}X_S$.

<u>Proof of (vi)</u> We claim that $X_T = \mathbb{E}[X_\infty \mid \mathcal{F}_T]$ a.s. If this holds, then

$$\mathbb{E}[X_T \mid \mathcal{F}_S] = \mathbb{E}[\mathbb{E}[X_\infty \mid \mathcal{F}_T] \mid \mathcal{F}_S] = \mathbb{E}[X_\infty \mid \mathcal{F}_S] = X_S.$$

In proof of (iii), we have shown $\mathbb{E}|X_T| < +\infty$. $\forall A \in \mathcal{F}_T$,

$$\mathbb{E}[X_T \mathbb{1}_A] = \sum_{n\in\mathbb{N}\cup\{\infty\}} \mathbb{E}[X_n \mathbb{1}_A \mathbb{1}_{\{T=n\}}] = \sum_{n\in\mathbb{N}\cup\infty} \mathbb{E}[\mathbb{E}[X_\infty \mid \mathcal{F}_n] \mathbb{1}_{A\cap\{T=n\}}]$$

$$= \sum_{n\in\mathbb{N}\cup\{\infty\}} \mathbb{E}[X_\infty \mathbb{1}_A \mathbb{1}_{\{T=n\}}] = \mathbb{E}[X_\infty \mathbb{1}_A] \Rightarrow X_T = \mathbb{E}[X_\infty \mid \mathcal{F}_T]. \ \square$$

An important example: <u>Gambler's ruin</u>



<u>Example</u> (Gambler's ruin) Consider a gambler making fair ~~coin tosses~~ bets on fair coin tosses. If Head, he wins one yuan; if tail, he loses one dollar. He will abandon the game when his fortune falls to 0 or rises to $N$. Let $X_n$ be gambler's fortune at time $n$. Let $\tau$ be the time when gambler stops the game. Assume that $X_0 = k$, $0 \le k \le n$. Calculate $\mathbb{P}(X_\tau = N)$ and $\mathbb{E}\tau$.

<u>Sol</u>: $\tau = \inf\{n : X_n = 0 \text{ or } N\}$ is a stopping time.

We claim that $X_n$ is a martingale. In fact,

$$\mathbb{E}[X_n \mid \sigma(X_0, X_1, \cdots, X_{n-1})] = \tfrac{1}{2}(X_{n-1}+1) + \tfrac{1}{2}(X_{n-1}-1) = X_{n-1}$$

By optional stopping theorem, we have

$$\mathbb{E}X_{\tau\wedge n} = \mathbb{E}X_0 = k \quad \forall n. \quad (*)$$

We claim that $\tau < \infty$ a.s. In fact,

$$\mathbb{P}(\tau = \infty) = \mathbb{P}[0 < X_n < N, \forall n] \le \liminf \mathbb{P}(0 < X_n < N).$$

But by CLT, $(X_n - X_0)/\sqrt{n}$ converges to standard normal in law. Then,

$$\mathbb{P}(0 < X_n < N) = \mathbb{P}\left(-\frac{k}{\sqrt{n}} < \frac{X_n - k}{\sqrt{n}} < \frac{N-k}{\sqrt{n}}\right) \to 0 \quad \text{as } n \to +\infty.$$

Now, noticing that $|X_{\tau\wedge n}| \le N$, appling DCT to $(*)$ gives that

$$\mathbb{E}X_\tau = \lim_{n\to+\infty} \mathbb{E}X_{\tau\wedge n} = k \Rightarrow N\mathbb{P}(X_\tau = N) = k \Rightarrow \mathbb{P}(X_\tau = N) = \frac{k}{N}.$$

To calculate $\mathbb{E}\tau$, we notice that $\{X_n^2 - n\}$ is a martingale. By OST, we have

$$\mathbb{E}[X_{n\wedge\tau}^2 - n\wedge\tau] = \mathbb{E}X_0^2 = k^2$$

By MCT, we have $\mathbb{E}(n\wedge\tau) \to \mathbb{E}\tau$ as $n \to +\infty$. By DCT, we have

$$\mathbb{E}[X_{n\wedge\tau}^2] \to \mathbb{E}X_\tau^2 \text{ as } n \to +\infty.$$

Thus, $\mathbb{E}[X_\tau^2 - \tau] = k^2 \Rightarrow \mathbb{E}\tau = N^2\mathbb{P}(X_\tau = N) - k^2 = N^2 \cdot \frac{k}{N} - k^2 = (N-k)k$. $\square$

<u>Example</u>: The above example can be regarded as a SRW with two absorbing states.

Consider SRW $S_n$ on $\mathbb{Z}$ with $S_n = \sum_{i=1}^n \xi_i + x$, $\mathbb{P}(\xi_i = \pm 1) = \frac{1}{2}$, $S_0 = x \in \mathbb{Z}$.

Let $a < x < b \in \mathbb{Z}$. Define $T_{a,b} = \inf\{n : n \in (a,b)\}$, the exit time of $(a,b) \cap \mathbb{Z}$. The above example tells that

$$\mathbb{P}_x(S_{T_{a,b}} = a) = 1 - \mathbb{P}_x(S_{T_{a,b}} = b) = \frac{b-x}{b-a}, \qquad \mathbb{E}_x T_{a,b} = (b-x)(x-a).$$

In particular, $\mathbb{E}_0 T_{a,b} = -ab$.

Let $T_a = \inf\{n : S_n = a\}$. Then, taking $b \to +\infty$, we get that

$$\mathbb{P}_x(T_a < +\infty) = \mathbb{P}_x(S_{T_{a,\infty}} = a) = 1. \quad (\text{Recurrent})$$

## Backwards Martingales

A backwards martingale is a martingale indexed by $\mathbb{Z}_-$, i.e. $X_n$, $n \le 0$, adapted to an increasing sequence of σ-fields $\mathcal{F}_n$ ($\cdots \subseteq \mathcal{F}_{n-1} \subseteq \mathcal{F}_n \subseteq \cdots \subseteq \mathcal{F}_{-2} \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}_0$) with $X_n \in L^1$ and

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n \quad \text{for all } n \le -1.$$

Note $X_n = \mathbb{E}[X_0 \mid \mathcal{F}_n]$ for all $n \le -1$, so $\{X_n\}$ is automatically UI. Moreover, the σ-fields ↓ as $n \downarrow -\infty$. The convergence theorem for backwards mg is simple.

