

# Probability Theory II (Fall 2016)

J.P.Kim

Dept. of Statistics

Finally modified at December 5, 2016

# Preface & Disclaimer

This note is a summary of the lecture Probability Theory II (326.516) held at Seoul National University, Fall 2016. Lecturer was S.Y.Lee, and the note was summarized by J.P.Kim, who is a Ph.D student. There are few textbooks and references in this course, which are following.

- *Probability: Theory and Examples, R.Durrett*

Also I referred to following books when I write this note. The list would be updated continuously.

- *Probability and Measures, P.Billingsley, 1995.*
- *Convergence in Probability Measures, P.Billingsley, 1999.*
- *Lecture notes on Financial Mathematics I & II (in course), Gerald Trutnau, 2015.*
- *Lecture notes on Topics in Mathematics I (in course), Gerald Trutnau, 2015.*
- *Lecture notes on Introduction to Stochastic Differential Equations (in course), Gerald Trutnau, 2015.*

If you want to correct typo or mistakes, please contact to: [joonpyokim@snu.ac.kr](mailto:joonpyokim@snu.ac.kr)

## Acknowledgement

I'm very grateful to my colleagues, Hyemin Yeon and Seonghun Cho, for comments and correction of errors. Their suggestions were very helpful to improve this note.

# Chapter 1

## Central Limit Theorems

In this chapter, we prove Central Limit Theorems in various cases, and find sufficient or necessary conditions to CLT be held.

### 1.1 i.i.d. case

Following lemma is very useful in our story.

**Lemma 1.1.1.** *Let  $X$  be a random variable with  $E|X|^n < \infty$  and  $\varphi(t) = Ee^{itX}$  be its characteristic function. Then*

$$\left| \varphi(t) - \sum_{k=0}^n \frac{(it)^k EX^k}{k!} \right| \leq E \min \left( \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right).$$

*Proof.* Note that, by Taylor's theorem, there exists  $\xi$  between 0 and  $x$  such that

$$e^{ix} = \sum_{k=0}^n \frac{(ix)^k}{k!} + \frac{(ix)^{n+1}}{(n+1)!} e^{i\xi},$$

so we can obtain that

$$\left| e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!} \right| \leq \frac{|x|^{n+1}}{(n+1)!}.$$

Similarly, there exists  $\xi'$  between 0 and  $x$  such that

$$e^{ix} = \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} + \frac{(ix)^n}{n!} e^{i\xi'} = \sum_{k=0}^n \frac{(ix)^k}{k!} + \frac{(ix)^n}{n!} e^{i\xi'} - \frac{(ix)^n}{n!},$$

so

$$\left| e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!} \right| \leq \frac{2|x|^n}{n!}$$

holds. Thus, we get

$$\left| e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!} \right| \leq \min \left( \frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!} \right),$$

and put  $tX$  into  $x$  then we get

$$\left| e^{itX} - \sum_{k=0}^n \frac{(itX)^k}{k!} \right| \leq \min \left( \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right).$$

Therefore, by Jensen  $|EX| \leq E|X|$  we get

$$\left| \varphi(t) - \sum_{k=0}^n \frac{(it)^k EX^k}{k!} \right| \leq E \left| e^{itX} - \sum_{k=0}^n \frac{(it)^k X^k}{k!} \right| \leq E \min \left( \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right).$$

□

**Corollary 1.1.2.** *For a random variable such that  $EX = 0$  and  $EX^2 = \sigma^2$ ,*

$$\varphi(t) = 1 - \frac{t^2 \sigma^2}{2} + o(|t|^2)$$

as  $t \approx 0$ .

*Proof.* Note that, if  $E|X|^n < \infty$ , by LDCT,

$$E \min \left( \frac{|t||X|^{n+1}}{(n+1)!}, \frac{2|X|^n}{n!} \right) \xrightarrow{|t| \rightarrow 0} 0$$

holds, so

$$E \min \left( \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right) = o(|t|^n)$$

and hence

$$\varphi(t) = \sum_{k=0}^n \frac{(it)^k EX^k}{k!} + o(|t|^n).$$

Now consider a special case  $n = 2$ , then

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(|t|^2)$$

is obtained, because  $EX = 0$ .

□

**Theorem 1.1.3** (CLT for i.i.d. case). *Let  $X_1, \dots, X_n$  be i.i.d. random variables such that  $EX_1 = 0$  and  $EX_1^2 = \sigma^2 > 0$ . Then, for  $S_n = X_1 + X_2 + \dots + X_n$ ,*

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

*Proof.* Let  $\varphi(t) = Ee^{itX_1}$  be a characteristic function of  $X_1$ . Then characteristic function of  $\frac{S_n}{\sigma\sqrt{n}}$  is

$$\begin{aligned} \varphi_{S_n/\sigma\sqrt{n}}(t) &= Ee^{it\frac{S_n}{\sigma\sqrt{n}}} \\ &= \left[ \varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n \\ &= \left[ 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right) \right]^n \\ &= \left[ 1 - \frac{t^2}{2n} + o(n^{-1}) \right]^n. \end{aligned}$$

Note that in here  $t$  is fixed, but  $\frac{t}{\sigma\sqrt{n}} \approx 0$ . Also note that, for a sequence  $c_n$  such that  $nc_n \xrightarrow[n \rightarrow \infty]{} c$ ,

$$\lim_{n \rightarrow \infty} (1 + c_n)^n = e^c$$

holds. Therefore,

$$\varphi_{S_n/\sigma\sqrt{n}}(t) = \left[ 1 - \frac{t^2}{2n} + o(n^{-1}) \right]^n \xrightarrow[n \rightarrow \infty]{} e^{-t^2/2},$$

and by Lévy's continuity theorem, we get the conclusion.  $\square$

## 1.2 Double arrays

**Definition 1.2.1** (Lindeberg's condition). *Let  $\{X_{nk} : k = 1, 2, \dots, r_n\}$  be a double array of r.v.'s where  $r_n \rightarrow \infty$  with*

1.  $X_{n1}, X_{n2}, \dots, X_{nr_n}$  are independent.
2.  $EX_{nk} = 0$  for  $k = 1, 2, \dots, r_n$ .
3.  $EX_{nk}^2 < \infty$ .

