Review of Probability Theory II

Park, Sihyung
naturale0@snu.ac.kr

Last Update: January 2021

Contents

1 Central Limit Theorem

1.1 Infinitely Divisible Distributions

A certain kind of well behaving distributions has characteristic functions that can be represented in canonical form. In this section we cover conditions that such distributions have and its canonical representation.

1.1.1 Infinitely divisible distributions

Definition 1 (infinitely divisible distribution)**.** *Let F be a distribution with characteristic function φ. F is inifinitely divisible (ID for short) if one of the followings hold.*
(i) There exists a squence of distributions (F_n) such that $F = F_n * \cdots * F_n$ for all $n \in \mathbb{N}$.
(ii) There exists random variables X, X_{nk} in a probability space $(\Omega, \mathcal{F}, P)$ such that $X \overset{d}{=} X_{n1} + \cdots + X_{nn}$ for all n, where $X \sim F$, $X_{nk} \sim F_n$ for all k and X_{nk}'s are rowwise independent.
(iii) There exists a sequence of characteristic functions (φ_n) such that $\varphi = (\varphi_n)^n$.

Here $*$ denotes convolution. In fact all three conditions are equivalent. As an example, we can easily check that a normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ is infinitely divisible since $X \overset{d}{=} X_{n1} + \cdots + X_{nn}$ for rowwise independent $X_{nk} \sim \mathcal{N}(\frac{\mu}{n}, \frac{\sigma^2}{n})$.

First important property is that characteristic functions of ID distributions always have non-zero values. For this, we need a lemma that applies to all characteristic functions.

Lemma 1. *For a ch.f. φ,*

$$1 - |\varphi(2t)|^2 \leq 4(1 - |\varphi(t)|^2).$$

Proof. Proof is simple using elementary trigonometrics. Notice that $|\varphi|^2$ is a real-valued ch.f. so it suffices to show that $\text{Re}(1 - \varphi(2t)) \leq 4\text{Re}(1 - \varphi(t))$ for a given φ.

$$\begin{aligned}
\text{Re}(1 - \varphi(t)) &= \int (1 - \cos tx) dF(x) \\
&= \int 2 \sin^2 \frac{t}{2} dF(x) \\
&= \int \frac{\sin^2 tx}{2 \cos^2 \frac{t}{2} x} dF(x) \\
&= \int \frac{1}{2} \sin^2 tx dF(x) \\
&= \int \frac{1}{4}(1 - \cos 2tx) dF(x) \\
&= \frac{1}{4}\text{Re}(1 - \varphi(2t)).
\end{aligned}$$

□

Theorem 1.

For an infinitly divisible φ, $\varphi(t) \neq 0, \ \forall t$.

Proof. Proof is by induction. Since φ is ID, there exists φ_n such that $\varphi = (\varphi_n)^n$. We know that $\varphi \to 1$ as $t \to 0$, so there exists $a > 0$ such that $|\varphi(t)| > 0$ for all $|t| \leq a$.

Given t that $|t| \leq a$,

$$|\varphi_n(t)| = |\varphi(t)^{\frac{1}{n}}| \geq \left(\inf_{|t| \leq a} |\varphi(t)| \right)^{\frac{1}{n}} \to_n 1.$$

so for all $0 \leq \epsilon \leq 1$, there exists $N > 0$ such that $|\varphi_n(t)| > 1 - \epsilon$ for all $n \geq N$. By the lemma, for $n \geq N$ and $|t| \leq a$,

$$1 - |\varphi_n(2t)|^2 \leq 4(1 - (1 - \epsilon)^2) \leq 8\epsilon.$$

Thus for large n, $|\varphi_n(2t)|^2 \geq 1 - 8\epsilon > 0$, for all $0 < \epsilon < 1/8$. This gives $\varphi(2t) \neq 0$ for all $|t| \leq a$. Repeatedly apply the process to get $\varphi \neq 0$ for all t. □

1.1.2 Canonical representation

Characteristic functions of Infinitely divisible distributions can be uniquely represented in a certain form. Furthermore, if a characteristic function can be written in such form, then it is infinitly divisible. We call it a *canonical form.* While there are several equivalent canonical representations, I would like to cover the one by Kolmogorov. The first theorem is about sufficiency of ID distribution.

Sufficiency

Theorem 2 (Kolmogorov's canonical representation i)**.** *Let φ be a characteristic function. If*

$$\varphi(t) = \exp\left\{\int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} d\mu(x)\right\}, \ \forall t$$

for some finite measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then φ is infinitely divisible with mean 0 and variance $\mu(\mathbb{R})$.

Proof. (case 1. μ has a mass only at 0.)
Let $\sigma^2 = \mu(\mathbb{R}) = \mu\{0\} > 0$ then $\varphi(t) = e^{\frac{t^2\sigma^2}{2}}$ which is the ch.f. of $\mathcal{N}(0, \sigma^2)$ so it is ID.
(case 2. μ has a mass only at $x \neq 0$.)
Let $\mu\{x\} = \lambda x^2$ for some $\lambda > 0$. Then $\varphi(t) = e^{\lambda(e^{itx}-1-itx)}$ which is a ch.f. of $x(Z_\lambda - \lambda)$ where $Z \sim \mathcal{P}(\lambda)$. Let $X_{nk} \overset{\text{iid}}{\sim} \mathcal{P}(\frac{\lambda}{n})$ for $1 \leq k \leq n$. $x(Z_\lambda - \lambda) \overset{d}{=} x\sum_{k=1}^{n}(X_{nk} - \frac{\lambda}{n})$ so it is ID.
(case 3. μ has masses at $x_1, \cdots, x_k$.)
Let $\mu\{x_i\} = \delta_i > 0$ and $\varphi_i(t) = \exp\{\int_{-\infty}^{\infty} \frac{e^{itx}-1-itx}{x^2} d\mu_i(x)\}$ where $\mu_i(\mathbb{R}) = \delta_i$. By case 2, φ_i is ID with mean 0 and variance δ_i. Thus for all n, there exists ch.f.s φ_{jn} such that $\varphi_n = (\varphi_{jn})^n$. It follows that $\varphi = \prod_{j=1}^{k} \varphi_j = \left(\prod_{j=1}^{k} \varphi_{jn}\right)^n$ thus φ is ID. Let $X \sim \varphi$ and $X_i \overset{\text{indep}}{\sim} \varphi_i$ then $X \overset{d}{=} X_1 + \cdots + X_k$ so $EX = 0, \mathrm{Var}(X) = \mu\{x_1, \cdots, x_k\}$.
(case 4. general finite μ.)
Let $\mu_k\{j \cdot 2^{-k}\} = \mu(j \cdot 2^{-k}, (j+1)2^{-k}]$, $j \in J_k = \{0, \pm1, \pm2, \cdots, \pm2^{2k}\}$. Then μ_k has masses on $\{j \cdot 2^{-k} : j \in J_k\}$. Since $\mu_k(\mathbb{R}) \to \mu(\mathbb{R}) > 0$ as $k \to \infty$, $\mu_k(\mathbb{R}) > 0$ for all large k.
Now assume that $f : \mathbb{R} \to \mathbb{R}$ is continuous and vanishes at infinity (i.e. $\lim_{|x|\to\infty} f(x) = 0$). Let

$$f_k = \begin{cases} f(j \cdot 2^{-k}) & , \ x \in (j \cdot 2^{-k}, (j+1)2^{-k}] \\ 0 & , \ \text{otherwise} \end{cases}$$

be a step function, then $\int f d\mu_k = \int f_k d\mu$. As $k \to \infty$, $f_k \to f$. Since $|f_k| \leq |f| \leq \sup_x |f(x)| < \infty$, apply BCT and we get $\int f_k d\mu \to \int f d\mu$.
By the case 3, $\varphi_k(t) = \exp\{\int_{-\infty}^{\infty} \frac{e^{itx}-1-itx}{x^2} d\mu_k(x)\}$ is ID. since the integrand is continuous and vanishes at infinity, $\varphi_k \to \varphi$ as $k \to \infty$. Since $\varphi(0) = 1$ and φ is continuous at 0, by continuity theorem φ is a ch.f. for some random variable.
In addition, $EX^2 \leq \liminf_k EX_k^2 < \infty$ for $X \sim \varphi$ and $X_k \sim \varphi_k$. By moment generating property of ch.f., $iEX = \varphi'(0) = 0$ and $-\mathrm{Var}(X) = -\mu(\mathbb{R})$. Let another $\psi_n(t) = \exp\{\int_{-\infty}^{\infty} \frac{e^{itx}-1-itx}{x^2} d\frac{\mu}{n}(x)\}$ then it is a ch.f. Observe that $\varphi = (\psi_n)^n$ so φ is ID. □

In other words,

$$\log \varphi(t) = \int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} d\mu(x).$$

We call the right hand side the *canonical representation of* φ and μ the *canonical measure.* Note that $\frac{|e^{itx}-1-itx|}{x^2} \leq t^2$ so the integral is well-defined. For $x = 0$, we define $\frac{e^{itx}-1-itx}{x^2}\Big|_{x=0} = -\frac{t^2}{2}$ by continuity. Also note that

$$\frac{|e^{itx} - 1 - itx|}{x^2} \leq t^2 \wedge \frac{2|t|}{|x|} \to 0 \text{ as } |x| \to \infty.$$

This follows from error estimation of the second-order Taylor series.

Necessity To show the necessity part for more general class of characteristic functions, we define the condition R.

Definition 2 (condition R). *A rowwise independent triangular array* $(X_{nk})_{k=1}^{r_n}$ *satisfies R if the followings hold.*
(i) $EX_{nk} = 0$, $\sigma_{nk}^2 = EX_{nk}^2 < \infty$, $s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2 > 0$.
(ii) $\sup_n s_n^2 < \infty$.
(iii) $\max_{1\le k\le r_n} \sigma_{nk}^2 \to 0$ *as* $n \to \infty$.

For the proof of the next theorem, we need the following lemma.