Theorem Let $\{X_n : n \le 0\}$ be a backwards martingale. Then, $X_\infty = \lim_{n \to -\infty} X_n$ exists a.s. and in $L^1$. Moreover, if $X_0 \in L^p$ for some $p \in (1,\infty)$, then $X_n \to X_{-\infty}$ in $L^p$.

Pf: Let $U_n$ be the # of upcrossings of $[a,b]$ by $X_{-n}, \cdots, X_0$. Then, applying Doob's upcrossing inequality, we get

$$(b-a)\,\mathbb{E}U_n \le \mathbb{E}(X_0 - a)^+ < +\infty.$$

Letting $n \to \infty$ and using MCT, we get $\mathbb{E}U_\infty < +\infty$, which implies that $X_n$ converges a.s. Since $X_n = \mathbb{E}[X_0 \mid \mathcal{F}_n]$, $\{X_n\}$ is UI and convergence occurs in $L^1$. Convergence in $L^p$ follows by applying Doob's maximal inequality. □

Theorem: Let $X_{-\infty} = \lim_{n \to -\infty} X_n$ and $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$, then $X_{-\infty} = \mathbb{E}[X_0 \mid \mathcal{F}_{-\infty}]$.



Pf: $X_{-\infty} \in \mathcal{F}_n$ $\forall n$, so $X_{-\infty} \in \mathcal{F}_{-\infty}$. $X_n = \mathbb{E}[X_0 \mid \mathcal{F}_n]$, so $\forall A \in \mathcal{F}_{-\infty} \subseteq \mathcal{F}_n$, there is

$$\mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[X_0 \mathbb{1}_A]$$

By convergence in $L^1$, $\mathbb{E}[X_0 \mathbb{1}_A] = \lim_{n \to -\infty} \mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[X_{-\infty} \mathbb{1}_A]$. □

Theorem: $\forall\, Y \in L^1$, we have $\mathbb{E}[Y \mid \mathcal{F}_n] \to \mathbb{E}[Y \mid \mathcal{F}_{-\infty}]$ a.s. and in $L^1$.

Pf: $X_n = \mathbb{E}[Y \mid \mathcal{F}_n]$ is a backwards mg, so $X_n \to X_{-\infty}$ a.s. and in $L^1$, where

$$X_{-\infty} = \mathbb{E}[X_0 \mid \mathcal{F}_{-\infty}] = \mathbb{E}[\mathbb{E}[Y \mid \mathcal{F}_0] \mid \mathcal{F}_{-\infty}] = \mathbb{E}[Y \mid \mathcal{F}_{-\infty}]. \quad \square$$

We now use backwards mg to give another proof of SLLN.

Example: Let $\xi_1, \xi_2, \cdots$ be i.i.d with $\mathbb{E}|\xi_i| < +\infty$. Let $S_n = \sum_{i=1}^n \xi_i$ and $X_{-n} = S_n / n$.

Let $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, \cdots) = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \cdots)$.

We claim that $\{X_n : n \le 0\}$ is a backwards martingale w.r.t. $\{\mathcal{F}_n : n \le 0\}$.

We can calculate that $\mathbb{E}[X_{-n+1} \mid \mathcal{G}_{-n}] = \mathbb{E}\left[\frac{S_{n-1}}{n-1} \,\middle|\, \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \cdots)\right]$

$$= \mathbb{E}\left[\frac{S_{n-1}}{n-1} \,\middle|\, \sigma(S_n)\right].$$

By symmetry, we have $\mathbb{E}[X_j \mid \sigma(S_n)] = \mathbb{E}[X_k \mid \sigma(S_n)]$ $\forall\, 0 \le j, k \le n$.

Thus, $\mathbb{E}[X_j \mid \sigma(S_n)] = \frac{1}{n}\sum_{j=1}^n \mathbb{E}[X_j \mid \sigma(S_n)] = \frac{S_n}{n}$. Hence, we get that

$$\mathbb{E}\left[\frac{S_{n-1}}{n-1} \,\middle|\, \sigma(S_n)\right] = \frac{1}{n-1}\sum_{j=1}^{n-1} \mathbb{E}[X_j \mid \sigma(S_n)] = \frac{S_n}{n} = X_{-n}.$$

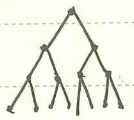
By the previous theorem, we have that $X_{-n} \to X_{-\infty}$ a.s. and in $L^1$ ~~to some~~ ~~with~~ $\cancel{X_{\infty} = \mathbb{E}[X_1 \mid \mathcal{F}_{-\infty}]}$ Note that $Y = \lim_{n \to \infty} \frac{1}{n}\sum_{i=k}^n \xi_i$, so $Y$ is $\mathcal{T}$-measurable, $\mathcal{T}$: tail σ-algebra. By Kolmogorov 0-1 law, the tail algebra is trivial, i.e., $\mathbb{P}(Y = c) = 0$ or $1$ $\forall\, c \in \mathbb{R}$. Hence, there exists $c \in \mathbb{R}$ s.t.

$$\mathbb{P}(Y = c) = 1.$$

Since we have convergence in $L^1$, $c = \mathbb{E}Y = \lim_{n \to \infty} \mathbb{E}X_{-n} = \mathbb{E}\xi_1$. □

## Section 1.4: Applications: Branching Processes (Galton-Watson Tree)

A tree is a connected graph without cycles. A rooted tree has a distinguished vertex, $o$, called the root. A leaf is a vertex with degree 1.



A binary tree.

The depth of a vertex is its graph distance to the root.

For any vertex $v \neq o$ ~~that is not the root~~, it has a unique "parent", i.e., the vertex the neighboring vertex $v$ ~~that~~ with smaller depth. All the other neighboring vertices of $v$ are called the "children" of $v$.

In this section, we study the simplest random tree — the Galton-Watson tree.

(*) Starting with an initial "ancestor", denoted by the root $o$, every vertex has a certain number of children following some distribution $(p_0, p_1, \cdots, p_k, \cdots)$ with $\sum_k p_k = 1$.

I.e., let $Z_n$ be the # of individuals ~~vertices~~ in the $n$-th generation (i.e., # of vertices with depth $= n$). Each member of the $n$-th generation gives birth independently to an identically distributed # of children:

$Z_n$, $n \geq 0$ with $Z_0 = 1$ and

$$Z_{n+1} = \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1}, & \text{if } Z_n > 0 \\ 0, & \text{if } Z_n = 0 \end{cases}$$

$(\xi_1^{n+1}, \xi_2^{n+1}, \cdots)$ are i.i.d. with $\mathbb{P}(\xi_i^{n+1} = k) = p_k$. $\vec{p} = (p_0, p_1, \cdots)$ is called the offspring distribution. This is called the GW tree or GW process.