*Then  $\{X_{nk}\}$  is said to satisfy Lindeberg's condition if*

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{|X_{nk}| \geq \epsilon s_n} X_{nk}^2 d\mathbb{P} = 0 \quad \forall \epsilon > 0$$

where  $s_n^2 = \sigma_{n1}^2 + \cdots + \sigma_{nr_n}^2 = \text{Var}(X_{n1} + \cdots + X_{nr_n})$  and  $\text{Var}(X_{nk}) = \sigma_{nk}^2$ .

**Theorem 1.2.2.** Let  $S_n = X_{n1} + \cdots + X_{nr_n}$ , where notations are those of definition 1.2.1. Then under Lindeberg's condition,

$$\frac{S_n}{s_n} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

**Remark 1.2.3.** Note that 2nd assumption in Lindeberg's condition is just for convenience. Also, this theorem and Lindeberg condition say that tail behavior (when  $|X_{nk}| \geq \epsilon s_n$ ) of random variables are important for central convergence. If the distribution of r.v.'s has heavy tail and so  $X_{nk}$  can have extreme values, summation may not cancel out extreme effects.

*Proof.* WLOG we assume  $s_n^2 = 1$ . Put  $\varphi_n(t) = Ee^{itS_n}$  and  $\varphi_{nk}(t) = Ee^{itX_{nk}}$ , then

$$\varphi_n(t) = \prod_{k=1}^{r_n} \varphi_{nk}(t)$$

holds. Now our goal is to show that:

**Claim.**  $\varphi_n(t) \rightarrow e^{-t^2/2}$

Note that for two sequences  $w_i$  and  $z_i$  of complex numbers, if  $|w_i|, |z_i| \leq 1$ , then

$$\left| \prod_{i=1}^m w_i - \prod_{i=1}^m z_i \right| \leq \sum_{i=1}^m |w_i - z_i|$$

by induction on  $m$ . Thus,

$$\begin{aligned} |\varphi_n(t) - e^{-t^2/2}| &\stackrel{s_n^2=1}{=} \left| \varphi_n(t) - e^{-\frac{t^2}{2} \sum_{k=1}^{r_n} \sigma_{nk}^2} \right| \\ &= \left| \prod_{k=1}^{r_n} \varphi_{nk}(t) - \prod_{k=1}^{r_n} e^{-\frac{t^2}{2} \sigma_{nk}^2} \right| \\ &\leq \sum_{k=1}^{r_n} \left| \varphi_{nk}(t) - e^{-\frac{t^2}{2} \sigma_{nk}^2} \right| \\ &\leq \underbrace{\sum_{k=1}^{r_n} \left| \varphi_{nk}(t) - \left(1 - \frac{t^2}{2} \sigma_{nk}^2\right) \right|}_{=: A_n} + \underbrace{\sum_{k=1}^{r_n} \left| 1 - \frac{t^2}{2} \sigma_{nk}^2 - e^{-\frac{t^2}{2} \sigma_{nk}^2} \right|}_{=: B_n} \end{aligned}$$

holds. Now by lemma 1.1.1,

$$\left| \varphi_{nk}(t) - \left(1 - \frac{t^2}{2} \sigma_{nk}^2\right) \right| \leq E \min(|tX_{nk}|^3, |tX_{nk}|^2)$$

holds, so

$$\begin{aligned}
A_n &\leq \sum_{k=1}^{r_n} E \min(|tX_{nk}|^3, |tX_{nk}|^2) \\
&= \sum_{k=1}^{r_n} \int \min(|tX_{nk}|^3, |tX_{nk}|^2) d\mathbb{P} \\
&\stackrel{(*)}{\leq} \sum_{k=1}^{r_n} \int_{|X_{nk}| < \epsilon} |tX_{nk}|^3 d\mathbb{P} + \sum_{k=1}^{r_n} \int_{|X_{nk}| \geq \epsilon} |tX_{nk}|^2 d\mathbb{P} \\
&\leq \sum_{k=1}^{r_n} \int |t|^3 \epsilon |X_{nk}|^2 d\mathbb{P} + \sum_{k=1}^{r_n} \int_{|X_{nk}| \geq \epsilon s_n} |tX_{nk}|^2 d\mathbb{P} \\
&= \underbrace{\sum_{k=1}^{r_n} |t|^3 \epsilon \sigma_{nk}^2}_{=|t|^3 \epsilon} + \underbrace{\sum_{k=1}^{r_n} \int_{|X_{nk}| \geq \epsilon s_n} |tX_{nk}|^2 d\mathbb{P}}_{\xrightarrow{n \rightarrow \infty} 0 \text{ (Lindeberg)}}
\end{aligned}$$

holds for sufficiently small  $\epsilon > 0$ . Letting  $\epsilon \searrow 0$  we get  $A_n \xrightarrow{n \rightarrow \infty} 0$  (For (\*), see next remark).

Next, note that,

$$\begin{aligned}
\sigma_{nk}^2 &= \int_{|X_{nk}| < \epsilon} X_{nk}^2 d\mathbb{P} + \int_{|X_{nk}| \geq \epsilon} X_{nk}^2 d\mathbb{P} \\
&\leq \epsilon^2 + \int_{|X_{nk}| \geq \epsilon} X_{nk}^2 d\mathbb{P}
\end{aligned}$$

so

$$\max_{1 \leq k \leq r_n} \sigma_{nk}^2 \leq \epsilon^2 + \underbrace{\sum_{k=1}^{r_n} \int_{|X_{nk}| \geq \epsilon} X_{nk}^2 d\mathbb{P}}_{\xrightarrow{n \rightarrow \infty} 0}$$

holds. It implies that,

$$\frac{\max_k \sigma_{nk}^2}{s_n^2} \xrightarrow{n \rightarrow \infty} 0. \tag{1.1}$$

Now note that  $\exists K > 0$  such that  $|e^x - (1+x)| \leq K|x|^2$  if  $|x| \leq 1$  (For this, see next remark).