Lemma 2. *Let* (μ_n) *be a sequence of finite measures with* $\sup_n \mu_n(\mathbb{R}) < \infty$. *There exists a subsequence* (μ_{nk}) *and a finite measure* μ *such that* $\mu_{nk} \xrightarrow{w} \mu$ *and* $\int hd\mu_{nk} \to \int hd\mu$ *for all* h *that is continuous and vanishes at infinity.*

Theorem 3 (Kolmogorov's canonical representation ii). *Let* F *be the limiting distribution of* $S_n = X_{n1} + \cdots + X_{nr_n}$ *for some* (X_{nk}) *that satisfies R. Then* φ, *the ch.f. of* F, *has a unique canonical representation:*

$$\varphi(t) = exp\left\{\int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} d\mu(x)\right\}.$$

Proof.

$$\begin{aligned}\Bigg|\underbrace{\prod_{k=1}^{r_n} \varphi_{nk}(t)}_{\text{(i)}} - \underbrace{\prod_{k=1}^{r_n} e^{\varphi_{nk}(t)-1}}_{\text{(ii)}}\Bigg| &\le \sum_{k=1}^{r_n} |\varphi_{nk}(t) - e^{\varphi_{nk}(t)-1}| \\ &\le \sum_{k=1}^{r_n} |\varphi_{nk}(t) - 1|^2 \\ &\le \sum_{k=1}^{r_n} (t^2\sigma_{nk}^2)^2 \\ &\le t^4 \max_{1\le k\le r_n} \sigma_{nk}^2 \cdot s_n^2 \to 0.\end{aligned}$$

The first inequality is from 3.4.3, the second is from 3.4.4, the third is from 3.3.19, and $\to 0$ is by condition R. In addition,

$$\text{(i)} \to \varphi \text{ as } n \to \infty$$

also by condition R.

$$\begin{aligned}\text{(ii)} &= \sum_{k=1}^{r_n} \int (e^{itx} - 1) dF_{nk}(x) \\ &= \sum_{k=1}^{r_n} \int \frac{e^{itx} - 1 - itx}{x^2} x^2 dF_{nk}(x) \\ &= \int \frac{e^{itx} - 1 - itx}{x^2} d\left(\sum_{k=1}^{r_n} x^2 F_{nk}(x)\right)\end{aligned}$$

Let $\mu_n(-\infty, x] = \sum_{k=1}^{r_n} \int_{-\infty}^{x} y^2 dF_{nk}(y)$, then

$$\text{(ii)} = \int \frac{e^{itx} - 1 - itx}{x^2} d\mu_n(x)$$

and $\mu_n(\mathbb{R}) = s_n^2$. So $\sup_n \mu_n(\mathbb{R}) < \infty$ and there exists $(\mu_{nj}), \mu$ such that $\mu_{nj} \overset{w}{\to} \mu$ and $\int h d\mu_{nj} \to \int h d\mu$ for all h that is continuous and vanishes at infinity. By the above mentioned fact,

$$\int \frac{e^{itx} - 1 - itx}{x^2} d\mu_{nj}(x) \to \int \frac{e^{itx} - 1 - itx}{x^2} d\mu(x).$$

By convergence of (i) and (ii), the existence part of the proof is done.
For the uniqueness part, we only need to show that such μ is unique. Suppose

$$\int \frac{e^{itx} - 1 - itx}{x^2} d\mu(x) = \int \frac{e^{itx} - 1 - itx}{x^2} d\nu(x), \ \forall t.$$

This implies $\int e^{itx} d\mu(x) = \int e^{itx} d\nu(x)$. Put $t = 0$ to both sides and we get $c := \mu(\mathbb{R}) = \nu(\mathbb{R})$. Dividing both sides with c, μ/c and ν/c becomes probabilty measures with identical ch.f.s and the proof is done. □

2 Martingales

2.1 Conditional Expectation

In this chapter we study convergence of a sequence of random variables with dependency. To be specific, I will cover theory of martingales. The first subsection is about conditional expectation which is essential for defining martingales.

2.1.1 Definition

Definition 3 (conditional expectation)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{F}_0 \subset \mathcal{F}$ be a sub σ-algebra. For a random variable $X \in \mathcal{F}_0$, $E|X| < \infty$, we say Y a version of $E(X|\mathcal{F}_\prime)$, conditional expectation of X given $\mathcal{F}$, if (i) $Y \in \mathcal{F}$ and (ii) $\int_A XdP = \int_A YdP$ for all $A \in \mathcal{F}$.*

The term "versions" means they are almost surely equivalent. So in the following sections, I will just call such Y a conditional expectation instead of a version.

Non-negative random variables We need to know the existence of such Y and if it is unique (in almost sure sense) if exists at all. For a non-negative X, it can be constructed as the Radon-Nikodym derivative.

Definition 4 (absolute continuity)**.** *For measures μ, ν on a measurable space $(\Omega, \mathcal{F})$, we say ν is absolutely continuous to μ and write $\nu \ll \mu$ if $\mu(A) = 0$ implies $\nu(A)$ for all $A \in \mathcal{F}$.*

Theorem 4 (Radon-Nikodym)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and μ, ν be σ-finite measures. If $\nu \ll \mu$, then there exists $f = \frac{d\nu}{d\mu} \in \mathcal{F}$ such that $f \geq 0$ almost everywhere and $\nu(A) = \int_A fd\mu$ for all $A \in \mathcal{F}$. $f = \frac{d\nu}{d\mu}$ is called the Radon-Nikodym derivative of ν with respect to μ.*

Let $Q(A) = \int_A XdP$ for all $A \in \mathcal{F}0$ then Q is a σ-finite measure such that $Q \ll P$. Thus by Radon-Nikodym theorem, there exists $\frac{dQ}{dP} \in \mathcal{F}_0$ such that $\int_A XdP = \int_A \frac{dQ}{dP}dP$ for all $A \in \mathcal{F}_0$. By definition $\frac{dQ}{dP}$ satisfies conditions for being a conditional expectation of X given $\mathcal{F}$.

Notice that for a non-negative random variable, conditional expectation exists even for random variables that are not integrable.

General case For a general X, let Y^+, Y^- be conditional expectations of X^+, X^- respectively. Let $E(X|\mathcal{F}_0) = Y^+ - Y^-$, then clearly $Y \in \mathcal{F}_0)$ and for given $A \in \mathcal{F}_0$,

$$\begin{aligned}
\int_A XdP &= \int_A X^+dP - \int_A X^-dP \\
&= \int_A Y^+dP - \int_A Y^-dP = \int_A YdP.
\end{aligned}$$

Uniqueness Suppose Y, Y' are $E(X|\mathcal{F}0)$ Then $\int_A (Y - Y')dP = 0$ for all $A \in \mathcal{F}_0$. Let $A_1 = Y - Y' \geq 0$ and $A_2 =$

$Y - Y' \le 0$, $A_1, A_2 \in \mathcal{F}_0$.

$$\int_{A_1} (Y - Y')dP = 0 \implies Y - Y' = 0 \text{ on } A_1.$$
$$\int_{A_2} (Y - Y')dP = 0 \implies Y - Y' = 0 \text{ on } A_2.$$

Thus $Y = Y'$ almost surely.

Not only we get $Y = Y'$ a.s. but we can also be sure that for any $X_1, X_2 \in \mathcal{F}$ that satisfy $\int_A X_1 dP = \int_A X_2 dP$ for all $A \in \mathcal{F}$, it always follows $X_1 = X_2$ a.s.

2.1.2 Examples and insight

Think of $\mathcal{F}_0 \subset \mathcal{F}$ as the information we have at our disposal. For $A \in \mathcal{F}_0$, we can interpret it as an event that we know whether A occurred or not. In this sense, $E(X|\mathcal{F}_0)$ is our best guess of X given the information we have.

Theorem 5 (best guess)**.** *Let X be a random variable such that $EX^2 < \infty$. Let $\mathcal{C} = \{Y \in \mathcal{F}_0 : \ EY^2 < \infty\} \subset L^2$. Then*

$$E[X - E(X|\mathcal{F}_0)]^2 = \inf_{Y \in \mathcal{C}} E(X - Y)^2.$$

The proof requires a property yet to be mentioned, so I will leave it until the end of the section. The following examples will help getting a grasp of the intuition behind conditional expectations. Proofs are clear so I will not mention it.

Proposition 1 (perfect information)**.**

$$X \in \mathcal{F}_0 \implies E(X|\mathcal{F}_0) = X \ \ a.s.$$

Proposition 2 (no information)**.**

$$X \perp \mathcal{F}_0 \implies E(X|\mathcal{F}_0) = EX \ \ a.s.$$

Here $X \perp \mathcal{F}_0$ means

$$P((X \in B) \cap A) = P(X \in B)P(A), \ \forall B \in \mathcal{B}(\mathbb{R}), A \in \mathcal{F}_0.$$

As an extension of undergraduate definition, we can define conditional probability.

Proposition 3 (conditional probability)**.** *(i) For $(\Omega, \mathcal{F}, P)$, suppose $\Omega = \cup_{i=1}^{\infty} \Omega_i$, where Ω_i's are disjoint and $P(\Omega_i) > 0$ for all i. Let $\mathcal{F}_0 = \sigma(\Omega_1, \Omega_2, \cdots)$, then*

$$E(X|\mathcal{F}_0) = \sum_{i=1}^{\infty} \frac{\int_{\Omega_i} X dP}{P(\Omega_i)} \mathbf{1}_{\Omega_i}.$$

i.e.

$$E(X|\mathcal{F}_0) = \frac{\int_{\Omega_i} X dP}{P(\Omega_i)} \ \ on \ \Omega_i.$$

(ii)

$$P(A|\mathcal{F}_0) := E(\mathbf{1}_A|\mathcal{F}_0).$$
$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

(ii) follows naturally from (i).

In undergraduate statistics, instead of giving σ-field, we gave random variables. This can be regarded as a special case of our definition.

Definition 5 (conditional expectation given random variable)**.**

$$E(Y|X) := E(Y|\sigma(X)).$$

Furthermore, we get some form of "conditional density".[1]

Proposition 4 (conditional density)**.** *(i) Suppose* X, Y *have a joint density* $f(x,y)$*. i.e.* $P((X,Y) \in B) = \int_B f(x,y)dxdy$ *for all* $B \in \mathcal{B}(\mathbb{R}^2)$*. If* $E|g(X)| < \infty$*, then*

$$E(g(X)|Y) = h(Y), \text{ where } h(y)\int f(x,y)dx = \int g(x)f(x,y)dx.$$

(ii) $X \perp Y$, $\varphi : \mathbb{R}^2 \to \mathbb{R}$ *is a Borel function such that* $E|\varphi(X,Y)| < \infty$*, then*

$$E(\varphi(X,Y)|X) = h(X), \text{ where } h(x) = E\varphi(x,Y).$$

Proof. (i) Since f, g are Borel, h is also a Borel function. Let (X,Y) be a random vector on a product space $(\Omega, \mathcal{F}, P)$ of $(\Omega_X, \mathcal{F}_X, P_X)$ and $(\Omega_Y, \mathcal{F}_Y, P_Y)$. Given $A \in \sigma(Y)$, let $B \in \mathcal{B}(\mathbb{R})$ so that $A = Y^{-1}(B)$.

$$\begin{aligned}
\int_A g(X)dP &= \int g(X)\mathbf{1}_A dP \\
&= \int g(X)\mathbf{1}_B(Y)dP \\
&= \int\int g(X)\mathbf{1}_B(Y)dP_X dP_Y \\
&= \int\int g(x)\mathbf{1}B(Y)f(x,y)dxdy \\
&= \int_B\int g(x)f(x,y)dxdy \\
&= \int_B h(y)\int f(x,y)dxdy \\
&= \int_A h(Y)dP.
\end{aligned}$$

The third and the fifth equality is from the Fubini's theorem.

(ii) By the Fubini's theorem, $h \in \sigma(X)$. Given $A \in \sigma(X)$, let $B \in \mathcal{B}(\mathbb{R})$ so that $A = X^{-1}(B)$. Similar to (i), we get

$$\begin{aligned}
\int_A h(X)dP_X &= \int\int \varphi(X,Y)dP_Y\mathbf{1}_B(X)dP_X \\
&= \int \varphi(X,Y)\mathbf{1}_B(X)dP \\
&= \int_A \varphi(X,Y)dP_X
\end{aligned}$$

The second equality is from the Fubini's theorem. □

[1] There is a formal notion of (regular) conditional distribution. The actual conditional distribution is a function defined on a product space of $\mathcal{B}(\mathbb{R})$ and Ω.

2.1.3 Properties

Next I would like to cover fundamental properties of conditional expectations. These will be used throughout this chapter.