A natural question is whether this process will become extinct, and if yes, what is the extinction probability:

$$\mathbb{P}(Z_n = 0 \text{ eventually}) = ?$$

Or, if we define the extinction time $T = \inf\{n : Z_n = 0\}$, what is $\mathbb{P}(T < \infty)$?

Of course, the answer depends on the law $\vec{p}$. To avoid trivial situations, throughout the following discussion, we always assume that $p_0 + p_1 < 1$ (otherwise, we trivially have that $\mathbb{P}(T < \infty) = 1$ unless $p_1 = 1$).

We first notice that $Z_n$ can be made to be a martingale after a proper scaling.

<u>Lemma</u>: Let $\mathcal{F}_n = \sigma(\xi_i^m : i \geq 1, 1 \leq m \leq n)$ and $\mu = \sum_k k p_k = \mathbb{E}\xi_i^m \in (0, \infty)$. Then, $X_n = Z_n / \mu^n$ is a martingale w.r.t. $\mathcal{F}_n$.

<u>Proof</u>: First, $Z_n \in \mathcal{F}_n$. Moreover, we have that on $A_k = \{Z_n = k\}$,

$$\cancel{\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\Big(\sum_{i=1}^{Z_n} \xi_i^{n+1} \mid \mathcal{F}_n\Big)}$$

$$\mathbb{E}(Z_n \mathbb{1}_{A_k} \mid \mathcal{F}_n) = \mathbb{E}\Big(\sum_{i=1}^{k} \xi_i^{n+1} \mathbb{1}_{A_k} \mid \mathcal{F}_n\Big) = \mathbb{1}_{A_k} \mathbb{E}\Big(\sum_{i=1}^{k} \xi_i^{n+1}\Big) = \mu k \mathbb{1}_{A_k}$$
$$= \mu Z_n \mathbb{1}_{A_k}.$$

Summing over $k$ and using MCT, we get that $\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = \mu Z_n \Rightarrow \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$.

This also gives that $\mathbb{E} X_n = \mathbb{E} X_0 = 1 < +\infty$, so $X_n \in L^1$. □

$X_n$ is a non-negative martingale, so it converges to a limit $X$ a.s. Note $\mathbb{E} Z_n = \mu^n \to 0$ if $\mu < 1$. So it is natural to have the following result.

<u>Theorem</u>: If $\mu < 1$, then $Z_n = 0$ for all $n$ sufficiently large, so $X_n = Z_n / \mu^n \to 0$.

<u>Pf</u>: $\mathbb{P}(Z_n > 0) \leq \mathbb{E}(Z_n ; Z_n > 0) = \mathbb{E}(Z_n) = \mu^n \to 0$.

By Borel-Cantelli, $\mathbb{P}(Z_n = 0 \text{ eventually}) = 1$. □

~~Intui~~ This is in accordance with our intuition: if each individual on average gives birth to less than one child, the species will finally die out. We will see that this holds even if $\mu = 1$ as long as we exclude the trivial case $p_1 = 1$.

**Theorem:** If $\mu=1$ and $p_1<1$, then $Z_n=0$ for all sufficiently large $n$, i.e., $\mathbb{P}(Z_n=0 \text{ eventually})=1$.

Pf: When $\mu=1$, $Z_n$ itself is a non-negative martingale, so it converges a.s. to a finite limit $Z_\infty$. But $Z_n$ is integer-valued, so we must have $Z_n=Z_\infty$ a.s. for large $n$. If $p_1<1$, then $\forall k>0$, $\mathbb{P}(Z_n=k$ for all $n\ge N)=0$ for any $N$. Hence, we must have $Z_\infty\equiv 0$. □

Next, we show that when $\mu>1$, then $\mathbb{P}(T=\infty)>0$, i.e., there is a positive probability that the process never dies out. To calculate $\rho=\mathbb{P}(T<\infty)$, we introduce the PGF (probability generating function) of $\vec{p}$:

$$f(s)=\mathbb{E}[s^{\xi_i^{n+1}}]=\sum_{k=0}^{\infty} s^k p_k \qquad (0^0=1 \text{ by convention})$$

Note $f(0)=p_0$, $f(1)=1$, $f'(1)=\sum_{k=0}^{\infty} k p_k=\mu$.

**Theorem:** Suppose $\mu>1$. If $Z_0=1$, then $\rho=\mathbb{P}(T<\infty)$ is the unique solution to the equation $f(x)=x$ in $[0,1)$.

Pf: We denote $f_n(s):=\mathbb{E}[s^{Z_n}]$, $s\in[0,1]$, $n\ge 1$.

We claim that $f_n=\underbrace{f\circ f\circ\cdots\circ f}_{n\text{ times}}$. In fact, we have

$$\mathbb{E}[s^{Z_{n+1}}]=\mathbb{E}[\mathbb{E}[s^{Z_{n+1}}\mid\mathcal{G}(Z_n)]]=\mathbb{E}[\mathbb{E}[s^{\sum_{i=1}^{Z_n}\xi_i^{n+1}}\mid\mathcal{G}(Z_n)]]$$

$$=\mathbb{E}[\prod_{i=1}^{Z_n}\mathbb{E}[s^{\xi_i^{n+1}}\mid\mathcal{G}(Z_n)]]=\mathbb{E}[\prod_{i=1}^{Z_n}f(s)]=\mathbb{E}[f(s)^{Z_n}]$$

$$=f_n(f(s)).$$

By induction, we get that $f_n=\overbrace{f\circ\cdots\circ f}^{n\text{ times}}$.

~~Intuitively, $f_n(s)$ will converge to~~ Note $\mathbb{P}(Z_n=0)=f_n(0)$. On the other hand, $\{Z_n=0\}\subseteq\{Z_{n+1}=0\}$, so $f_n(0)=\mathbb{P}(Z_n=0)\le\mathbb{P}(Z_{n+1}=0)=f_{n+1}(0)$.

Thus, $\rho=\mathbb{P}(\bigcup_n\{Z_n=0\})=\lim_{n\to\infty}\mathbb{P}(Z_n=0)=\lim_{n\to\infty}f_n(0)$.