Thus

$$\begin{aligned}
B_n &\leq K \sum_{k=1}^{r_n} \left( \frac{t^2}{2} \sigma_{nk}^2 \right)^2 \\
&= K \cdot \frac{t^4}{4} \sum_{k=1}^{r_n} \sigma_{nk}^4 \\
&\leq K \cdot \frac{t^4}{4} \max_{1 \leq k' \leq r_n} \sigma_{nk'}^2 \sum_{k=1}^{r_n} \sigma_{nk}^2
\end{aligned}$$

$$= K \cdot \frac{t^4}{4} \max_{1 \leq k' \leq r_n} \sigma_{nk'}^2 \xrightarrow{n \rightarrow \infty} 0$$

holds, and it implies the conclusion.  $\square$

**Remark 1.2.4.**

- (a) In (\*), following fact is used. Note that  $\min(|x|^3, |x|^2) = |x|^3$  if  $|x| < 1$ , and  $= |x|^2$  otherwise. Thus if  $\epsilon < 1/t$ , we get

$$|tx|^3 I(|x| < \epsilon) + |tx|^2 I(|x| \geq \epsilon) \geq \min(|tx|^3, |tx|^2).$$

For this, see figure 1.1.

- (b) Note that  $\frac{|e^x - (1+x)|}{|x^2|}$  converges as  $|x| \rightarrow 0$ , so

$$\left\{ \frac{|e^x - (1+x)|}{|x^2|} : |x| \leq 1 \right\}$$

is a bounded set. Thus there exists  $K > 0$  such that  $|e^x - (1+x)| \leq K|x|^2$  if  $|x| \leq 1$ .

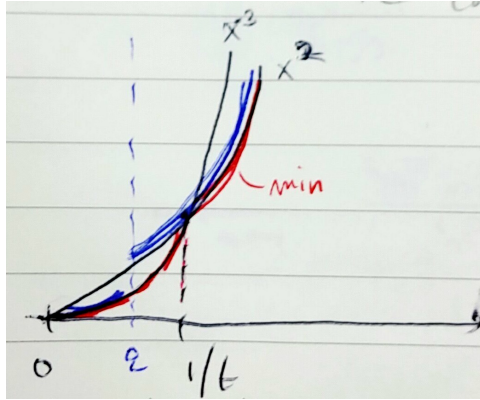


Figure 1.1: The graph of  $\min(|tx|^3, |tx|^2)$ .

**Definition 1.2.5** (Lyapunov's condition). Let  $\{X_{nk}\}$  be a double array such that  $X_{n1}, \dots, X_{nr_n}$  are independent.  $\{X_{nk}\}$  satisfies Lyapunov condition if for some  $\delta > 0$ ,

(a)  $EX_{nk} = 0$

(b)  $E|X_{nk}|^{2+\delta} < \infty$

(c)  $\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{E|X_{nk}|^{2+\delta}}{s_n^{2+\delta}} = 0.$



**Proposition 1.2.6.** *Lyapunov condition implies Lindeberg condition.*

*Proof.*

$$\begin{aligned}
 \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk}| \geq \epsilon s_n} 1 \cdot X_{nk}^2 d\mathbb{P} &\leq \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk}| \geq \epsilon s_n} \left( \frac{|X_{nk}|}{\epsilon s_n} \right)^\delta \cdot X_{nk}^2 d\mathbb{P} \\
 &= \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \int_{|X_{nk}| \geq \epsilon s_n} \frac{|X_{nk}|^{2+\delta}}{\epsilon^\delta} d\mathbb{P} \\
 &\leq \sum_{k=1}^{r_n} \frac{E|X_{nk}|^{2+\delta}}{s_n^{2+\delta}} \frac{1}{\epsilon^\delta} \xrightarrow[n \rightarrow \infty]{\text{Lyapunov}} 0.
 \end{aligned}$$

□

We showed that Lindeberg condition implies CLT. However, next example says that converse does not hold.

**Example 1.2.7.** Let  $\sigma_1^2 > 0$  be a real number and  $\sigma_n^2 = \sigma_1^2 + \cdots + \sigma_{n-1}^2$  for  $n = 2, 3, \dots$ . Let  $X_n \sim N(0, \sigma_n^2)$ , and note that  $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2 = 2\sigma_n^2$ . Then

$$\frac{X_1 + \cdots + X_n}{s_n} \sim N(0, 1)$$

so CLT holds. But for  $Z \sim N(0, 1)$ ,

$$\begin{aligned}
 \frac{1}{s_n^2} \sum_{k=1}^n \int_{|X_k| > \epsilon s_n} X_k^2 d\mathbb{P} &\geq \int_{|X_n| > \epsilon s_n} \left( \frac{X_n}{s_n} \right)^2 d\mathbb{P} \\
 &= \int_{|X_n|/\sigma_n > \sqrt{2}\epsilon} \frac{1}{2} \left( \frac{X_n}{\sigma_n} \right)^2 \\
 &= \frac{1}{2} E[Z^2 I(Z > \sqrt{2}\epsilon)]
 \end{aligned}$$

so Lindeberg condition does not hold.

Now our interest is: what is an equivalent condition for CLT? Fortunately, following Feller's theorem is well known.