Proposition 5. *Suppose* $E|X| < \infty$, $E|Y| < \infty$.
(i) $E(aX + bY|\mathcal{F}_0) = aE(X|\mathcal{F}_0) + bE(Y|\mathcal{F}_0)$.
(ii) $X \geq 0$ *a.s.* $\implies$ $E(X|\mathcal{F}_0) \geq 0$ *a.s.*

Notable result from (ii) is that $|E(X|\mathcal{F}_0)| \leq E(|X||\mathcal{F}_0)$.

Inequalities These are conditional version of some of the inequalities that we covered earlier in chapter 1.

Theorem 6 (Markov)**.** *Suppose* $E|X| < \infty$, $X \geq 0$.

$$P(X \geq a|\mathcal{F}_0) \leq \frac{1}{a}E(X|\mathcal{F}_0).$$

Proof.

$$P(X \geq a|\mathcal{F}_0) \leq E(\mathbf{1}_{X \geq a}\frac{X}{a}|\mathcal{F}_0) \leq \frac{1}{a}E(X|\mathcal{F}_0).$$

□

Similarly, Chebyshev's inequality also holds for conditional expectation.

Theorem 7 (Jensen)**.** $E|X| < \infty$, $\varphi : \mathbb{R} \to \mathbb{R}$ *is convex,* $E|\varphi(X)| < \infty$. *Then* $E(\varphi(X)|\mathcal{F}_0) \geq \varphi(E(X|\mathcal{F}_0))$.

Proof. Note that $\varphi(x) = \sup\{ax + b : \ (a, b) \in S\}$ where $S = \{(a, b) : \ ax + b \leq \varphi(x), \forall x\}$. So $\varphi(X) \geq aX + b$ for all $(a, b) \in S$.

$$\begin{aligned} E(\varphi(X)|\mathcal{F}_0) &\geq aE(X|\mathcal{F}_0) + b, \ \forall (a, b) \in S. \\ E(\varphi(X)|\mathcal{F}_0) &\geq \sup\{aE(X|\mathcal{F}0) + b : \ (a, b) \in S\} \\ &= \varphi(E(X|\mathcal{F}_0)). \end{aligned}$$

□

Convergence theorems

Theorem 8 (MCT)**.** *If* $X_n \geq 0$ *a.s. and* $X_n \uparrow X$ *a.s. with* $E|X| < \infty$, *then* $E(X_n|\mathcal{F}_0) \uparrow E(X|\mathcal{F}_0)$ *a.s.*

In fact, the condition $E|X| < \infty$ is not required since we can always define conditional expectation for non-negative random variables as the Radon-Nikodym derivative. I wrote the condition only because Durrett did so.

Proof. Note that $E(X_n|\mathcal{F}_0) \le E(X_{n+1}|\mathcal{F}_0) \le E(X|\mathcal{F}_0)$ for all n. Given $A \in \mathcal{F}_0$, by using MCT twice,

$$\begin{aligned}\int_A \lim_n E(X_n|\mathcal{F}_0)dP &= \lim_n \int_A E(X_n|\mathcal{F}_0)dP \\ &= \lim_n \int_A X_n dP \\ &= \int_A X dP \\ &= \int_A E(X|\mathcal{F})dP.\end{aligned}$$

□

Theorem 9 (DCT)**.** *$X_n \to X$ a.s. and $|X_n| \le Y$ for all n where $EY < \infty$. Then $E(X_n|\mathcal{F}_0) \to E(X|\mathcal{F}_0)$ a.s.*

The proof is similar to that of conditional MCT.

Theorem 10 (Fatou)**.** *$X \ge 0$ a.s., Then $E(\liminf_n X_n|\mathcal{F}_0) \le \liminf_n E(X_n|\mathcal{F}_0)$.*

Proof. Given $M > 0$, $X_n \wedge M$ is dominated by M. There exists a subsequence (X_{n_k}) such that $X_{n_k} \to \liminf_n X_n$. By conditional DCT,

$$\begin{aligned}E(\liminf_n X_n \wedge M|\mathcal{F}) &= \lim_k E(X_{n_k} \wedge M|\mathcal{F}_0) \\ &\le \liminf_n E(X_n|\mathcal{F}),\ \forall M > 0.\end{aligned}$$

By conditional MCT, letting $M \uparrow \infty$ gives the result. □

The obvious consequences are

$$B_n \subset B_{n+1} \uparrow B,\ B = \cup_{n=1}^{\infty} B_n \implies P(B_n|\mathcal{F}_0) \uparrow P(B|\mathcal{F}_0).$$

and

$$C_n \in \mathcal{F}_0 \text{ are disjoint } \implies P(\cup_{n=1}^{\infty} C_n|\mathcal{F}_0) = \sum_{n=1}^{\infty} P(C_n|\mathcal{F}_0).$$

Smoothing property

Theorem 11 (smoothing property)**.** *(i) $X \in \mathcal{F}_0$, $E|Y| < \infty$, $E|XY| < \infty$. Then $E(XY|\mathcal{F}_0) = XE(Y|\mathcal{F}_0)$.*
(ii) $\mathcal{F}_1 \subset \mathcal{F}_2$ are sub σ-fields. $E|X| < \infty$. Then

$$\begin{aligned}E[E(X|\mathcal{F}_1)|\mathcal{F}_2] &= E(X|\mathcal{F}_1) \\ \text{and } E[E(X|\mathcal{F}_2)|\mathcal{F}_1] &= E(X|\mathcal{F}_1).\end{aligned}$$

(i) is clear by using the standard machine. (ii) is also clear by the definition of (nested) conditional expectations.

Finishing the section, let me prove the second theorem of this section.

Proof of the best guess.

$$\begin{aligned} E(X-Y)^2 &= E[X - E(X|\mathcal{F}_0) + E(X|\mathcal{F}_0) - Y]^2 \\ &= E[X - E(X|\mathcal{F}_0)]^2 + E[E(X|\mathcal{F}_0) - Y]^2 \\ &\quad + 2E[(E(X|\mathcal{F}_0) - Y)E((X - E(X|\mathcal{F}_0))|\mathcal{F}_0)] \end{aligned}$$

The canceled term in the second equality is by the smoothing property. Thus $E(X|\mathcal{F}_0) = \arg\min_{Y\in\mathcal{C}} E(X-Y)^2$. □

2.2 Martingales

Remaining sections in chapter 4 is about martingales and convergence of it. Regarding martingales, our first topic will be convergence in almost sure sense. Next we will look into convergence in L^p, with $p > 1$ and $p = 1$ separately. In the meantime the theory of optional stopping will be covered.

2.2.1 Martingales

Definition 6 (martingale)**.** *Let $(\mathcal{F}_n)_{n=1}^{\infty}$ be a sequence of sub σ-fields of $\mathcal{F}$, (X_n) be a sequence of random variables with $X_n \in \mathcal{F}_n$, $E|X_n| < \infty$ for all n. $(X_n, \mathcal{F}_n)$ is a martingale if $E(X_{n+1}|\mathcal{F}_n) = X_n$ a.s., a submartingale if $E(X_{n+1}|\mathcal{F}_n) \geq X_n$ a.s., or a supermartingale if $E(X_{n+1}|\mathcal{F}_n) \leq X_n$ a.s.*

We say X_n is adapted to $\mathcal{F}_n$ if $X_n \in \mathcal{F}_n$ for all n. For simplicity instead of denoting $\mathcal{F}_n$ together, we could just say X_n is a (sub/super)martingale if the adapted σ-fields are clear. If X_n is a martingale, $\int_A X_{n+1}dP = \int_A X_n dP$ for all $A \in \mathcal{F}n$, so trivially $EX_{n+1} = EX_n$ for all n. X_n is a martingale if and only if X_n is both a submartingale and a supermartingale. In addition, if X_n is a submartingale, then $-X_n$ is a supermartingale.

The easiest but important examples are random walks and square martingales.

Example 1. *Suppose $\xi_1, \xi_2, \cdots$ are i.i.d. with mean 0 and variance σ^2. Let $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$. Then*
(i) $X_n := \xi_1 + \cdots \xi_n$ is a martingale.
(ii) $X_n := (\xi_1 + \cdots + \xi_n)^2 - n\sigma^2$ is a martingale.

Though we cannot guarantee that functions of martingales are also martingales, we can say for sure that a function of martingale is a submartingale if the function is convex.

Theorem 12 (4.2.6)**.** *For a martingale X_n, if φ is convex and $E|\varphi(X_n)| < \infty$ for all n, then $\varphi(X_n)$ is a submartingale.*

The proof is direct by conditional Jensen's inequality. The obvious corollary is for submartingales.

Corollary 1 (4.2.7)**.** *For a submartingale X_n, if φ is convex, increasing and $E|\varphi(X_n)| < \infty$ for all n, then $\varphi(X_n)$ is a submartingale.*

The following two examples will be useful in the section comes later.

Example 2. *(i) If X_n is a submartingale, then $(X_n - a)^+$ is a submartingale.*
(ii) If X_n is a supermartingale, then $X_n \wedge a$ is a supermartingale.

2.2.2 Martingale convergence theorems

For martingale convergence theorems, we need to define and prove predictable sequences, stopping times, upcrossing inequality and related properties.

Upcrossing inequality

Definition 7 (filtration)**.** *Let $\mathcal{F}_n$ be a sequence of sub σ-fields of $\mathcal{F}$. $\mathcal{F}_n$ is a filtration if $F_n \subset \mathcal{F}_{n+1}$ for all n.*

Definition 8 (predictable sequence)**.** *For a filtration $(\mathcal{F}_n)_{n\geq 0}$, a sequence of random variables H_n is predictable if $H_{n+1} \in \mathcal{F}n$ for all $n \geq 0$.*

Intuitively, consider n as time index. The term "predictable" is from the fact that we knows every information about the behavior of H_{n+1} in the time point n.

We get the result that the sum of submartingale increments weighted by a bounded predictable sequence is also a submartingale.

Theorem 13 (4.2.8)**.** *Let X_n be a submartingale adapted to a filtration $(\mathcal{F}_n)_{n\geq 0}$. Let H_n be a non-negative predictable sequence with $|H_n| \leq M_n$ for some $M_n > 0$ for all n. Then*

$$(H \cdot X)_n := \sum_{m=1}^{n} H_m(X_m - X_{m-1})$$

is a submartingale.

Proof. (i) $E|(H \cdot X)_n| \leq \sum_{m=1}^{n} M_n E(|X_m| + |X_{m-1}|) < \infty$ for all n.
(ii) Clearly, $(H \cdot X)_n \in \mathcal{F}_n$ for all n.
(iii)

$$\begin{aligned} E((H \cdot X)_{n+1}|\mathcal{F}_n) &= (H \cdot X)_n + E(H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n) \\ &= (H \cdot X)_n + H_{n+1}\{E(X_{n+1}|\mathcal{F}_n) - X_n\}. \end{aligned}$$

∎

We already get a glimpse of stopping times while studying coupon collector's problem and renewal theory. They were random variables that specifies the time that an event occurs. Here, we define it formally.

Definition 9 (stopping time)**.** *Let N be a random variable taking values $0, 1, \cdots, \infty$. N is a stopping time if $\{N = n\} \in \mathcal{F}_n$ for all $n = 0, 1, \cdots, \infty$.*

It is highly useful to define a predictable sequence as an indicator function related to stopping times. With such sequence, we can easily derive the following theorem.