We know $\underbrace{f\circ f\circ\cdots\circ f}_{n}$ should converge to a fixed point of $f(x)$, i.e., a point $x$ with $f(x)=x$.

We now study the properties of $f(x)$.

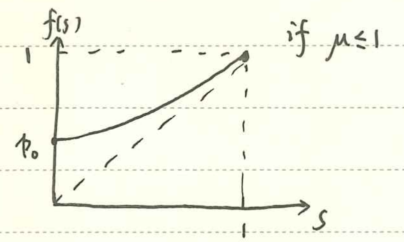
First, $f'(s)=\sum_{k=1}^{\infty}k p_k s^{k-1}>0$ for $s\in(0,1)$, so $f$ is strictly increasing.

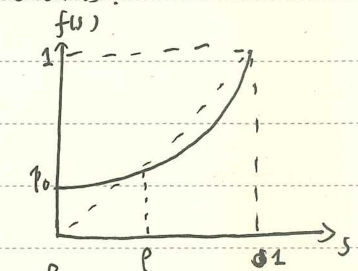
$f''(s)=\sum_{k=2}^{\infty}k(k-1)s^{k-2}>0$ for $s\in(0,1)$ (because $\mu>1$ implies that $p_k>0$ for some $k\ge2$).

Moreover,

$$\lim_{s\uparrow1}f'(s)=\sum_k k p_k=\mu.$$

Thus, we have the following two scenarios:




Hence, there is a unique fixed point $\rho$ inside $(0,1)$ when $\mu>1$.

Now, $f(0)=p_0$, $f_2(0)=f(p_0)$, $f_3(0)=$ …, $f_{n+1}^{(0)}=f(f_n(0))$, … is an increasing sequence in $(0,1)$. We show that $f_n(0)=\mathbb{P}(Z_n=0)\le\rho$. (So it converges to a point, say $\tilde{\rho}$.)

First, we have $p_0\le\rho$. Moreover, if $f_{m-1}(0)\le\rho$, then

$$f_m(0)=f(f_{m-1}(0))\le f(\rho)=\rho.$$

Now, taking $n\to\infty$ in the equation $f_n(0)=f(f_{n-1}(0))$, we get $\tilde{\rho}=f(\tilde{\rho})$, i.e., $\tilde{\rho}$ is a fixed point of $f$. Since $\tilde{\rho}\le\rho$, it follows that $\tilde{\rho}=\rho$.

The above argument also shows that $\mathbb{P}(Z_n=0)\to 0$ when $\mu\le1$ (because in this case, the only fixed point of $f(x)$ is $x=1$.) □

when $\mu > 1$, $X_n = Z_n/\mu^n$ has a chance of converging to a non-zero limit. Whether this happens is due to the following result by Kesten-Stigum.

**Theorem** (Kesten-Stigum)

$$\mathbb{E}[X_\infty] = 1 \iff \mathbb{P}(X_\infty > 0 \mid \text{survival}) = 1 \iff \mathbb{E}(Z_1 \log^+ Z_1) < \infty.$$

Hence, $X_\infty$ is not $\equiv 0$ if and only if $\sum_{k \geq 2} p_k k \log k < +\infty$.

## Chapter II. Markov chains (discrete time)

### 1. Markov chains on countable state space S: examples

Markov chain on a countable space $S$ is a stochastic process $X_n : n \in \mathbb{N}$ satisfying the "Markov property": $\forall$ states $i_0, i_1, \cdots, i_{n-1}, i$ and $j$,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

In words, given the present state, the rest of the past is irrelevant for predicting the future. If the "transition probability" $p(i,j) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ does not depend on $n$, we say the Markov chain is time-homogeneous.

**Examples**: (Random walks on graphs) A graph $G = (V, E)$ consists of a vertex set $V$ and an edge set $E$. The edges in $E$ are undirected. We write $x \sim y$ ($y$ is a neighbor of $x$) if $\{x, y\} \in E$. The degree of $x$, $\deg(x)$, is the # of neighbors of $x$.

The we define the simple Random walk on $G$ to be the Markov chain with state space $V$ and transition matrix

$$P(x,y) = \begin{cases} \frac{1}{\deg(x)} & \text{if } y \sim x, \\ 0 & \text{otherwise} \end{cases}$$



**Example** (The branching process) $S = \mathbb{N}$, $X_{n+1} = \sum_{i=1}^{X_n} \xi_i$, $\xi_i$ are i.i.d. non-neg. integer-valued random variables. Then, we have

$$p(i,j) = \mathbb{P}\left(\sum_{m=1}^{i} \xi_m = j\right)$$

**Example** (Ehrenfest chain) There is a total of $r$ balls in two urns; $k$ in the first and $r-k$ in the second. We pick one of the $r$ balls at random and move it to the other urn. $S = \{0, 1, \cdots, r\}$

$$p(k, k+1) = (r-k)/r$$
$$p(k, k-1) = k/r$$
$$p(i,j) = 0 \quad \text{otherwise}$$

When the state space is finite, the transition probabilities can be arranged into a matrix

$P$ : $P_{ij} = p(i,j)$ transition matrix

Graphical notation: $i \xrightarrow{p(i,j)} j$

**Example** (Gambler's Ruin / SRW with absorbing boundaries)

The gambler wins / ~~does~~ loses with probability $p$ / $1-p$, and the game ends when $X_n = 0$ or $N$. $\tau = \inf\{n \geq 0 : X_n = 0 \text{ or } N\}$.