**Theorem 1.2.8** (Feller's theorem). *Lindeberg condition  $\Leftrightarrow$  CLT +  $\left[ \frac{\max_{1 \leq k \leq r_n} \sigma_{nk}^2}{s_n^2} \xrightarrow[n \rightarrow \infty]{} 0 \right]$ .*

*Proof.*  $\Rightarrow$  part was already done. To show  $\Leftarrow$  part, WLOG  $s_n^2 = 1$ . By the CLT,

$$\prod_{k=1}^{r_n} \varphi_{nk}(t) \xrightarrow[n \rightarrow \infty]{} e^{-t^2/2}$$

holds, where  $\varphi_{nk}(t) = Ee^{itX_{nk}}$ . Recall that: since  $EX_{nk} = 0$  and  $EX_{nk}^2 = \sigma_{nk}^2$ , by lemma 1.1.1,

$$|\varphi_{nk}(t) - 1| \leq t^2 \sigma_{nk}^2$$

holds, so

$$\max_{1 \leq k \leq r_n} |\varphi_{nk}(t) - 1| \leq \max_{1 \leq k \leq r_n} t^2 \sigma_{nk}^2 \xrightarrow{n \rightarrow \infty} 0$$

is obtained. Meanwhile, note that

$$|e^z - 1 - z| \leq K|z|^2 \quad \forall z \text{ s.t. } |z| \leq 2$$

holds for some  $K$ . Hence, we get

$$\begin{aligned} \sum_{k=1}^{r_n} \left| e^{\varphi_{nk}(t)-1} - 1 + 1 - \varphi_{nk}(t) \right| &\leq K \sum_{k=1}^{r_n} |\varphi_{nk}(t) - 1|^2 \\ &\leq K \max_{1 \leq k \leq r_n} |\varphi_{nk}(t) - 1| \underbrace{\sum_{k'=1}^{r_n} |\varphi_{nk'}(t) - 1|}_{\leq t^2} \\ &\leq K t^4 \max_{1 \leq k \leq r_n} \sigma_{nk}^2 \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Now since  $|e^z| = e^{\operatorname{Re} z} \leq e^{|z|}$ ,

$$\left| e^{\varphi_{nk}(t)-1} \right| \leq e^{-1} e^{|\varphi_{nk}(t)|} < 1$$

holds, so by lemma,

$$\left| e^{\sum_{k=1}^{r_n} (\varphi_{nk}(t)-1)} - \prod_{k=1}^{r_n} \varphi_{nk}(t) \right| \leq \sum_{k=1}^{r_n} \left| e^{\varphi_{nk}(t)-1} - \varphi_{nk}(t) \right| \xrightarrow{n \rightarrow \infty} 0$$

is obtained. Thus by CLT, we get

$$e^{\sum_{k=1}^{r_n} (\varphi_{nk}(t)-1)} \xrightarrow{n \rightarrow \infty} e^{-t^2/2},$$

which implies

$$\left| e^{\sum_{k=1}^{r_n} (\varphi_{nk}(t)-1)} \right| \xrightarrow{n \rightarrow \infty} \left| e^{-t^2/2} \right| = e^{-t^2/2}.$$

Note that

$$|e^z| = \left| e^{\operatorname{Re}(z) + i\operatorname{Im}(z)} \right| = e^{\operatorname{Re}(z)}$$

holds, so it implies that

$$e^{\mathcal{R}e(\sum_{k=1}^{r_n}(\varphi_{nk}(t)-1))} \xrightarrow{n \rightarrow \infty} e^{-t^2/2},$$

and hence

$$\mathcal{R}e \left( \sum_{k=1}^{r_n} (\varphi_{nk}(t) - 1) \right) \xrightarrow{n \rightarrow \infty} -\frac{t^2}{2}$$

holds. Thus,

$$\mathcal{R}e \left( \sum_{k=1}^{r_n} (\varphi_{nk}(t) - 1) \right) + \frac{t^2}{2} = \sum_{k=1}^{r_n} (E \cos tX_{nk} - 1) + \frac{t^2}{2} \xrightarrow{n \rightarrow \infty} 0.$$

Now, since  $EX_{nk}^2 = \sigma_{nk}^2$ , and by our assumption, it is equivalent to

$$\sum_{k=1}^{r_n} E \left( \cos tX_{nk} - 1 + \frac{t^2}{2} X_{nk}^2 \right) \xrightarrow{n \rightarrow \infty} 0.$$

Note that for any real number  $y$ ,  $\cos y - 1 + y^2/2 \geq 0$  holds. Therefore,

$$\begin{aligned} \sum_{k=1}^{r_n} E \underbrace{\left( \cos tX_{nk} - 1 + \frac{t^2}{2} X_{nk}^2 \right)}_{\geq 0} &\geq \sum_{k=1}^{r_n} E \left( \underbrace{\cos tX_{nk} - 1}_{\geq -2} + \frac{t^2}{2} X_{nk}^2 \right) I(|X_{nk}| \geq \epsilon) \\ &\geq \sum_{k=1}^{r_n} E \left( \frac{t^2}{2} X_{nk}^2 I(|X_{nk}| \geq \epsilon) - \underbrace{2I(|X_{nk}| \geq \epsilon)}_{\leq 2X_{nk}^2 \epsilon^{-2} I(|X_{nk}| \geq \epsilon)} \right) \\ &\geq \left( \frac{t^2}{2} - \frac{2}{\epsilon^2} \right) \sum_{k=1}^{r_n} EX_{nk}^2 I(|X_{nk}| \geq \epsilon) \end{aligned}$$

holds for any arbitrarily given  $\epsilon > 0$ . Letting  $t$  such that  $\frac{t^2}{2} - \frac{2}{\epsilon^2} > 0$ , we get

$$\sum_{k=1}^{r_n} EX_{nk}^2 I(|X_{nk}| \geq \epsilon).$$

□

### 1.3 Poisson convergence

**Theorem 1.3.1.** *For each  $n$ ,  $X_{nm}$  are independent r.v.'s with  $P(X_{nm} = 1) = p_{nm}$  and  $P(X_{nm} = 0) = 1 - p_{nm}$ . Assume that*

$$(i) \sum_{m=1}^n p_{nm} \rightarrow \lambda \in (0, \infty)$$

$$(ii) \max_{1 \leq m \leq n} p_{nm} \xrightarrow{n \rightarrow \infty} 0$$

Then  $S_n := X_{n1} + \cdots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} Poi(\lambda)$ .