Theorem 14 (4.2.9)**.** *Let N be a stopping time, X_n be a submartingale. Then $X_{n\wedge N}$ is a submartingale.*

Proof. Let $H_m = \mathbf{1}_{m\leq N}$ then it is a non-negative bounded predictable sequence since $\{m \leq N\} = \{N \leq m-1\}^c \in \mathcal{F}_{m-1}$. By theorem 4.2.8 $X_{n\wedge N} - X_0$ is a submartingale, so $X_{n\wedge N}$ is also a submartingale. ∎

As an example and a lemma for our main theorem - martingale convergence - I will state and prove the upcrossing inequality.

Theorem 15 (upcrossing inequality)**.** *Let $(X_n, \mathcal{F}_n)_{n\geq 0}$ be a submartingale. For $a < b$, define*

$$\begin{aligned} N_{2k-1} &:= \inf\{m > N_{2k-2} :\ X_m \leq a\}, \\ N_{2k} &:= \inf\{m > N_{2k-1} :\ X_m \geq b\}, \\ N_0 &:= -1, \\ U_n &= \sup\{k :\ N_{2k} \leq n\}. \end{aligned}$$

For a submartingale $(X_n)_{n\geq 0}$,

$$(b-a)EU_n \leq E(X_n - a)^+ - E(X_0 - a)^+.$$

Proof. First we show that N_m's are stopping times. For given n,

$$\{N_1 = n\} = \{X_0 > a, \cdots, X_{n-1} > a, X_n \leq a\} \in \mathcal{F}_n.$$

$$\{N_1 = n\} = \bigcup_{\ell=1}^{n-1} \{N_1 = \ell, X_{\ell+1} < b, \cdots, X_{n-1} < b, X_n \geq b\} \in \mathcal{F}_n.$$

$$\cdots$$

Thus N_m's are stopping times. Next, we define $Y_n = a + (X_n - a)^+$ so that $Y_{N_{2k}} \geq b$ and $Y_{N_{2k-1}} = a$ for all k. Since $x \mapsto a + (x-a)^+$ is increasing and convex, Y_n is also a submartingale.

$$\begin{aligned} (b-a)EU_n &\leq \sum_{k=1}^{U_n} (Y_{N_{2k}} - Y_{N_{2k-1}}) \\ &= \sum_{k=1}^{U_n} \sum_{i\in J_k} (Y_i - Y_{i-1}), \\ &\quad \text{where } J_k = \{N_{2k-1}+1, \cdots, N_{2k}\} \\ &= \sum_{m\in J} (Y_m - Y_{m-1}), \\ &\quad \text{where } J = \cup_{k=1}^{U_n} J_k \\ &\leq \sum_{m=1}^{n} \mathbf{1}_{m\in J}(Y_m - Y_{m-1}). \end{aligned}$$

Let $H_m = \mathbf{1}_{m\in J}$, then since

$$\{m \in J\} = \{N_{2k-1} < m \leq N_{2k} \text{ for some } k\}$$

H_m is a bounded, non-negative predictable sequence. Thus

$$(b-a)U_n \leq (H\cdot Y)_n$$

and the right hand side is a submartingale. Let $K_m = 1 - H_m$ then similarly $(K \cdot Y)_n$ is a submartingale and $E(K\cdot Y)_n \geq 0$. Hence

$$\begin{aligned} E(Y_n - Y_0) &= E(H\cdot Y)_n + E(K\cdot Y)_n \\ &\geq E(H\cdot Y)_n \geq (b-a)EU_n. \end{aligned}$$

□

We call U_n the number of upcrossings. An important fact directly follows from the theorem is $EU_n \le \frac{1}{b-a}(EX_n^+ + |a|)$. This will be the key to prove the martingale convergence.

Martingale convergence theorems We get our first convergence theorem for dependent sequence.

Theorem 16 (submartingale convergence)**.** *For a submartingale X_n, if $\sup_n X_n^+ < \infty$, then there exists $X \in L^1$ such that $X_n \to X$ a.s.*

Proof. Given $a < b$, let $U_n[a, b]$ be the number of upcrossings of $X_1, \cdots, X_n$ over $[a, b]$. By the upcrossing inequality, $EU_n[a, b] \le \frac{EX_n^+ + |a|}{b-a}$. Let $U[a, b] = \lim_n U_n[a, b]$ then

$$EU[a, b] = \lim_n EU_n[a, b] \le \sup_n \frac{EX_n^+ + |a|}{b-a} < \infty.$$

Thus by Markov's inequality, $0 \le U[a, b] \le \infty$ a.s.
Now suppose $\liminf_n X_n < \limsup_n X_n$. Then for some $a < b$, $X_n < a$ and $X_n > b$ infinitely often. Thus

$$\begin{aligned}
P(\liminf_n X_n < \limsup_n X_n) &= P(\liminf_n X_n < a < b < \limsup_n X_n \text{ for some } a, b \in \mathbb{Q}) \\
&\le \sum_{a,b\in\mathbb{Q}} P(\liminf_n X_n < a < b < \limsup_n X_n) \\
&= \sum_{a,b\in\mathbb{Q}} P(U[a, b] = \infty) = 0
\end{aligned}$$

so there exists X such that $X_n \to X$ a.s. We now need to show that such X is integrable. By Fatou's lemma,

$$\begin{aligned}
EX^+ &\le \liminf_n EX_n^+ \le \sup_n EX_n^+ < \infty. \\
EX^- &\le \liminf_n EX_n^- = \liminf_n E(X_n^+ - X_n) \\
&\le \sup_n EX_n^+ - EX_0 < \infty.
\end{aligned}$$

□

As a corollary, we get supermartingale convergence and closability of negative submartingales.

Corollary 2 (supermartingale convergence)**.** *Let $X_n \ge 0$ be a supermartingale. There exists $X \in L^1$ such that $X_n \to X$ a.s. and $EX_n \le EX_0$.*

Corollary 3 (closability)**.** *If $X_n, n = 1, 2, \cdots$ is a negative submartingale, then $X_n, n = 1, 2, \cdots, \infty$ is also a negative submartingale.*

The next example show that even if a martingale converges almost surely, we cannot guarantee L^p convergence. The following sections will be about in which condition does a martingale converges in L^p.

Example 3. *Let $\xi_1, \cdots$ be i.i.d. with $P(\xi_1 = 1) = P(\xi_1 = -1) = \frac{1}{2}$. Let $S_n = \xi_1 + \cdots + \xi_n$, $S_0 = 1$ and $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$, $\mathcal{F}_0 = \{\phi, \Omega\}$ then S_n is a martingale. Let $N = \inf\{n \ge 1 : \; S_n = 0\}$ be a stopping time, then $X_n := S_{n\wedge N} \ge 0$ is also a martingale. $X_n \to 0$ a.s. but $X_n \to 1$ in L^1.*

Proof. By supermartingale convergence, $X_n \to X$ for some $X \in L^1$. Note that on $(N = \infty)$, $X_n = S_n$. By the law of iterated logarithm, $P(\liminf_n S_n = -\infty, \limsup_n S_n = \infty) = 1$. It follows that

$$\begin{aligned} P(N = \infty) &= P(N = \infty, \liminf_n S_n = -\infty, \limsup_n S_n = \infty) \\ &= P(N = \infty, \liminf_n X_n = -\infty, \limsup_n X_n = \infty) \\ &\leq P(\liminf_n X_n = -\infty, \limsup_n X_n = \infty) = 0. \end{aligned}$$

and $N < \infty$ a.s. Hence $X = \lim_n S_{n\wedge N} = S_N = 0$ a.s.
However, $E|X_n| = ES_{n\wedge N} = ES_0 = 1$ for all n since X_n is a martingale. □

2.3 Applications of Martingales

For applications of martingales, I would like to cover the case of martingales with bounded increments and the branching process.

2.3.1 Martingales with bounded increments

Before getting to the topic, I would like to state a very useful theorem when constructing a (sub)martingale.

Theorem 17 (Doob's decomposition)**.** *Let (X_n) be a submartingale. There uniquely exists (M_n) and (A_n) where the former is a martingale and the latter is an increasing predictable sequence with $A_0 = 0$.*

The uniqueness in the statement is in almost sure sense.

Proof. Let $A_n = A_{n+1} + (E(X_n|\mathcal{F}_{n-1}) - X_{n-1}$, $A_0 = 0$. It is clear that A_n is an increasing predictable sequence. Let $M_n = X_n - A_n$ accordingly, then it is a martingale.
Now suppose $X_n = M_n + A_n = M'_n + A'_n$. Then $M_n - M'_n = A'_n - A_n \in \mathcal{F}_{n-1}$ and

$$\begin{aligned} &M_n - M'_n \\ &= E(M_n - M'_n|\mathcal{F}_{n-1}) \\ &= M_{n-1} - M'_{n-1}. \end{aligned}$$

Thus $M_n - M'_n = A'_0 - A_0 = 0$ for all n and the uniqueness follows. □

The theorem insists that every submartingales can be decomposed into an increasing sequence and a martingale. The important part is where we constructed A_n. Since $A_0 = 0$,

$$\begin{aligned} A_n &= \sum_{m=1}^{n} (E(X_m|\mathcal{F}_{m-1}) - X_{m-1}) \\ &= \sum_{m=1}^{n} E(X_m - X_{m-1}|\mathcal{F}_{m-1}). \end{aligned}$$

This gives us a form of *conditional increment.* In quite a lot of situations constructing a sequence like this leads to a (sub)martingale with bounded increments.
The main theorem of this subsection is a dichotomy that applies to martingales with bounded increments.

Theorem 18 (4.3.1)**.** *Let (X_n) be a martingale with $|X_{n+1} - X_n| \leq M < \infty$ for all n. Let*

$$C = \{X_n \text{ converges}\},$$
$$D = \{\liminf_n X_n = -\infty, \limsup_n X_n = \infty\}.$$

Then $P(C \cup D) = 1$.

Proof. Without loss of generality, let $X_0 = 0$. For $k > 0$, let $N_k = \inf\{n : X_n \leq -k\}$ be a stopping time so that $X_{n \wedge N_k}$ also be a martingale. If $N_k = \infty$, $X_{n \wedge N_k} = X_n > -k$ for all n. If $N_k < \infty$, $X_{N_k} \leq -k$ and $X_t > -k$ for $t = 1, 2, \cdots, N_k - 1$, thus $X_{N_k} = X_{N_k - 1} + (X_{N_k} - X_{N_k - 1}) \geq -k - M$. Since $X_{n \wedge N_k} + k + m$ is a non-negative martingale, by supermartingale convergence $X_{n \wedge N_k}$ converges a.s.
This implies X_n converges on $\{N_k = \infty\}$. Since $\liminf_n X_n > -\infty$ implies $X_n \geq -k'$ for all but finite n's, for some k' and so $N_{k'+1} = \infty$, we get

$$\{\liminf_n X_n > -\infty\} \subset \bigcup_{k=1}^{\infty} \{N_k = \infty\}.$$

Apply the same to $(-X_n)$ and we get

$$\{\limsup_n X_n < \infty\} \subset \bigcup_{k=1}^{\infty} \{N_k = \infty\}.$$

Hence $D^c \subset C$ and it follows that $P(C \cup D) = 1$. □

As a corollary we get an extension of the second Borel-Cantelli lemma for dependent sequence.