Then $X_n$ is a Markov chain with

$$p(i,j) = \begin{cases} p, & \text{if } j = i+1, \quad 0 < i < N \\ 1-p, & \text{if } j = i-1, \quad 0 < i < N \\ 1, & \text{if } i = j = 0 \ / \ i = j = N \\ 0, & \text{otherwise} \end{cases}$$

e.g., for $N = 4$,

$$\begin{array}{cc} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ P = & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \end{array}$$

When $p = \frac{1}{2}$, we have shown that $\mathbb{P}(X_\tau = N \mid X_0 = k) = \frac{k}{N}$. We now give a different proof using the technique of Markov chains. Denote

$$f(k) := \mathbb{P}(X_\tau = N \mid X_0 = k) = \mathbb{P}_k(X_\tau = N)$$

For any $1 \le k \le N-1$, we have

$$f(k) = \mathbb{P}(X_\tau = N, X_1 = k+1 \mid X_0 = k) + \mathbb{P}(X_\tau = N, X_1 = k-1 \mid X_0 = k)$$

$$= \mathbb{P}(X_\tau = N \mid X_1 = k+1, X_0 = k)\,\mathbb{P}(X_1 = k+1 \mid X_0 = k) +$$

$$\mathbb{P}(X_\tau = N \mid X_1 = k-1, X_0 = k)\,\mathbb{P}(X_1 = k-1 \mid X_0 = k)$$

Markov property $\Leftarrow$ $= p f(k+1) + (1-p) f(k-1)$,

and $f(0) = 0$, $f(N) = 1$

$$\Rightarrow p[f(k+1) - f(k)] = (1-p)[f(k) - f(k-1)]$$

$$\Rightarrow f(k+1) - f(k) = \left(\frac{1-p}{p}\right)^k (f(1) - f(0)), \quad 0 \le k \le N-1$$

$$\Rightarrow 1 = f(N) - f(0) = \left[\left(\frac{1-p}{p}\right)^{N-1} + \cdots + \frac{1-p}{p} + 1\right](f(1) - f(0)) = \frac{1-\rho^N}{1-\rho}(f(1) - f(0)),$$

$\rho = \frac{1-p}{p}$. we understand the case $\rho = 1$ as

$$\lim_{\rho \to 1} \frac{1-\rho^N}{1-\rho} = N$$

Thus, we get $f(1) - f(0) = \frac{1-\rho}{1-\rho^N}$ and we can calculate that $\forall k \ge 1$,

$$f(k) = (f(1) - f(0)) + (f(2) - f(1)) + \cdots + (f(k) - f(k-1))$$

$$= [f(1) - f(0)](1 + \rho + \cdots + \rho^{k-1}) = \frac{1-\rho^k}{1-\rho^N}.$$

Let $\rho \to 1$, we get $f(k) = \frac{k}{N}$, which is the result we get before.

When $p = \frac{1}{2}$, we can also evaluate $g(k) := \mathbb{E}[\tau \mid X_0 = k] = \mathbb{E}_k(\tau)$. For $1 \le k \le N-1$,

$$g(k) = \mathbb{E}_k[\tau] = \frac{1}{2}\mathbb{E}_k(\tau \mid X_1 = k+1) + \frac{1}{2}\mathbb{E}_k(\tau \mid X_1 = k-1)$$

$$\cancel{= p\, g(k+1) + (1-p)\, g(k-1) = \frac{1}{2} g}$$

$$= \frac{1}{2}[\mathbb{E}_{k+1}(\tau) + 1] + \frac{1}{2}[\mathbb{E}_{k-1}(\tau) + 1] = \frac{1}{2}(g(k+1) + g(k-1)) + 1.$$

~~$\mathbb{E}(B) = a$, VE~~ $g(0) = g(N) = 0$.

One can check that there exists a unique solution: $g(k) = k(N-k)$, $0 \le k \le N$. □

2. Construction of Markov chains, Markov properties

Let $(S, \mathcal{F})$ be a measurable space. $S$ is the state space (countable / uncountable).

A function $p: S \times \mathcal{F} \to \mathbb{R}_{\ge 0}$ is said to be a transition probability if

(i) For each $x \in S$, $p(x, \cdot)$ is a probability measure on $(S, \mathcal{F})$.

(ii) For each $A \in \mathcal{F}$, $x \to p(x, A)$ is a measurable function.

We say $X_n$ is a Markov chain with transition probability $p$ if

$$\mathbb{P}(X_{n+1} \in B \mid \mathcal{F}_n) = p(X_n, B) \quad \text{where } \mathcal{F}_n = \sigma(X_1, X_2, \cdots, X_n).$$

Construction of a Markov chain

Given a transition probability $p$ and an initial distribution $\mu$ on $(S, \mathcal{F})$, we can define

$$\mathbb{P}(X_j \in B_j, 0 \le j \le n) = \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \int_{B_2} p(x_1, dx_2) \cdots \int_{B_n} p(x_{n-1}, dx_n) \quad (\dagger)$$

This gives a sequence of consistent probability measures $\mu_n$ with

$$\mu_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Kolmogorov's extension theorem allows us to construct a probability measure $\mathbb{P}_\mu$ on sequence space $(S^{\mathbb{N}}, \sigma(\mathcal{F}^{\mathbb{N}}))$ so that ~~the coordinate~~ for $\omega \in S^{\mathbb{N}}$, the coordinate maps $X_n(\omega) = \omega_n$ has the desired distribution $(\dagger)$.

Note we have a family of measures, one for each initial distribution $\mu$.

But these measures are all built from $\mathbb{P}_x \equiv \mathbb{P}_{\delta_x}$:

$$\mathbb{P}_\mu(A) = \int \mu(dx)\, \mathbb{P}_x(A).$$

On the sequence space, we can define the shift operator:

$$\theta_n(\omega_0, \omega_1, \omega_2, \cdots) = (\omega_n, \omega_{n+1}, \omega_{n+2}, \cdots).$$

Thm: $\{X_n\}$ defined above is a Markov chain w.r.t. $\mathcal{F}_n$ with transition probability $p$, i.e., $\mathbb{P}_\mu(X_{n+1} \in B \mid \mathcal{F}_n) = p(X_n, B)$.

Pf: We let $A=\{X_0\in B_0, X_1\in B_1, X_2\in B_2, \cdots, X_n\in B_n\}$, $B_{n+1}=B$. By the constuction, we have that

$$\int_A \mathbb{1}_{X_{n+1}\in B}\, d\mathbb{P}_\mu = \mathbb{P}_\mu(A, X_{n+1}\in B) = \mathbb{P}_\mu(X_0\in B_0, X_1\in B_1, \cdots, X_n\in B_n, X_{n+1}\in B_{n+1})$$

$$= \int_{B_0}\mu(dx_0)\int_{B_1}p(x_0,dx_1)\int_{B_2}p(x_1,dx_2)\cdots\int_{B_n}p(x_{n-1},dx_n)\ \ p(x_n,B_{n+1}).$$

Need to show: RHS $=\int_A p(X_n,B)\,d\mathbb{P}_\mu$.