*Proof.* Let  $\varphi_{nm}(t) = Ee^{itX_{nm}} = (1 - p_{nm}) + p_{nm}e^{it}$ . Then

$$Ee^{itS_n} = \prod_{m=1}^n ((1 - p_{nm}) + p_{nm}e^{it}).$$

Note that

$$\left| e^{p_{nm}(e^{it}-1)} \right| = e^{\operatorname{Re}(p_{nm}(e^{it}-1))} = e^{p_{nm}(\cos t - 1)} \leq 1$$

and

$$\left| (1 - p_{nm}) + p_{nm}e^{it} \right| \leq (1 - p_{nm}) + p_{nm}|e^{it}| = 1,$$

so we get

$$\begin{aligned} \left| e^{\sum_{m=1}^n p_{nm}(e^{it}-1)} - \prod_{m=1}^n ((1 - p_{nm}) + p_{nm}e^{it}) \right| &\leq \sum_{m=1}^n \left| e^{p_{nm}(e^{it}-1)} - ((1 - p_{nm}) + p_{nm}e^{it}) \right| \\ &\stackrel{(*)}{\leq} K \sum_{m=1}^n \left( p_{nm} \underbrace{|e^{it} - 1|}_{\leq 2} \right)^2 \\ &\leq 4K \sum_{m=1}^n p_{nm}^2 \\ &\leq 4K \underbrace{\max_{1 \leq m' \leq n} p_{nm'}}_{\xrightarrow{n \rightarrow \infty} 0} \underbrace{\sum_{m=1}^n p_{nm}}_{\xrightarrow{n \rightarrow \infty} \lambda} \\ &\xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

In (\*), we used  $|e^z - 1 - z| \leq K|z|^2$  ( $\because p_{nm}|e^{it} - 1| \leq 2p_{nm} \leq 2$ ). Note that

$$e^{\sum_{m=1}^n p_{nm}(e^{it}-1)} \xrightarrow{n \rightarrow \infty} e^{\lambda(e^{it}-1)} = \varphi_Z(t),$$

where  $\varphi_Z(t)$  is ch.f of  $Poi(\lambda)$ , and therefore

$$Ee^{itS_n} = \prod_{m=1}^n ((1 - p_{nm}) + p_{nm}e^{it}) \xrightarrow{n \rightarrow \infty} \varphi_Z(t),$$

and Lévy continuity theorem ends the proof.  $\square$

**Corollary 1.3.2.** *Let  $X_{nm}$  be independent nonnegative integer valued random variables for  $1 \leq m \leq n$ , with*

$$P(X_{nm} = 1) = p_{nm}, \quad P(X_{nm} \geq 2) = \epsilon_{nm}.$$

*Assume that*

$$(i) \quad \sum_{m=1}^n p_{nm} \rightarrow \lambda \in (0, \infty)$$

$$(ii) \quad \max_{1 \leq m \leq n} p_{nm} \xrightarrow{n \rightarrow \infty} 0$$

$$(iii) \quad \sum_{m=1}^n \epsilon_{nm} \xrightarrow{n \rightarrow \infty} 0$$

*Then  $S_n := X_{n1} + \cdots + X_{nm} \xrightarrow[n \rightarrow \infty]{d} Poi(\lambda)$ .*

*Proof.* Let  $X'_{nm} = I(X_{nm} = 1)$  and  $S'_n = X'_{n1} + \cdots + X'_{nn}$ . Then since  $P(X'_{nm} = 1) = p_{nm}$ , by previous theorem,

$$S'_n \xrightarrow[n \rightarrow \infty]{d} Poi(\lambda)$$

holds. Now, note that

$$\begin{aligned} P(S_n \neq S'_n) &\leq P\left(\bigcup_{m=1}^n (X_{nm} \neq X'_{nm})\right) \\ &\leq \sum_{m=1}^n P(X_{nm} \neq X'_{nm}) \\ &= \sum_{m=1}^n P(X_{nm} \geq 2) \\ &= \sum_{m=1}^n \epsilon_{nm} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

With this, we get

$$P(\underbrace{|S_n - S'_n|}_{\text{integer}} \geq \epsilon) \leq P(S_n \neq S'_n) \xrightarrow{n \rightarrow \infty} 0$$

so  $S_n - S'_n \xrightarrow[n \rightarrow \infty]{P} 0$ . Therefore, the assertion holds. □

## Chapter 2

# Martingales

### 2.1 Hilbert space

Recall that Hilbert space is a “complete inner product space.”

**Definition 2.1.1.** Let  $E$  be a  $\mathbb{C}$ -vector space. Inner product  $\langle \cdot, \cdot \rangle : E \times E \rightarrow \mathbb{C}$  is a function satisfies followings.

$$(i) \quad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$$

$$(ii) \quad \langle \alpha x, y \rangle = \alpha \langle x, y \rangle$$

$$(iii) \quad \langle y, x \rangle = \overline{\langle x, y \rangle}$$

$$(iv) \quad \langle x, x \rangle \geq 0, \quad \langle x, x \rangle \Leftrightarrow x = 0$$

**Definition 2.1.2.** Let  $\|x\| = \sqrt{\langle x, x \rangle}$  be the norm.

**Proposition 2.1.3.** Followings hold.

$$(a) \quad \|x + y\| \leq \|x\| + \|y\|$$

$$(b) \quad |\langle x, y \rangle| \leq \|x\| \cdot \|y\|$$

$$(c) \quad 2\|x\|^2 + 2\|y\|^2 = \|x + y\|^2 + \|x - y\|^2$$

**Theorem 2.1.4** (Projection). Suppose that  $M$  is a closed convex subset of Hilbert space  $E$ . Then  $\forall y \in E, \exists! w \in M$  such that

$$\|y - w\| = d(y, M) := \inf\{\|y - z\| : z \in M\}.$$

We may denote it as  $\mathcal{P}_M y = w$ .