Corollary 4 (the second B-C lemma (2))**.** *Let $(\mathcal{F}_n)_{n \geq 0}$ be a filtration with $\mathcal{F}_0 = \{\phi, \Omega\}$. Suppose $A_n \in \mathcal{F}_n$ for all $n \geq 1$. Then*

$$\{A_n \text{ } i.o.\} = \{\sum_{n=1}^{\infty} P(A_n | \mathcal{F}_{n-1}) = \infty\}.$$

Proof. Let $X_n = \sum_{m=1}^{n} (\mathbf{1}_{A_m} - P(A_m | \mathcal{F}_{m-1}))$, $X_0 = 0$. Then it is easy to check that X_n is a martingale with bounded increment.By the dichotomy, we get C or D almost surely.
On C, in order to make X_n convergent,

$$\sum_n \mathbf{1}_{A_n} = \infty \iff \sum_n P(A_n | \mathcal{F}_{n-1}) = \infty.$$

On D,

$$\sum_n \mathbf{1}_{A_n} \geq \limsup_n X_n = \infty,$$
$$\sum_n P(A_n | \mathcal{F}_{n-1}) \geq \limsup_n (-X_n) = \infty.$$

Thus in any case, the desired result follows. □

Notice that X_n in the proof is in the form of A_n from Doob's decomposition.

2.3.2 Braching process

Definition 10 (branching process). *Let ξ_i^n be i.i.d. non-negative integer-valued random variables. Let*

$$Z_0 = 1,\ Z_{n+1} = \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1} & , \ Z_n \geq 0 \\ 0 & , \ Z_n = 0 \end{cases}$$

and $\mathcal{F}_n = \sigma(\xi_i^m : i \geq 0, 1 \leq m \leq n)$. (Z_n) is called a branching process.

Think of ξ_i^n as the number of offsprings that nth individual produce in ith generation. Z_n naturally be the total number of offsprings in nth generation. By construction, Z_n's are independent.

Lemma 3 (4.3.10). *Let $\mu = E\xi_i^n$, then $(\frac{Z_n}{\mu^n}, \mathcal{F}_n)$ is a martingale.*

Proof. It is clear that $Z_n/\mu^n \in \mathcal{F}_n$ and is integrable for all n.

$$\begin{aligned} E(Z_{n+1}|\mathcal{F}_n) &= E(Z_{n+1}\sum_{k=0}^{\infty} \mathbf{1}_{Z_n=k}|\mathcal{F}_n) \\ &= \sum_{k=0}^{\infty} E(Z_{n+1}\mathbf{1}_{Z_n=k}|\mathcal{F}_n) \\ &= \sum_{k=0}^{\infty} E(\sum_{i=1}^{k} \xi_i^{n+1}\mathbf{1}_{Z_n=k}|\mathcal{F}_n) \\ &= \sum_{k=0}^{\infty} \mathbf{1}_{Z_n=k} k\mu \\ &= \sum_{k=0}^{\infty} \mathbf{1}_{Z_n=k} Z_n\mu = Z_n\mu. \end{aligned}$$

□

Using this, we can confirm our naturale guess that the population will be extinct if the average number of offsprings per individual is below 1.

Theorem 19 (4.3.11). *If $\mu < 1$ then $Z_n = 0$ a.s. for all but finite n's.*

Proof. $P(Z_n > 0) = E\mathbf{1}_{Z_n>0} \leq EZ_n\mathbf{1}_{Z_n>0} = EZ_n$. By the lemma, $E(\frac{Z_n}{\mu^n}) = E(\frac{Z_0}{\mu^0}) = 1$ thus $EZ_n = \mu^n$.

$$\sum_{n=1}^{\infty} P(Z_n > 0) \leq \sum_{n=1}^{\infty} \mu^n < \infty.$$

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

2.4 Convergence in L^p, $p > 1$

In this section we look into the condition that makes a martingale converges in L^p, $p > 1$ in detail. We start by proving *Doob's inequality.* By using this result we prove *martingale inequalities* which will then be used to prove *Doob's L^p maximal inequality.* L^p convergence is direct from them. Lastly, as an extension of Doob's inequality, I will brief a version of optional stopping.

2.4.1 Martingale inequalities

Theorem 20 (Doob's inequality)**.** *Let X_n be a submartingale, N be a stopping time such that $N \leq k$ a.s. Then*

$$EX_0 \leq EX_N \leq EX_k.$$

Proof. (i) Observe that $X_{n\wedge N}$ is also a submartingale. Thus $EX_{0\wedge N} \leq EX_{k\wedge N}$ and we get the first inequality.
(ii) Let $K_n = \mathbf{1}_{N\leq n-1}$ be a non-negative bounded predictable sequence then $(K\dot{X})_n = X_n - X_{n\wedge N}$ is a submatringale. Thus $0 = E(K \cdot X)_0 \leq E(K \cdot X)_k$ which leads to the second inequality. □

This natural result will be the foundation of numerous theorems that will be introduced from now on. For simplicity, I will call stopping times with almost sure upper bound *bounded stopping times.*

Theorem 21 (submartingale inequality)**.** *Let X_n be a submartingale. Define $\bar{X}_n = \max_{0\leq m\leq n} X_m$. For $\lambda > 0$,*

$$\lambda P(\bar{X}_n \geq \lambda) \leq EX_n\mathbf{1}_{\bar{X}_n\geq\lambda}.$$

Proof. Let $A = \{\bar{X}_n \geq \lambda\}$. Let $N = \inf\{m : X_m \geq \lambda\} \wedge n$ be a bounded stopping time. Since $\lambda\mathbf{1}_A \leq X_N\mathbf{1}_A$, $\lambda P(A) \leq EX_N\mathbf{1}_A$.
On A, $EX_N \leq EX_n$ by Doob's inequality. On A^c, $N = n$ a.s. Thus in either case $EX_N\mathbf{1}_A \leq EX_n\mathbf{1}_A$ and we get the result. □

A more comprehensive form might be

$$P(\bar{X}_n \geq \lambda) \leq \frac{1}{\lambda}EX_n\mathbf{1}_{\bar{X}_n\geq\lambda},$$

which can be viewed as a version of inequality that resembles Chebyshev's inequality.
Similarly, we can also derive supermartingale inequality.

Theorem 22 (supermartingale inequality)**.** *Let X_n be a supermartingale. For $\lambda > 0$,*

$$\lambda P(\bar{X}_n \geq \lambda) \leq EX_0 - EX_n\mathbf{1}_{\bar{X}_n<\lambda}.$$

Proof. Let A and N as in the proof of submartingale inequality. The result is direct from

$$EX_0 \geq EX_N = EX_N\mathbf{1}_A + EX_N\mathbf{1}_{A^c}.$$

□

2.4.2 L^p convergence theorem

With the help of submartingale inequality, we get the following theorem.

Theorem 23 (Doob's maximal inequality)**.** *Let X_n be a non-negative submartingale. For $1 < p < \infty$,*

$$E\bar{X}_n^p \leq \left(\frac{p}{p-1}\right)^p EX_n^p.$$

Proof. Let $M > 0$. By properly using Foubini's theorem

$$\begin{aligned}
E(\bar{X}_n \wedge M)^p &= \int_0^\infty P((\bar{X}_n \wedge M)^p \geq t)dt \\
&= \int_0^\infty P(\bar{X}_n \wedge M \geq \lambda)p\lambda^{p-1}d\lambda \\
&= \int_0^M P(\bar{X}_n \geq \lambda)p\lambda^{p-1}d\lambda \\
&\leq \int_0^M \frac{1}{\lambda} EX_n \mathbf{1}_{\bar{X}_n \geq \lambda} p\lambda^{p-1}d\lambda \\
&= \int_0^M \int_\Omega X_n \mathbf{1}_{\bar{X}_n \geq \lambda} dP p\lambda^{p-2}d\lambda \\
&= \frac{p}{p-1} EX_n(\bar{X}_n \wedge M)^{p-1} \\
&\leq \frac{p}{p-1}(EX_n^p)^{1/p}(E(\bar{X}_n \wedge M)^p)^{1/q}
\end{aligned}$$

The first inequality is follows submartingale inequality and the second one is from Holder's inequality. Transposition and applying MCT ($M \uparrow \infty$) leads to the result. □

It is often called L^p *maximal inequality.* Note that we used $\bar{X}_n \wedge M$ in order to prove that the inequality holds even if $E\bar{X}_n$ is not finite. L^p convergence of a martingale is derived from this.

Theorem 24 (L^p convergence)**.** *Let X_n be a martingale with $\sup_n E|X_n|^p < \infty$. For $p > 1$, there exists X such that $X_n \to X$ a.s. and in L^p.*

Proof. By submartingale convergence, there exists $X \in L^1$ such that $X_n \to X$ a.s. By MCT and L^p maximal inequality,

$$\begin{aligned}
E \sup_n |X_n|^p &= \lim_n E \max_{0 \leq m \leq n} |X_m|^p \\
&\leq \lim_n \left(\frac{p}{p-1}\right)^p E|X_n|^p \\
&\leq \left(\frac{p}{p-1}\right)^p \sup_n E|X_n|^p < \infty.
\end{aligned}$$

Thus $|X_n - X|^p \leq (2\sup_n |X_n|^p)$ is integrable and by DCT, the result follows. □

2.4.3 Bounded optional stopping

As a sidenote, I would like to cover the fact that bounded stopping times preserve submartingale properties.

Definition 11. *For a stopping time τ,*

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap (\tau = n) \in \mathcal{F}_n, \forall n\}$$

It is not difficult to check that $\mathcal{F}_\tau$ is a *sigma*-field with $\tau \in \mathcal{F}_\tau$.

Theorem 25 (bounded optional stopping)**.** *Let X_n be a submartingale, σ, τ be stopping times that $0 \leq \sigma \leq \tau \leq k$ a.s. Then $E(X_\tau|\mathcal{F}_\sigma) \geq X_\sigma$ a.s.*

The proof can be done in two different ways. The first proof uses Doob's inequality.