We again apply the 4-step strategy: for any $C\in\mathcal{F}$, we have

$$\int_{B_0}\mu(dx_0)\int_{B_1}p(x_0,dx_1)\cdots\int_{B_n}p(x_{n-1},dx_n)\,\mathbb{1}_C(x_n)$$

$$=\int_{B_0}\mu(dx_1)\int_{B_1}p(x_0,dx_1)\cdots\int_{B_n\cap C}p(x_{n-1},dx_n) = \mathbb{P}_\mu(A, X_n\in C)$$

$$=\int_A \mathbb{1}_C(X_n)\,d\mathbb{P}_\mu.$$

Then extend it to simple functions, then positive measurable functions by MCT, and finally general measurable functions, i.e. we have obtained that

$$\int_{B_0}\mu(dx_0)\int_{B_1}p(x_0,dx_1)\cdots\int_{B_n}p(x_{n-1},dx_n)\,f(x_n) = \int_A f(X_n)\,d\mathbb{P}_\mu.$$

In particular, applying it to $f(x_n)=p(x_n,B)$ gives

$$(\dagger)\quad \int_A \mathbb{1}_{X_{n+1}\in B}\,d\mathbb{P}_\mu = \int_A p(X_n,B)\,d\mathbb{P}_\mu.$$

The collection of sets $A$ for which $(\dagger)$ holds is a $\lambda$-system, and ~~we~~ we have proved $(\dagger)$ holds for a $\pi$-system that generates $\mathcal{F}_n$. Thus, by $\pi$-$\lambda$ thm, $(\dagger)$ holds for any $A\in\mathcal{F}_n$. Hence, we have shown that

$$\mathbb{P}(X_{n+1}\in B\mid\mathcal{F}_n) = p(X_n,B).$$ □

The next result is a simple but very useful one.

Thm: If $X_n$ is a Markov chain with transition probability $p$, then $\forall$ bounded measurable function $f$ we have

$$\mathbb{E}[f(X_{n+1})\mid\mathcal{F}_n] = \int p(X_n,dy)\,f(y). \qquad (*)$$

Pf: Let $\mathcal{H}$ be the collection of all bounded functions for which $(*)$ holds. Then, we will use the following Meta-theorem:

Thm (Monotone class theorem) Let $\mathcal{A}$ be a $\pi$-system that contains $\Omega$ and let $\mathcal{H}$ be a collection of real-valued function that satisfies:

(i) If $A\in\mathcal{A}$, then $\mathbb{1}_A\in\mathcal{H}$.

(ii) If $f,g\in\mathcal{H}$, then $f+g$ and $cf\in\mathcal{H}$ for any $c\in\mathbb{R}$.

(iii) If $f_n\in\mathcal{H}$ are non-negative and $f_n\uparrow f$, then $f\in\mathcal{H}$.

Then, $\mathcal{H}$ contains all bounded functions measurable with respect to $\sigma(\mathcal{A})$.

Pf: The assumption $\Omega\in\mathcal{A}$, (ii) and (iii) imply that $\mathcal{G}=\{A:\mathbb{1}_A\in\mathcal{H}\}$ is a $\lambda$-system. Then, by $\pi$-$\lambda$ theorem, $\sigma(\mathcal{A})\subset\mathcal{G}$. Then, using (ii), we get that $\mathcal{H}$ contains all simple $\sigma(\mathcal{A})$-measurable functions. Next, using (iii), we get that $\mathcal{H}$ contains all bounded measurable functions. □

Then, $(*)$ is a direct consequence of this theorem. □

Next, we will prove two important extensions of the Markov property in which $\{X_{n+1}\in B\}$ is replaced by a bounded function of the future, $h(X_n, X_{n+1}, \cdots)$, and $n$ is replaced by a stopping time $N$.

Theorem (Markov property) Let $Y$ be a bounded measurable function. Then,

$$\mathbb{E}_\mu(Y\circ\theta_m\mid\mathcal{F}_m) = \mathbb{E}_{X_m}Y.$$

Here, $\mathbb{E}_\mu$ means conditional expectation w.r.t. $\mathbb{P}_\mu$. The RHS is a RV. It should be thought of as the function $\varphi(x)=\mathbb{E}_xY$ evaluated at $x=X_m$.

$Y\circ\theta_m(\omega) = Y(\omega_m, \omega_{m+1}, \cdots)$, $Y$ evaluated at shifted sequences.

Pf: Again we prove the result in a special case and then extend it using the $\pi$-$\lambda$ and monotone class theorems.

Let $A = \{w: w_0 \in A_0, w_1 \in A_1, \cdots, w_m \in A_m\}$ and $g_0, \cdots, g_n$ be bounded measurable.

Lemma: For bounded measurable $f_0, f_1, \cdots, f_n$,

$$\mathbb{E}\Big(\prod_{m=0}^{n} f_m(X_m)\Big) = \mathbb{E}\Big[\mathbb{E}\Big(\prod_{m=0}^{n} f_m(X_m) \,\Big|\, \mathcal{F}_{n-1}\Big)\Big]$$

$$= \mathbb{E}\Big[\prod_{m=0}^{n-1} f_m(X_m)\, \mathbb{E}\big(f_n(X_n) \mid \mathcal{F}_{n-1}\big)\Big]$$

$$= \mathbb{E}\Big[\prod_{m=0}^{n-1} f_m(X_m) \underbrace{\int p_{n-1}(X_{n-1}, dy) f_n(y)}_{\text{This is a bounded measurable function of } X_{n-1}}\Big]$$

By induction, we get

$$\mathbb{E}\Big(\prod_{m=0}^{n} f_m(X_m)\Big) = \int \mu(dx_0) f_0(x_0) \int p_0(x_0, dx_1) f_1(x_1) \cdots \int p_{n-1}(x_{n-1}, dx_n) f(x_n). \quad \square$$

Applying this lemma with $f_k = \mathbb{1}_{A_k}$ for $k < m$, $f_m = \mathbb{1}_{A_m} g_0$, and $f_k = g_{k-m}$ for $m < k \le m+n$, we get that

$$\mathbb{E}_\mu\Big[\prod_{k=0}^{n} g_k(X_{k+m}); A\Big] = \int_{A_0} \mu(dx_0) \int_{A_1} p(x_0, dx_1) \cdots \int_{A_m} p(x_{m-1}, dx_m)\, \underline{g_0(x_m) \int p(x_m, dx_{m+1})\, g_1(x_{m+1}) \cdots \int p(x_{m+n-1}, dx_{m+n})\, g_n(x_{m+n})}$$

$$= \mathbb{E}_\mu\Big(\mathbb{E}_{X_m}\Big(\prod_{k=0}^{n} g_k(X_k)\Big); A\Big). \quad (**)$$

We can show that the collection of sets $\mathcal{L}$ for which $(**)$ holds is a $\lambda$-system, and the collection for which it has been proven is a $\pi$-system generating $\mathcal{F}_m$. Hence, by $\pi$-$\lambda$ theorem, $(**)$ holds for all $A \in \mathcal{F}_m$. In other words, we have shown that

$$\mathbb{E}_\mu(Y \circ \theta_m \mid \mathcal{F}_m) = \mathbb{E}_{X_m} Y,$$

for $Y$ of the form $Y = \prod_{m=0}^{n} f_m(X_m)$.