*Proof.* Let  $d := \inf\{\|y - z\| : z \in M\}$ . For  $n \geq 1$ ,  $\exists z_n \in M$  such that

$$d \leq \|y - z_n\| < d + \frac{1}{n}.$$

Then, since

$$2(\|y + z_n\|^2 + \|y - z_n\|^2) = \|2y - z_n - z_m\|^2 + \|z_n - z_m\|^2,$$

we get

$$\begin{aligned} \|z_n - z_m\|^2 &= 2\|y - z_n\|^2 + 2\|y + z_n\|^2 - 4\left\|y - \frac{z_n + z_m}{2}\right\|^2 \\ &\leq 2\|y - z_n\|^2 + 2\|y + z_n\|^2 - 4d^2 \quad (\because M \text{ is convex, and } d \text{ is minimum distance}) \\ &\xrightarrow{m,n \rightarrow \infty} 0 \quad (\because \|y - z_n\|, \|y - z_m\| \rightarrow d) \end{aligned}$$

and hence  $\{z_n\}$  is Cauchy sequence. Since  $M$  is Hilbert,  $\exists w = \lim_n z_n \in M$ , which makes  $\|y - w\| = d$ . For uniqueness, let  $\exists z \in M$  such that  $\|y - z\| = d$ . Then

$$d^2 \leq \left\|y - \frac{z + w}{2}\right\|^2 = 2\left\|\frac{y - z}{2}\right\|^2 + 2\left\|\frac{y - w}{2}\right\|^2 - \left\|\frac{z - w}{2}\right\|^2 = d^2 - \frac{\|z - w\|^2}{4} \leq d^2$$

and therefore we get  $z = w$ . □

**Theorem 2.1.5.** Let  $M \subseteq E$  be a closed subspace. Then  $\forall y \in E$ ,  $\exists! w \in M$  and  $v \in M^\perp$  such that  $y = w + v$ , where  $M^\perp = \{u : \langle u, v \rangle = 0 \ \forall v \in M\}$ .

*Proof.* By previous theorem, there exists  $w \in M$  such that  $\|y - w\| = d(y, M) =: d$ . Let  $z \in M, z \neq 0$ . Then for any  $\lambda \in \mathbb{C}$ ,

$$d^2 \leq \|y - (w + \lambda z)\|^2 = \|(y - w) - \lambda z\|^2$$

holds. Using

$$\|x + y\|^2 = \|x\|^2 + 2\operatorname{Re}\langle x, y \rangle + \|y\|^2,$$

we obtain

$$d^2 \leq \|(y - w) - \lambda z\|^2 = \|y - w\|^2 - 2\operatorname{Re}\bar{\lambda}\langle y - w, z \rangle + |\lambda|^2\|z\|^2$$

and hence

$$2\operatorname{Re}\bar{\lambda}\langle y - w, z \rangle \leq |\lambda|^2\|z\|^2$$

is obtained. Especially take  $\bar{\lambda} = \overline{r\langle y - w, z \rangle}$  for  $r \in \mathbb{R}$ , and then

$$2r|\langle y - w, z \rangle|^2 \leq r^2|\langle y - w, z \rangle|^2\|z\|^2$$

holds, which implies  $\langle y - w, z \rangle = 0$ . (To show this, assume not, and yield contradiction.) Since  $z$  was arbitrary,  $y - w \in M^\perp$ , and then  $y = w + (y - w)$  is the desired decomposition. For uniqueness, let  $y = w + v, w' + v'$  such that  $w, w' \in M$  and  $v, v' \in M^\perp$ . Then

$$w - w' = v' - v$$

holds. Note that  $w - w' \in M$  and  $v' - v \in M^\perp$ , and since  $M \cap M^\perp = \{0\}$ , we obtain  $w = w'$  and  $v = v'$ .  $\square$

## 2.2 Conditional Expectation

Now let's go back to the space of random variables.

**Theorem 2.2.1.** *Let  $\mathcal{L}^2 = \{X : EX^2 < \infty\}$ . Then  $\mathcal{L}^2$  is a Hilbert space with inner product  $\langle X, Y \rangle = EXY$ .*

*Proof.* It's enough to show completeness. First we need a lemma.

**Lemma 2.2.2.** *If  $\{X_n\} \subseteq \mathcal{L}^2$  and  $\|X_n - X_{n+1}\| \leq 2^{-n}$  for any  $n = 1, 2, \dots$ , then  $\exists X \in \mathcal{L}^2$  such that*

$$(1) P(X_n \rightarrow X \text{ as } n \rightarrow \infty) = 1.$$

$$(2) \|X_n - X\| \xrightarrow{n \rightarrow \infty} 0.$$

*Proof of lemma.* Put  $X_0 \equiv 0$ . Note

$$\begin{aligned} E\left(\sum_{j=1}^{\infty} |X_j - X_{j+1}|\right) & \stackrel{\text{MCT}}{=} \sum_{j=1}^{\infty} E|X_{j+1} - X_j| \\ & \leq \sum_{j=1}^{\infty} (E|X_{j+1} - X_j|^2)^{1/2} \\ & \leq \sum_{j=1}^{\infty} 2^{-j} < \infty. \end{aligned}$$



Thus  $\sum_{j=1}^{\infty} |X_{j+1} - X_j| < \infty$  (Note that  $E|X| < \infty \Rightarrow |X| < \infty$  a.s.), and hence  $\sum_{j=1}^{\infty} (X_{j+1} - X_j)$  converges  $P$ -a.s.. Let

$$X := X_1 + \sum_{j=1}^{\infty} (X_{j+1} - X_j) = \sum_{j=0}^{\infty} (X_{j+1} - X_j).$$

Then  $\lim_n X_n = X$   $P$ -a.s. and because

$$\|X\| \leq \sum_{j=0}^{\infty} \|X_{j+1} - X_j\| < \infty$$

we get  $X \in \mathcal{L}^2$ . Therefore

$$\|X_n - X\| = \left\| \sum_{j=n}^{\infty} (X_{j+1} - X_j) \right\| \leq \sum_{j=n}^{\infty} \|X_{j+1} - X_j\| \xrightarrow{n \rightarrow \infty} 0.$$