Proof. Since $Y_{n\wedge\tau}$ is a submartingale, by Doob's inequality $EY_\sigma \le EY_\tau$. For given $A \in \mathcal{F}_\sigma$, let

$$N = \begin{cases} \sigma & \text{on } A \\ \tau & \text{on } A^c \end{cases}$$

Then N is a stopping time since

$$(N = n) = ((\sigma = n) \cap A) \cup ((\tau = n) \cap (\sigma \le n) \cap A^c) \in \mathcal{F}_n.$$

Hence

$$EY_N = EY_\sigma \mathbf{1}_A + EY_\tau \mathbf{1}_A^c \le EY_\tau.$$

$$\int_A Y_\sigma dP \le \int_A Y_\tau dP = \int_A E(Y_\tau|\mathcal{F}_\sigma)dP.$$

□

The second approach uses the lemma and inductive process:

Lemma 4.

$$E(X_\tau|\mathcal{F}_\sigma)\mathbf{1}_{\sigma=n} = E(X_\tau|\mathcal{F}_n)\mathbf{1}_{\sigma=n} \ a.s.$$

Proof. We first show that the right hand side is $\mathcal{F}_\sigma$-measurable. Given $a \in \mathbb{R}$ and $k \ge 0$,

$$\begin{aligned}&(E(X_\tau|\mathcal{F}_n)\mathbf{1}_{\sigma=n} \le a) \cap (\sigma = k) \\ &= \begin{cases} (E(X_\tau|\mathcal{F}n) \le a) \cap (\sigma = k) \in \mathcal{F}_k & , \ k = n \\ (0 \le a) \cap (\sigma = k) \in \mathcal{F}_k & , \ \text{otherwise} \end{cases}\end{aligned}$$

Next for given $A \in \mathcal{F}_\sigma$,

$$\begin{aligned}&\int_A E(X_\tau|\mathcal{F}_\sigma)\mathbf{1}_{\sigma=n}dP \\ &= \int_{A\cap(\sigma=n)} E(X_\tau|\mathcal{F}_\sigma)dP \\ &= \int_{A\cap(\sigma=n)} X_\tau dP \\ &= \int_{A\cap(\sigma=n)} E(X_\tau|\mathcal{F}_n)dP \\ &= \int_A E(X_\tau|\mathcal{F}_n)\mathbf{1}_{\sigma=n}dP.\end{aligned}$$

□

Proof of bounded optional stopping. it sufficies to show that for all $A \in \mathcal{F}_n$

$$\int_A E(X_\tau|\mathcal{F}_\sigma)\mathbf{1}_{\sigma=n}dP \ge E(X_\tau|\mathcal{F}_n)\mathbf{1}_{\sigma=n}.$$

Given $A \in \mathcal{F}_n$,

$$
\begin{aligned}
&\int_A E(X_\tau|\mathcal{F}_\sigma)\mathbf{1}_{\sigma=n}dP - E(X_\tau|\mathcal{F}_n)\mathbf{1}_{\sigma=n} \\
&= \int_{A\cap(\sigma=n)} E(X_\tau|\mathcal{F}_n) - X_n dP \\
&= \int_{A\cap(\sigma=n)} X_\tau - X_n dP \\
&= \int_{A\cap(\sigma=n)\cap(\tau\geq n+1)} X_\tau - X_n dP \\
&\geq \int_{A\cap(\sigma=n)\cap(\tau\geq n+1)} X_\tau - X_{n+1} dP \\
&= \int_{A\cap(\sigma=n)\cap(\tau\geq n+2)} X_\tau - X_{n+1} dP \\
&\cdots \\
&\geq \int_{A\cap(\sigma=n)\cap(\tau=k)} X_\tau - X_k dP = 0.
\end{aligned}
$$

□

2.5 Convergence in L^1

In the previous section, we covered the condition where martingales converges in L^p. We only covered the case where $p > 1$. In this section, the notions of uniform integrability is introduced to compensate convergence in $p = 1$ case.

2.5.1 Uniform integrability

If a random variable X is integrable, $\int_{-}\|X\| \geq a\|X\|dP < \epsilon$ for all $\epsilon > 0$ for large a and vice versa. Intuitively, in order for a random variable to be integrable, integration of its tail part should be bounded for any small ϵ. Uniform integrability is defined accordingly.

Definition 12 (uniform integrability). $(X_t)_{t\in T}$ *is uniformly integrable if* $\lim_a \sup_{t\in T} \int_{|X_t|\geq a} |X_t|dP = 0$.

If $X_t \leq X$ for all $t \in T$ where X is integrable, (X_t) is uniformly integrable. If $(X_t), (Y_t)$ are uniformly integrable, then $(X_t + Y_t)$ is uniformly integrable since for given $a > 0$

$$
\begin{aligned}
&\int_{|X_t+Y_t|\geq a} |X_t + Y_t|dP \\
&\leq \int_{|X_t|+|Y_t|\geq a, |X_t|\geq|Y_t|} |X_t| + |Y_t|dP \\
&\quad + \int_{|X_t|+|Y_t|\geq a, |X_t|<|Y_t|} |X_t| + |Y_t|dP \\
&\leq \int_{2|X_t|\geq a} 2|X_t|dP + \int_{2|Y_t|\geq a} 2|Y_t|dP.
\end{aligned}
$$

The next theorem which sometimes is referred to as Vitali's lemma is about necessary and sufficient condition for uniform integrability.

Theorem 26. $(X_t)_{t\in T}$ *is uniformly integrable if and only if the followings hold.*
(i) $\sup_{t\in T} E|X_t| < \infty$.
(ii) $\forall \epsilon > 0, \exists \delta > 0$ *such that* $\sup_{t\in T}\int_A |X_t|dP \le \epsilon$ *for all* $A \in \mathcal{F}$ *where* $P(A) \le \delta$.

Proof. ($\Rightarrow$) (i) is clear. Given $A \in \mathcal{F}$, $a > 0$,

$$\begin{aligned}&\int_A |X_t|dP\\ &= \int_{A\cap\{|X_t|\ge a\}} |X_t|dP + \int_{A\cap\{|X_t|<a\}} |X_t|dP\\ &\le \int_{|X_t|\ge a} |X_t|dP + aP(A)\end{aligned}$$

Thus $\sup_t \int_A |X_t|dP \le \epsilon/2 + a\delta$.
($\Leftarrow$) Let $M = \sup_t E|X_t| < \infty$, $a_0 = M/\delta$. Since $P(|X_t| \ge a_0) \le E|X_t|/a_0 \le M/a_0 = \delta$, $\sup_t \int_{|X_t|\ge a_0} |X_t|dP \le \epsilon$. □

We state our main theorem of this subsection.

Theorem 27 (Vitali)**.** *Suppose* $X_n \to X$, $X_n \in L^p$, $p \ge 1$. *The followings are equivalent.*
(i) $(|X_n|^p)$ *is uniformly integrable.*
(ii) $X_n \to X$ *in* L^p.
(iii) $E|X_n|^p \to E|X|^p < \infty$.

Proof. ((i)$\Rightarrow$(ii)) By Fatou's lemma, $E|X|^p \le \infty$. $|X_n - X|^p \le 2^p(|X_n|^p + |X|^p)$ makes $|X_n - X|^p$ uniformly integrable. By the theorem, given $\epsilon > 0$ there exists $\delta > 0$ such that $\sup_{t\in T}\int_A |X_t|dP \le \epsilon$ for all $A \in \mathcal{F}$ where $P(A) \le \delta$. There exists N such that for all $n \ge N$, $P(|X_n - X|^p \ge \epsilon) \le \delta$. Thus

$$E|X_n - X|^p = E|X_n - X|^p \mathbf{1}_{|X_n - X|^p \ge \epsilon} + E|X_n - X|^p \mathbf{1}_{|X_n - X|^p < \epsilon} \le 2\epsilon.$$

((ii)$\Rightarrow$(iii)) Trivial by $|\|X_n\|_p - \|X\|_p| \le \|X_n - X\|_p$.
((iii)$\Rightarrow$(i)) Given $a \in \mathbb{R}$ such that $P(|X|^p = a) = 0$.
claim: $|X_n|^p \mathbf{1}_{|X_n|^p \le a} \xrightarrow{P} |X|^p \mathbf{1}_{|X|^p \le a}$.
For all $\delta > 0$,

$$\begin{aligned}&P(|\mathbf{1}_{|X_n|^p\le a} - \mathbf{1}_{|X|^p \le a}| > \epsilon)\\ &\le P(|X_n|^p \le a, |X|^p > a) + P(|X_n|^p > a, |X|^p \le a)\\ &\le P(|X_n|^p \le a, |X|^p > a + \delta) + P(|X_n|^p > a, a - \delta < |X|^p \le a)\\ &\quad + P(|X_n|^p > a, |X|^p \le a - \delta) + P(|X_n|^p > a, a - \delta < |X|^p \le a)\\ &\le 2P(||X_n|^p - |X|^p| > \delta) + P(a < |X|^p \le a + \delta) + P(a - \delta < |X|^p \le a)\end{aligned}$$

Thus as $\delta \to 0$,

$$\limsup_n P(|\mathbf{1}_{|X_n|^p\le a} - \mathbf{1}_{|X|^p\le a}| > \epsilon) \le 0 + P(|X|^p = a) = 0.$$

By the claim and since $|X_n|^p \mathbf{1}_{|X_n|^p \le a}$ is bounded by a, $(|X_n|^p \mathbf{1}_{|X_n|^p \le a})$ is uniformly integrable. By (i)$\Rightarrow$(ii)$\Rightarrow$(iii), $E|X_n|^p \mathbf{1}_{|X_n|^p \le a} \to E|X|^p \mathbf{1}_{|X|^p \le a}$. In addition, by the assumption $E|X_n|^p \mathbf{1}_{|X_n|^p > a} \to E|X|^p \mathbf{1}_{|X|^p > a}$. For a given $\epsilon > 0$, there exists $a_0 > 0$ such that $E|X|^p \mathbf{1}_{|X|^p > a_0} < \epsilon/2$ and $P(|X|^p = a_0) = 0$. Pick N such that $\left|E|X_n|^p \mathbf{1}_{|X_n|^p > a_0} - E|X|^p \mathbf{1}_{|X|^p > a_0}\right| < \epsilon/2$ for all $n \le N$. Then for $n \ge N$, $E|X_n|^p \mathbf{1}_{|X_n|^p > a_0} < \epsilon$. For $n < N$, there exists a_1 such that $\max_{n<N} E|X_n|^p \mathbf{1}_{|X_n|^p > a_0} < \epsilon$. □

2.5.2 L^1 convergence of martingales

With uniform integrability we get L^1-convergence of martingales. First we define regular and closable martingale for simplicity of the statement.

Definition 13. *A martingale (X_n) is regular if there exists a random variable $X \in L^1$ such that $X_n = E(X|\mathcal{F}_n)$ a.s. (X_n) is closable if there exists a random variable $X_\infty \in L^1$ such that $X_n \to X_\infty$ a.s. and $E(X_\infty|\mathcal{F}_n) = X_n$ a.s. for all n.*

If X_n is closable, then it is clearly a regular martingale.

Theorem 28 (4.6.7)**.** *Let X_n be a martingale. The followings are equivalent.*
(i) X_n is regular.
(ii) X_n is uniformly integrable.
(iii) X_n converges a.s. and in L^1
(iv) X_n is closable.

Proof. ((i)⇒(ii)) There exists $X \in L^1$ such that $X_n = E(X|\mathcal{F}_n)$ a.s.

$$\int_{|X_n|\geq a} |X_n| dP \leq \int_{|X_n|\geq a} E(|X||\mathcal{F}_n) dP \leq \int_{E(|X||\mathcal{F}_n)\geq a} |X| dP.$$

Since X is integrable, for a given $\epsilon > 0$ there exists $\delta > 0$ such that $\int_A |X| dP < \epsilon$ for all A such that $P(A) \leq \delta$ and there exists $a > 0$ such that $P(E(|X||\mathcal{F}_n) \geq a) \leq \frac{1}{a} E|X| < \delta$.
((ii)⇒(iii)) Uniform integrability implies $\sup_n |X_n| < \infty$ so by submartingale convergence we get convergence in probability. By Vitali's lemma, we get the result.
((iii)⇒(iv)) There exists $X \in L^1$ such that $E|X_n - X| \to 0$ as $n \to \infty$. Then $E|X_n| \to E|X|$ and $\sup_n E|X_n| < \infty$. By submartingale inequality, there exists $X_\infty \in L^1$ such that $X_n \to X_\infty$ a.s. Notice that $X = X_\infty$ a.s. Let $m \geq n$ then

$$\begin{aligned} &E\left|E(X_\infty|\mathcal{F}_n) - X_n\right| \\ &= E\left|E(X_\infty|\mathcal{F}_n) - E(X_m|\mathcal{F}_n)\right| \\ &\leq E\left|E(|X_\infty - X_m||\mathcal{F}_n)\right| \\ &= E|X_\infty - X_m| \to 0 \end{aligned}$$

as $m \to \infty$. Hence $E(X_\infty|\mathcal{F}_n) = X_n$ a.s.
((iv)⇒(i)) Trivial. □

Consider a sequence of conditional expectations $E(X\|\mathcal{F}_n)$ with fixed X. By using the theorem from previous subsection we can determine convergence of this sequence as well.