Fix any $A \in \mathcal{F}_m$ and let $\mathcal{H}$ be the collection of ~~sets~~ bounded measurable $Y$ for which

$$\mathbb{E}_\mu(Y \circ \theta_m; A) = \mathbb{E}_\mu(\mathbb{E}_{X_m} Y; A) \quad (**) \quad \text{holds.}$$

We have shown that $(**)$ holds when $Y(w) = \prod_{k=0}^{n} g_k(w_k)$.

Let $\mathcal{A}$ be the collection of sets of the form $\{w: w_0 \in A_0, w_1 \in A_1, \cdots, w_n \in A_n\}$.

$\mathcal{A}$ is a $\pi$-system containing $\Omega$ and taking $g_k = \mathbb{1}_{A_k}$ shows (i) holds. $\mathcal{H}$ obviously satisfies (ii) and (iii). Hence, by monotone class theorem, $\mathcal{H}$ contains all bounded measurable function w.r.t. $\sigma(\mathcal{A})$. $\square$

Corollary (Chapman - Kolmogorov equation)

$$\mathbb{P}_x(X_{m+n} = z) = \sum_y \mathbb{P}_x(X_m = y)\, \mathbb{P}_y(X_n = z)$$

Pf: $\mathbb{P}_x(X_{n+m} = z) = \mathbb{E}_x[\mathbb{P}_x(X_{n+m} = z \mid \mathcal{F}_m)] = \mathbb{E}_x[\mathbb{P}_{X_m}(X_n = z)]$ by the Markov property, with $\mathbb{1}_{X_n = z} \circ \theta_m = \mathbb{1}_{X_{n+m} = z}$. $\square$

Recall that $N$ is a stopping time if $\{N = n\} \in \mathcal{F}_n$, and define the $\sigma$-algebra

$$\mathcal{F}_N = \{A: A \cap \{N = n\} \in \mathcal{F}_n \text{ for all } n\}$$

Define the random shift operator

$$\theta_N w = \begin{cases} \theta_n w & \text{on } \{N = n\} \\ \Delta & \text{on } \{N = \infty\} \end{cases} \quad \text{where } \Delta \text{ is an extra point added to } \Omega.$$

We will explicitly restrict our attention to $\{N < \infty\}$.

☆ Theorem (Strong Markov Property) Suppose that for each $n$, $Y_n: \Omega \to \mathbb{R}$ is measurable and $|Y_n| \le M$ for all $n$. Then,

$$\mathbb{E}_\mu(Y_N \circ \theta_N \mid \mathcal{F}_N) = \mathbb{E}_{X_N} Y_N \quad \text{on } \{N < \infty\}.$$

The RHS is $\varphi(x, n) = \mathbb{E}_x Y_n$ evaluated at $x = X_N$ and $n = N$.

Proof: Let $A \in \mathcal{F}_N$. Then,

$$\mathbb{E}_\mu(Y_N\circ\theta_N;\ A\cap\{N<\infty\}) = \sum_{n=0}^{\infty}\mathbb{E}_\mu(Y_n\circ\theta_n;\ \overbrace{A\cap\{N=n\}}^{\in\mathcal{F}_n})$$

$$= \sum_{n=0}^{\infty}\mathbb{E}_\mu(\mathbb{E}_{X_n}Y_n;\ A\cap\{N=n\})$$

$$= \mathbb{E}_\mu(\mathbb{E}_{X_N}Y_N;\ A\cap\{N<\infty\}).\quad\square$$

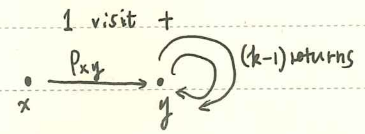
Rmk: For discrete time Markov chain, the proof of strong MP is trivial by breaking things down according to the value of $N$. For continuous time Markov process, this is much more difficult.

We now consider several applications of the strong MP.

① Return times: $T_y^0=0$, and for $k\ge1$, $T_y^k := \inf\{n>T_y^{k-1}: X_n=y\}$.

$T_y^k$ is the time of $k$-th return to $y$. (Note $T_y^1>0$ so any visit at time 0 does not count.) Let $T_y := T_y^1$ and $\rho_{xy} = \mathbb{P}_x(T_y<+\infty)$. Then:

Theorem: $\mathbb{P}_x(T_y^k<\infty) = \rho_{xy}\rho_{yy}^{k-1}$.




Proof: We prove it by induction and strong MP. $k=1$: trivial. Suppose $k\ge2$.
Let $Y(\omega)=1$ if $\omega_n=y$ for some $n\ge1$, $Y(\omega)=0$ otherwise. If $N=T_y^{k-1}$, then $Y\circ\theta_N=1$ if $T_y^k<+\infty$. The strong MP then implies that

$$\mathbb{E}_x(Y\circ\theta_N\,|\,\mathcal{F}_N) = \mathbb{E}_{X_N}Y \quad\text{on } \{N<+\infty\}.$$

On $\{N<+\infty\}$, $X_N=y$, so $\mathbb{E}_{X_N}Y = \mathbb{E}_yY = \mathbb{P}_y(T_y<+\infty) = \rho_{yy}$. Hence,

$$\mathbb{P}_x(T_y^k<+\infty) = \mathbb{E}_x(Y\circ\theta_N;\ N<+\infty)$$

$$= \mathbb{E}_x(\mathbb{E}(Y\circ\theta_N|\mathcal{F}_N);\ N<+\infty) = \rho_{yy}\mathbb{E}_x(N<+\infty)$$

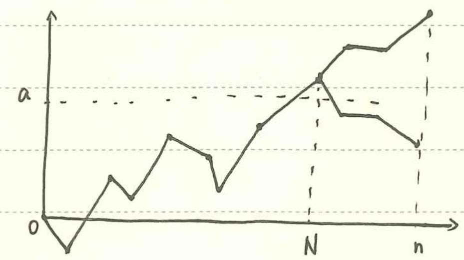
$$= \rho_{yy}\mathbb{P}_x(T_y^{k-1}<+\infty).\quad\square$$

② Reflection principle: (Here, we need to allow $Y$ to depend on $n$.)