□ (Lemma)

Now suppose that  $\{X_n\} \subseteq \mathcal{L}^2$  is a Cauchy sequence. Then for any  $\epsilon > 0$  there is  $N(\epsilon)$  such that

$$n, m \geq N(\epsilon) \Rightarrow \|X_n - X_m\| < \epsilon.$$

Put  $k_n = \max(N(2^{-1}), N(2^{-2}), \dots, N(2^{-n})) + 1$ . Then  $k_n \leq k_{n+1}$  for any  $n$ , and  $k_n, k_{n+1} \geq N(2^{-n})$  so

$$\|X_{k_{n+1}} - X_{k_n}\| \leq \frac{1}{2^n}.$$

Thus by lemma, there exists  $X \in \mathcal{L}^2$  such that  $X = \lim_{n \rightarrow \infty} X_{k_n}$ . To show for general  $n$ , note that

$$\|X_n - X\| \leq \underbrace{\|X_n - X_{k_n}\|}_{\rightarrow 0 \text{ (Cauchy)}} + \|X_{k_n} - X\| \xrightarrow{n \rightarrow \infty} 0.$$

□

**Theorem 2.2.3.** *Let  $X \in \mathcal{L}^2$  and let*

$$\mathcal{L}^2(X) = \{h(X) : h : \mathbb{R} \rightarrow \mathbb{R} \text{ is a Borel function and } E[h(X)]^2 < \infty\}.$$

*Then  $\mathcal{L}^2(X)$  is a closed subspace.*

*Proof.* Since subspace is trivial (show  $(\alpha h + \beta \tilde{h})(X) \in \mathcal{L}^2(X)$ ), so closedness is left. Let  $\{h_n(X)\} \subseteq \mathcal{L}^2(X)$  be a convergent sequence. Then since it is Cauchy, there is a subsequence  $\{k_n\}$  such that

$\|h_{k_n}(X) - h_{k_{n+1}}(X)\| \leq 2^{-n}$ , so by previous lemma, there exists  $Y$  such that

$$Y = \lim_{n \rightarrow \infty} h_{k_n}(X).$$

Note that  $\|Y - h_{k_n}(X)\| \xrightarrow{n \rightarrow \infty} 0$ . (“converge” means that  $\|Y - h_n(X)\| \xrightarrow{n \rightarrow \infty} 0$ .) Letting

$$M = \{x : -\infty < \liminf_{n \rightarrow \infty} h_{k_n}(x) = \limsup_{n \rightarrow \infty} h_{k_n}(x) < \infty\}$$

and

$$h(x) := \limsup_{n \rightarrow \infty} h_{k_n}(x) I_M(x),$$

we obtain  $Y = h(X)$   $P$ -a.s.. Therefore  $Y = h(X) \in \mathcal{L}^2(X)$ .  $\square$

Note that since  $\mathcal{L}^2(X)$  is closed subspace (subspace is convex!) of  $\mathcal{L}^2$ , there exists a “projection” of  $Y \in \mathcal{L}^2$  on  $\mathcal{L}^2(X)$ , and if we define

$$E(Y|X) = \mathcal{P}_{\mathcal{L}^2(X)} Y,$$

it will satisfy

$$\|Y - E(Y|X)\| = \inf_{h(X) \in \mathcal{L}^2(X)} \|Y - h(X)\|.$$

Furthermore, since  $Y - E(Y|X)$  is orthogonal to  $h(X)$ ,  $E(Y|X)$  should satisfy

$$E[(Y - E(Y|X))h(X)] = 0 \quad \forall h(X) \in \mathcal{L}^2(X).$$

Also note that such  $E(Y|X)$  is unique by previous theorems.

**Definition 2.2.4** (Temporary definition). *Let  $X, Y \in \mathcal{L}^2$ . Then  $E(Y|X)$  is defined as the only function of  $X$  satisfying*

$$E[(Y - E(Y|X))h(X)] = 0 \quad \forall h(X) \in \mathcal{L}^2(X).$$

**Proposition 2.2.5.** *Followings hold.*

- (a)  $E(c|X) = c$  for a constant  $c$ .
- (b)  $E(\alpha Y + \beta Z|X) = \alpha E(Y|X) + \beta E(Z|X)$ .
- (c) If  $EXY = EXEY$ ,  $E(Y|X) = EY$ .

(d) If  $g$  is bounded,  $E[g(X)Y|X] = g(X)E[Y|X]$ .

(e)  $EE(Y|X) = EY$ .

*Proof.* Trivial from the definition. Note that in (d), to be well-defined,  $g(X)Y$  should be in  $\mathcal{L}^2$ . Verifying this may be difficult for general  $g$ . If  $g$  is bounded, it is easily checked. (e) can be proved with definition, considering the case  $h(X) \equiv 1$ .  $\square$

Note that, in particular we choose  $h(X) = I(X \in A)$  for a Borel set  $A$ , then definition becomes

$$E(YI(X \in A)) = E(E(Y|X)I(X \in A)),$$

i.e.,

$$\int_{(X \in A)} Y d\mathbb{P} = \int_{(X \in A)} E(Y|X) d\mathbb{P}.$$

Note that since  $\sigma(X) = \{(X \in A) : A \in \mathcal{B}(\mathbb{R})\}$ , if  $Z$  is a  $\sigma(X)$ -measurable r.v. such that

$$\int_{(X \in A)} Z d\mathbb{P} = \int_{(X \in A)} E(Y|X) d\mathbb{P},$$

then  $Z = E(Y|X)$   $\mathbb{P}$ -a.s.. (Note that  $\int_B f d\mu = \int_B g d\mu \forall B \Rightarrow f = g$   $\mu$ -a.e.) Thus if we define conditional expectation using this property, we can omit the assumption that  $E(Y|X)$  is in  $\mathcal{L}^2$ . In other words, we can *extend* the definition.