2.5.3 Levy's theorem

Theorem 29 (Levy's theorem)**.** *Let X be an integrable random variable and $(\mathcal{F}_n)$ be a filtration. then $E(X|\mathcal{F}_n) \to E(X|\mathcal{F}_\infty)$ a.s. where $\mathcal{F}_\infty = \sigma\left(\cup_n \mathcal{F}_n\right)$.*

Proof. Let $X_n = E(X|\mathcal{F}_n)$ then X_n is a closable, thus regular, martingale and there exists X_∞ such that $X_n \to X_\infty$ a.s. It suffices to show that $X_\infty = E(X|\mathcal{F}_\infty)$ a.s. We show this with π-λ theorem. Let $\mathcal{L} = \{A : \int_A X_\infty dP = \int_A X dP\}$ be a λ-system. Then $\cup_n \mathcal{F}n \subset \mathcal{L}$ and $\cup_n \mathcal{F}_n$ is a π-system. By π-λ theorem $\mathcal{F}_\infty \subset \mathcal{L}$ thus $X_\infty = E(X|\mathcal{F}_\infty)$ a.s. □

Similar result holds for a sequence (X_n) uniformly dominated by an integrable random variable.

Theorem 30. *Suppose $X_n \to X$ a.s., $|X_n| \le Z, \forall n$, $E|Z| < \infty$ then $E(X_n|\mathcal{F}_n) \to E(X|\mathcal{F}_\infty)$ a.s.*

Proof. Let $W_n = \sup_{k,l \ge n} |X_k - X_l|$ then $W_n \downarrow 0$ a.s. and $|X_n - X| \le W_m$ for all $m \le n$ and $W_n \le 2Z$. By the previous theorem we only need to show $E(|X_n - X|\,|\mathcal{F}_n) \to 0$ a.s. Given m,

$$\limsup_n E(|X_n - X|\,|\mathcal{F}_n) \le \lim_n E(W_m|\mathcal{F}_n) = E(W_m|\mathcal{F}_\infty).$$

By conditional DCT, $E(W_m|\mathcal{F}_\infty) \to 0$ a.s. as $m \to \infty$. Thus $E(|X_n - X|\,|\mathcal{F}_n) \to 0$ a.s. With Levy's theorem and triangle inequality the desired result follows. □

2.5.4 Riez's decomposition

We know that any submartingales can be decomposed into a martingale and a predictable sequence (Doob's decomposition). Riez's decomposition allows us to do the similar to uniformly integrable non-negative supermartingales.

Definition 14 (potential)**.** *A supermartingale (X_n) is a potential if it is non-negative and $EX_n \to 0$ a.s.*

Two notable properties of potentials is that (i) $X_n \to 0$ a.s. and (ii) (X_n) is uniformly integrable. (i) is from supermartingale convergence and Fatou's lemma. (ii) follows from $E\|X_n\| \le \epsilon$ for large n for all $\epsilon > 0$.

Theorem 31 (Riez)**.** *For a non-negative uniformly integrable supermartingale (X_n), there uniquely exist a uniformly integrable martingale (M_n) and a potential (V_n) so that $X_n = M_n + V_n$.*

Proof. By supermartingale convergence, there exists X_∞ such that $X_n \to X_\infty$ a.s. Let $M_n = E(X_\infty|\mathcal{F}_n)$ be a regular, thus uniformly integrable martingale. It is enough to show that $V_n := X_n - M_n$ is a potential.

$$E(V_{n+1}|\mathcal{F}_n) = E(X_{n+1}|\mathcal{F}_n) - E(M_{n+1}|\mathcal{F}_n) \le X_n - M_n = V_n \text{ a.s.}$$

Thus V_n is a supermartingale.

$$E(X_\infty|\mathcal{F}_n) \le \liminf_m E(X_m|\mathcal{F}_n) \le X_n \text{ a.s.}$$

for all fixed n. Thus $V_n \ge 0$ for all n. Now by Levy's theorem,

$$\lim_n V_n = X_\infty - \lim_n E(X_\infty|\mathcal{F}_n) = 0 \text{ a.s.}$$

Since $(X_n), (M_n)$ are uniformly integrable, (V_n) is also. By Vitali's lemma $EV_n \to E\lim_n V_n = 0$. Thus V_n is a potential.
For the uniqueness part, let $M_n + V_n = M_n' + V_n'$, $M_n = E(\eta_1|\mathcal{F}_n)$ a.s. and $M_n' = E(\eta_2|\mathcal{F}_n)$ a.s.

$$M_n - M_n' = V_n' - V_n = E(\eta_1|\mathcal{F}_n) - E(\eta_2|\mathcal{F}_n) \to 0 \text{ a.s.}$$

since V_n, V_n' are potentials. By Levy's theorem this implies $E(\eta_1|\mathcal{F}_\infty) - E(\eta_2|\mathcal{F}_\infty) = 0$ a.s.

$$\begin{aligned} M_n &= E\left(E(\eta_1|\mathcal{F}_\infty)|\mathcal{F}_n\right) \\ &= E\left(E(\eta_2|\mathcal{F}_\infty)|\mathcal{F}_n\right) \\ &= E(\eta_2|\mathcal{F}_n) = M_n' \text{ a.s.} \end{aligned}$$

Equivalence of V_n, V_n' directly follows. □

2.6 Square Integrable Martingales

In this section, we look into martingales with special property - square integrability. Square integrability gives martingale an upper bound for maximal expectation so that it can further be used to determine the convergence of the sequence.

2.6.1 Square integrable martingales

Definition 15 (square integrable martingale)**.** *A martingale X_n is square integrable if $EX_n^2 < \infty$ for all n.*

In the following discussion, we assume $X_0 = 0$. Notice that X_n^2 is a submartingale and if we let $A_n = A\_n - 1 + E(X_n^2 \| \mathcal{F}_{n-1}) - X\_n - 1^2$, $A_0 = 0$, which is from Doob's decomposition, then $EX_n^2 = EA_n$ and

$$\begin{aligned} A_n &= \sum_{m=1}^{n} \left(E(X_m^2 | \mathcal{F}_{m-1}) - X_{m-1}^2 \right) \\ &= \sum_{m-1}^{n} E\left((X_m - X_{m-1})^2 | \mathcal{F}_{m-1} \right). \end{aligned}$$

Theorem 32. *For a square integrable martingale X_n, let $A_\infty = \lim_n A_n$. The followings hold.*
(i) $E \sup_n X_n^2 \le 4EA_\infty$.
(ii) $E \sup_n |X_n| \le 3EA_\infty^{\frac{1}{2}}$
(iii) $\lim_n X_n$ exists and is almost surely finite on $\{A_\infty < \infty\}$.
(iv) If $f : \mathbb{R} \to \mathbb{R}$ is increasing and $\int_0^\infty f^{-2}(t)dt < \infty$, $f(t) \ge 1, \forall t$, then $\frac{X_n}{f(A_n)} \to 0$ a.s. on $\{A_\infty = \infty\}$.

Proof. (i) is direct from L^p maximal inequality.
(ii) Let $N_a = \inf\{n : A_{n+1} > a^2\}$, then it is a stopping time.

$$\begin{aligned} P(\sup_n |X_n| > a) &= P(\sup_n |X_n| > a, N < \infty) + P(\sup_n |X_n| > a, N = \infty) \\ &\le P(N < \infty) + P(\sup_n |X_{n \wedge N} > a) \\ &= P(N < \infty) + \lim_n P(\sup_{m \le n} |X_{m \wedge N}| > a) \\ &\le P(N < \infty) + \frac{1}{a^2} \lim_n E|X_{n \wedge N}|^2 \\ &= P(N < \infty) + \frac{1}{a^2} \lim_n EA_{n \wedge N} \\ &\le P(N < \infty) + \frac{1}{a^2} E(A_\infty \wedge a^2) \\ &= P(A_\infty > a^2) + \frac{1}{a^2} E(A_\infty \wedge a^2). \end{aligned}$$

The last inequality is from the fact that

$$EA_{n \wedge N} \le EA_N \le a^2 \text{ on } \{N < \infty\}, EA_n \le EA_\infty \le a^2 \text{ on } \{N = \infty\}.$$

Using this, Fubini's theorem and integration by substitution, we get

$$\begin{aligned}
E\sup_n |X_n| &= \int P(\sup_n |X_n| \geq a)da \\
&\leq \int_0^\infty P(A_\infty^{1/2} > a)da + \int_0^\infty \frac{1}{a^2}E(A_\infty \wedge a^2)da \\
&= EA_\infty^{1/2} + \int_0^\infty \frac{1}{a^2}\int_0^{a^2} P(A_\infty > b)dbda \\
&= 3EA_\infty^{1/2}.
\end{aligned}$$

(iii) Given $a > 0$, by (i), $E\sup_n X_{n\wedge N_a} \leq 4a^2 < \infty$. By submartingale convergence, $X_{n\wedge N_a}$ converges a.s. and in L^2. Now let $C_k = \{X_{n\wedge N_k} \text{ converges}\}$, then $P(C_k) = 1$ and $P(\cap_k C_k) = 1$ as well. For an arbitrary $\omega \in (\cap_k C_k) \cap (A_\infty < \infty)$, $N_k(\omega) = \inf\{n : A_n(\omega) \geq k\} = \infty$ for large enough k since $A_\infty(\omega) < \infty$. Hence $X_{n\wedge N_k}(\omega) = X_n(\omega)$ converges.
(iv) Let $H_m = \frac{1}{f(A_m)}$ be a bounded predictable sequence. Then $Y_n := (H\cdot X)_n = \sum\limits_{m=1}^n \frac{X_m - X_{m-1}}{f(A_m)}$ is a square integrable martingale. Let $B_n = \sum\limits_{m=1}^n E\left((Y_m - Y_{m-1})^2|\mathcal{F}_{m-1}\right)$, then

$$\begin{aligned}
B_\infty &= \sum_{m=0}^\infty \frac{A_{m+1} - A_m}{f(A_{m+1})^2} \\
&\leq \sum_{m=0}^\infty \int_{A_m}^{A_{m+1}} f^{-2}(t)dt \\
&\leq \int_0^\infty f^{-2}(t)dt < \infty \text{ a.s.}
\end{aligned}$$

By (iii), $\lim_n Y_n$ exists and is finite almost surely. By Kronecker's lemma, it suffices to show that $f(A_n) \uparrow \infty$. Since $\int_0^\infty f^{-2}(t)dt < \infty$, $\lim_t f(t)$ should be ∞ otherwise it gives contradiction. Since A_n, f is increasing and $f(A_\infty) = \infty$ on $(A_\infty = \infty)$, this is true. □

From the facts, we get another form of conditional Borel-Cantelli lemma.