Theorem: Let $\xi_1,\xi_2,\cdots$ be ~~independent~~ i.i.d. with a distribution that is symmetric about 0. Let $S_n=\xi_1+\xi_2+\cdots+\xi_n$. If $a\ge0$, then

$$\mathbb{P}(\sup_{m\le n}S_m\ge a) \le 2\,\mathbb{P}(S_n\ge a).$$

Pf by picture:




$N=\inf\{m\le n: S_m>a\}$ (with $\inf\phi=\infty$), then on $\{N<\infty\}$, $S_n-S_N$ is independent of $S_N$ and has $\mathbb{P}(S_n-S_N\ge0)\ge\frac12$.

Hence, $\mathbb{P}(S_n>a)\ge\frac12\mathbb{P}(N\le n) = \frac12\mathbb{P}(\sup_{m\le n}S_m\ge a)$.

Rigorous proof with strong MP: Define $Y_m(\omega)=1$ if $m\le n$ and $\omega_{n-m}\ge a$, $Y_m(\omega)=0$ otherwise.

If $\omega_n\ge a$ (so $N\le n$) and $N=k\le n$, then

$$Y_k\circ\theta_k(\omega) = Y_k(\theta_k(\omega)) = 1,\quad\text{since } (\theta_k\omega)_{n-k}=\omega_n\ge a.$$

$$\Rightarrow\ Y_N\circ\theta_N(\omega)=1 \quad\text{if } \omega_n\ge a.$$

If $\omega_n<a$ and $N=k<n$, then

$$Y_k\circ\theta_k(\omega) = Y_k(\theta_k(\omega)) = 0,\quad\text{since } (\theta_k\omega)_{n-k}=\omega_n<a.$$

~~If $\omega_n<a$ and $N=+\infty>n$, then $Y_N\circ\theta_N(\omega)=0$.~~

Thus, the strong MP gives that $\mathbb{E}_0(Y_N\circ\theta_N\,|\,\mathcal{F}_N) = \mathbb{E}_{S_N}Y_N$ on $\{N<\infty\}=\{N\le n\}$.

For the RHS, note if $y>a$, then

$$\mathbb{E}_yY_m = \mathbb{P}_y(S_{n-m}\ge a) \ge \mathbb{P}_y(S_{n-m}\ge y)\ge\tfrac12.$$

Since $\{N\le n\}\in\mathcal{F}_N$, integrating over $\{N\le n\}$ and using the def. of conditional expectation gives that $\frac12\mathbb{P}(N\le n) \le \mathbb{E}_0(\mathbb{E}_0(Y_N\circ\theta_N|\mathcal{F}_N);\ N\le n) = \mathbb{E}_0(Y_N\circ\theta_N;\ N\le n)$

$$= \mathbb{E}_0(\mathbb{1}_{\{S_n\ge a\}};\ N\le n) = \mathbb{P}(S_n\ge a).\quad\square$$

3. Recurrence and Transience

Recall $T_y^0 = 0$ and the def. of $T_y^k$, $\rho_{xy} = \mathbb{P}_x(T_y < +\infty)$.

A state $y$ is said to be <u>recurrent</u> if $\underline{\rho_{yy} = 1}$ and <u>transient</u> if $\underline{\rho_{yy} < 1}$.

* If $y$ is recurrent, then we have $\mathbb{P}_y(T_y^k < +\infty) = \rho_{yy}^k = 1$ for all $k$, so

$\mathbb{P}_y(X_n = y \text{ i.o.}) = 1$.

* If $y$ is transient and let $N(y) = \sum_{n=1}^{\infty} \mathbb{1}(X_n = y)$ be the # of visits to $y$ at positive times,

then $\mathbb{E}_x N(y) = \sum_{k=1}^{\infty} \mathbb{P}_x(N(y) \geq k) = \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^k < +\infty) = \sum_{k=1}^{\infty} \rho_{xy} \rho_{yy}^{k-1} = \dfrac{\rho_{xy}}{1 - \rho_{yy}} < +\infty$.

These give the following conclusion:

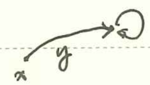
<u>Theorem</u>: A state $y$ is recurrent if and only if $\mathbb{E}_y N(y) = \infty$.

<u>Theorem</u>: If $x$ is recurrent and $\rho_{xy} > 0$ (or $x \rightsquigarrow y$), then $y$ is recurrent and $\rho_{yx} = 1$.

<u>Pf</u>: We first show $\rho_{yx} = 1$. If $\rho_{xy} > 0$ and $\rho_{yx} < 1$, then $\overset{\text{WTS}}{\underline{\rho_{xx} < 1}}$.

To see this, let $K = \inf\{k : p^k(x,y) > 0\}$.

There is a sequence $y_1, \cdots, y_{K-1}$ so that




$$p(x, y_1)\, p(y_1, y_2) \cdots p(y_{K-1}, y) > 0.$$

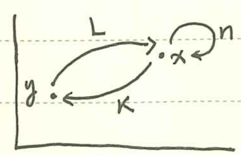
So $K$ is minimal, $y_i \neq x \;\; \forall\, 1 \leq i \leq K-1$. If $\rho_{yx} < 1$, then

$$\mathbb{P}_x(T_x = \infty) \geq p(x, y_1) \cdots p(y_{K-1}, y) \underbrace{\mathbb{P}_y(T_x = \infty)}_{1 - \rho_{yx}} > 0,$$ a contradiction.

Hence, $\rho_{yx} = 1$.

It remains to show $y$ is recurrent. Since $\rho_{yx} > 0$, there is an $L > 0$ so that $p^L(y,x) > 0$.

Then, $p^{L+n+K}(y,y) \geq p^L(y,x)\, p^n(x,x)\, p^K(x,y)$




Summing over $n$ gives $\sum_{n=1}^{\infty} p^{L+n+K}(y,y) \geq p^L(y,x)\, p^K(x,y) \underbrace{\sum_{n=1}^{\infty} p^n(x,x)}_{=\infty}$

Hence, the above theorem implies that $y$ is recurrent. $\square$