Theorem 33 (the second B-C lemma (3))**.** *Let $B_n \in \mathcal{F}_n$ for all $n \geq 0$ and $p_n = P(B_n|\mathcal{F}_{n-1}), n \geq 1$. Then*

$$\frac{\sum_{n=1}^\infty \mathbf{1}B_n}{\sum_{n=1}^\infty p_n} \to 1 \text{ a.s. on } \left\{\sum_{n=1}^\infty p_n = \infty\right\}.$$

Proof. Let $X_n = X_{n-1} + \not\Vdash_{B_n} - P(B_n|\mathcal{F}_{n-1})$, $X_0 = 0$ be a square integrable martingale. Then A_n from Doob's decomposition yields $A_m - A_{m-1} = p_m - p_m^2$ and $A_n = \sum\limits_{m=1}^n p_m - p_m^2 \leq \sum\limits_{m=1}^n p_n$.
On $(A_\infty < \infty)$, X_n converges a.s.

$$\frac{X_n}{\sum_{m=1}^n p_m} = \frac{\sum_{m=1}^n \mathbf{1}_{B_m}}{\sum_{m=1}^n p_m} - 1 \to 0 \text{ a.s. on } (\sum_{n=1}^\infty p_n = \infty).$$

On $(A_\infty = \infty)$, let $f(t) = 1 \vee t$ so that such f satisfies conditions in (iv) of the previous theorem. Then $\frac{X_n}{f(A_n)} = \frac{X_n}{A_n \vee 1} \to 0$ a.s. on $(A_\infty = \infty)$. Since $A_n \leq \sum_{m=1}^n p_m$, we get $\frac{X_n}{\sum_{m=1}^n p_m} \to 0$ a.s. on $(A_\infty = \infty)$. □

2.7 Optional Stopping Theorem

In this section, we generalize the bounded version of optional stopping. After that as an example we will cover theorem regarding assymetric random walk.

2.7.1 Optional stopping theorem

Our first theorem will be the extension of theorem 4.2.9.

Theorem 34 (4.8.1)**.** *Let (X_n) be a uniformly integrable submartingale and N be a stopping time. Then $(X_{n\wedge N})$ is a uniformly integrable submartingale.*

Proof. It is shown that $(X_{n\wedge N})$ is a submartingale in theorem 4.2.9. By Vitali's lemma X_n converges almost surely and in L^1 to some X_∞. Since $x \mapsto x^+$ is convex and increasing, $X_n^+, X_{n\wedge N}^+$ are submartingales. Let $\tau = n, \sigma = n \wedge N$ then τ, σ are bounded stopping times. By Doob's inequality, $EX_{n\wedge N}^+ \le EX_n^+$ and

$$\sup_n EX_{n\wedge N}^+ \le \sup_n EX_n^+ \le \sup_n E|X_n| < \infty.$$

By Submartingale convergence, $X_{n\wedge N} \to X_N$ a.s. and $E|X_N| < \infty$.

$$\begin{aligned}
&E|X_{n\wedge N}|\mathbf{1}_{|X_{n\wedge N}|\ge a} \\
&\le E|X_{n\wedge N}|\mathbf{1}_{|X_{n\wedge N}|\ge a, N\le n} + E|X_{n\wedge N}|\mathbf{1}_{|X_{n\wedge N}|\ge a, N>n} \\
&= E|X_N|\mathbf{1}_{|X_N|\ge a} + E|X_n|\mathbf{1}_{|X_n|\ge a}.
\end{aligned}$$

Since both terms on the right-hand side goes to 0 as $a \to \infty$, $X_{n\wedge N}$ is uniformly integrable. □

Next theorem is the unbounded version of Doob's inequality.

Theorem 35 (4.8.3)**.** *Let (X_n) be a uniformly integrable submartingale, N be a stopping time. Then*

$$EX_0 \le EX_N \le EX_\infty$$

where $X_\infty = \lim_n X_n$ a.s.

Proof. By the previous theorem $X_{n\wedge N}$ is a uniformly integrable submartingale. By Doob's inequality

$$EX_0 \le EX_{n\wedge N} \le EX_n.$$

By Vitali's lemma, $EX_n \to EX_\infty$ and

$$\lim_n X_{n\wedge N} = \begin{cases} X_N & , \ N < \infty \\ X_\infty = X_N & , \ N = \infty \end{cases}$$

Thus $X_{n\wedge N} \to X_N$ a.s. with $E|X_N| < \infty$ by Vitali's lemma and the desired result follows. □

Finally we state and prove the main theorem.

Theorem 36 (optional stopping)**.** *Let $L \le M$ be stopping times and $(Y_{n\wedge M})$ be a uniformly integrable submartingale. Then $EY_L \le EY_M$ and $Y_L \le E(Y_M|\mathcal{F}_L)$ a.s.*

Proof. Let $X_n = Y_{n\wedge M}$ then it directly follows that $EY_L \le EY_M$. The rest of the proof is the same as the first proof of bounded stopping theorem. □

Note that we do not need uniform integrability of Y_n. The next theorem guarantees uniform integrability of stopped martingale of submartingale with uniformly bounded conditional increment.

Theorem 37 (4.8.5)**.** *Let X_n be a submartingale with $E\left(|X_{n+1}-X_n|\,|\mathcal{F}_n\right) \le B$ a.s. and N be a stopping time with $EN < \infty$. Then $X_{n\wedge N}$ is uniformly integrable and $EX_0 \le EX_N$.*

Proof.

$$X_{n\wedge N} = X_0 + \sum_{m=1}^{n} (X_m - X_{m-1})\mathbf{1}_{m\le N} |X_{n\wedge N}| \le |X_0| + \sum_{m=1}^{n} |X_m - X_{m-1}|\mathbf{1}_{m\le N}$$

Let Z be the right-hand side of the inequality.

$$\begin{aligned} E|Z| &\le E|X_0| + \sum_m |X_m - X_{m-1}|\mathbf{1}_{m\le N} \\ &\le E|X_0| + \sum_m E\left(\mathbf{1}_{m\le N} E(|X_m - X_{m-1}|\,|\mathcal{F}_{m-1})\right) \\ &\le E|X_0| + B \cdot \sum_m P(m \le N) \\ &= E|X_0| + B \cdot EN < \infty. \end{aligned}$$

Thus Z is integrable and $X_{n\wedge N}$ is uniformly integrable. $EX_0 \le EX_N$ follows directly. □

2.7.2 Assymetric random walk

As an application of optional stopping, we look into properties of assymetric random walk. We define assymetric random walk $S_n = \xi_1 + \cdots + \xi_n$, $S_0 = 0$ where ξ_i's are i.i.d. with $P(\xi_1 = 1) = p$, $P(\xi_1 = -1) = q$, $p + q = 1$. Let
$textVar(\xi_1) = \sigma^2 < \infty$ and $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$ for $n \ge 1$, $\mathcal{F}_0$ be a trivial σ-field. Let $\varphi(x) = (\frac{1-p}{p})^x$.

Theorem 38 (4.8.9)**.** *(a) $0 < p < 1 \implies \varphi(S_n)$ is a martingale.*
(b) $T_x := \inf\{n : S_n = x\}$, $x \in \mathbb{Z}$ is a stopping time and $P(T_a < T_b) = \frac{\varphi(b)-\varphi(0)}{\varphi(b)-\varphi(a)}$ for $a < 0 < b$.
(c) $1/2 < p < 1$ and $a < 0 < b \implies T_b < \infty$ a.s. and $P(T_a < \infty) < 1$.
(d) $1/2 < p < 1 \implies ET_b = \frac{b}{2p-1}$, $b > 0$.

Proof of (b). (b) Let $T_a \wedge T_b$ be a stopping time. By law of iterated logarithm,

$$\limsup_n \frac{S_n - n(p-q)}{\sigma\sqrt{2n\log\log n}} = 1 \text{ a.s.}$$
$$\liminf_n \frac{S_n - n(p-q)}{\sigma\sqrt{2n\log\log n}} = -1 \text{ a.s.}$$

thus $S_n \approx n(p-q) \pm \sigma\sqrt{2n\log\log n}$. If $p > q$, $\lim_n S_n = \infty$ a.s. and $T_b < \infty$ a.s. Similarly T_a or T_b is almost surely finite in any cases, so $N < \infty$ a.s. If $N \ge n$, $a \le S_{n\wedge N} = S_n \le b$. If $N < n$, $S_{n\wedge N} = S_N = a$ or b. $\varphi(S_{n\wedge N})$ is a bounded, thus uniformly integrable and closable martingale. Note that $S_N = a\mathbf{1}_{T_a<T_b} + b\nVdash_{T_a>T_b}$. Also note that $E\varphi(S_N) = 1$ since $1 = E\varphi(S_0) = E\varphi(S_{n\wedge N}) \to E\varphi(S_N)$.

$$\begin{aligned} 1 = E\varphi(S_N) &= \varphi(a)P(T_a < T_b) + \varphi(b)P(T_a > T_b) \\ &= (\varphi(a) - \varphi(b))P(T_a < T_b) + \varphi(b). \end{aligned}$$

Organizing both sides gives the result. □

Proof of (c). Observe that $T_\alpha < T_\beta$ for all $\beta < \alpha < 0$. Thus $\lim_{a\to-\infty} T_a = \infty$.

$$\begin{aligned} P(T_b < \infty) &= \lim_{a\to-\infty} P(T_b < T_a) \\ &= \lim_{a\to-\infty} \left(1 - \frac{\varphi(b) - 1}{\varphi(b) - \varphi(a)}\right) \\ &= \lim_{a\to-\infty} \frac{1 - \varphi(a)}{\varphi(b) - \varphi(a)} = 1. \end{aligned}$$

Similarly, $P(T_a < \infty) = 1/\varphi(a) < 1$. □

Proof of (d). Observe that if $a < 0$, $(\inf_n S_n \le a) = (T_a < \infty)$. Since

$$P(\inf_n S_n \le a) = P(T_a < \infty) = \begin{cases} \left(\frac{1-p}{p}\right)^{-a} & , \ a < 0 \\ 1 & , \ a \ge 0 \end{cases}$$

we get

$$\begin{aligned} E|\inf_n S_n| &= \sum_{a=-\infty}^{\infty} |a| P(\inf_n S_n = a) \\ &= \sum_{a=-\infty}^{\infty} |a| \left(\left(\frac{1-p}{p}\right)^{-a} - \left(\frac{1-p}{p}\right)^{-(a-1)} \right) \\ &= \sum_{a=-\infty}^{\infty} |a| \left(\frac{1-p}{p}\right)^{-a} \left(1 - \frac{1-p}{p}\right) < \infty. \end{aligned}$$

Thus $\inf_n S_n$ is integrable. Let $X_n = S_n - n(p-q)$ then X_n is a martingale. Since $T_b < \infty$ a.s., $X_{n\wedge T_b}$ is also a martingale.

$$\begin{aligned} ES_{n\wedge T_b} &= EX_{n\wedge T_b} + (p-q)E(T_b \wedge n) \\ &= \cancel{EX_0} + (p-q)E(T_b \wedge n). \end{aligned}$$

Note that $\inf_n S_n \le S_{n\wedge T_b} \le b$ and $|S_{n\wedge T_b}| \le |\inf_n S_n| + b$ for all n. By DCT, $ES_{n\wedge T_b} \to ES_{T_b} = b$. By MCT, $E(T_b \wedge n) \uparrow ET_b$. Thus the desired result follows. □