We can also interpret the conditional expectation as Radon-Nikodym derivative.

**Theorem 2.2.6** (Radon-Nikodym theorem). *Let  $(\Omega, \mathcal{F})$  be a measurable space and let  $\mu, \nu$  be  $\sigma$ -finite measures with  $\nu \ll \mu$ . (It means that  $\mu(A) = 0 \Rightarrow \nu(A) = 0$ ) Then there exists a ( $\mu$ -a.e.) nonnegative  $\mathcal{F}$ -measurable function  $f$  such that*

$$\nu(A) = \int_A f d\mu \quad \forall A \in \mathcal{F}$$

and denote it as  $f = \frac{d\nu}{d\mu}$ .  $f$  is called **Radon-Nikodym derivative**.

Now we are ready to define a conditional expectation.

**Theorem 2.2.7.** *Let  $(\Omega, \mathcal{F}_0, P)$  be a probability space and  $\mathcal{F} \subseteq \mathcal{F}_0$  be a sub- $\sigma$ -field. Consider  $X \in \mathcal{L}^1$ . Then there exists a unique r.v.  $Y$  satisfying*

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

(ii) For any  $A \in \mathcal{F}$ ,  $\int_A X dP = \int_A Y dP$ .

*Proof.* (Existence) Let  $X = X^+ - X^-$ . Letting

$$Q^+(A) = \int_A X^+ dP \text{ and } Q^-(A) = \int_A X^- dP$$

for any  $A \in \mathcal{F}$ , by Radon-Nikodym theorem, there are  $\mathcal{F}$ -measurable random variables

$$\frac{dQ^+}{dP} \text{ and } \frac{dQ^-}{dP} \text{ satisfying } Q^+(A) = \int_A \frac{dQ^+}{dP} dP, \quad Q^-(A) = \int_A \frac{dQ^-}{dP} dP \quad \forall A \in \mathcal{F}.$$

Note that

$$\frac{dQ^+}{dP} \text{ and } \frac{dQ^-}{dP} \text{ are integrable because } Q^+(\Omega) = \int_{\Omega} \frac{dQ^+}{dP} dP < \infty \text{ and similar for } \frac{dQ^-}{dP}.$$

Therefore, we get

$$\int_A X dP = \int_A (X^+ - X^-) dP = \int_A \left( \frac{dQ^+}{dP} - \frac{dQ^-}{dP} \right) dP \quad \forall A \in \mathcal{F}.$$

(Uniqueness) If  $Y'$  also satisfies (i) and (ii), then

$$\int_A Y dP = \int_A Y' dP \quad \forall A \in \mathcal{F}.$$

Taking  $A = \{Y - Y' \geq \epsilon\}$  for  $\epsilon > 0$ , and then

$$0 = \int_A (Y - Y') dP \geq \int_A \epsilon dP = \epsilon P(A)$$

holds, hence  $P(A) = 0$ . Since  $\epsilon > 0$  was arbitrary, we get  $Y \leq Y'$   $P$ -a.s., and by symmetry, we get  $Y = Y'$   $P$ -a.s..  $\square$

**Definition 2.2.8.** Such  $Y$  is called a **conditional expectation** of  $X$ , and denoted as  $Y = E(X|\mathcal{F})$ . Also, if  $\mathcal{F} = \sigma(X)$ , we denote

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

for integrable r.v.'s  $X, Y$ .

**Remark 2.2.9.** Note that  $E(X|\mathcal{F})$  is also  $\mathcal{L}^1$ . To show this, letting  $A = (E(X|\mathcal{F}) > 0) \in \mathcal{F}$ ,

we get

$$0 \leq \int_A E(X|\mathcal{F})dP = \int_A XdP \leq \int_A |X|dP$$

and

$$0 \leq \int_{A^c} -E(X|\mathcal{F})dP = \int_{A^c} -XdP \leq \int_{A^c} |X|dP$$

so we have  $E|E(X|\mathcal{F})| \leq E|X|$ .

**Definition 2.2.10.** We define

$$P(A|\mathcal{F}) := E[I_A|\mathcal{F}]$$

for any  $A \in \mathcal{F}_0$ .

**Proposition 2.2.11.** Followings hold. In here,  $X \in \mathcal{L}^1$ . Also, for convenience, I omitted “P-a.s.”

(a)  $E(c|\mathcal{F}) = c$ .

(b) For  $Y \in \mathcal{L}^1$ , and constants  $a, b$ ,  $E(aX + bY|\mathcal{F}) = aE(X|\mathcal{F}) + bE(Y|\mathcal{F})$ .

(c) For Borel function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ , if  $E|\varphi(X)| < \infty$ , then  $E[\varphi(X)|\mathcal{F}] = \varphi(X)$ .

(d) If  $\mathcal{F} = \{\phi, \Omega\}$ , then  $E(X|\mathcal{F}) = EX$ . (“trivial  $\sigma$ -field”)

(e) If  $\Omega = \bigcup_{i=1}^{\infty} \Omega_i$  for  $\Omega_i \cap \Omega_j = \phi \ \forall i \neq j$ , and

$$\mathcal{F} = \sigma(\Omega_i : i \in \mathbb{N}) = \left\{ \bigcup_{i \in I} \Omega_i : I \subseteq \mathbb{N} \right\},$$

then

$$E(X|\mathcal{F}) = \sum_{i=1}^{\infty} \frac{E[XI_{\Omega_i}]}{P(\Omega_i)} I_{\Omega_i}.$$

(f) If  $E|Y| < \infty$  and  $E|XY| < \infty$ , and  $X$  is  $\mathcal{F}$ -mb, then

$$E(XY|\mathcal{F}) = X \cdot E(Y|\mathcal{F}).$$

(g) (Tower property) If  $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_0$ , then

$$E[E(X|\mathcal{F}_1)|\mathcal{F}_2] = E[E(X|\mathcal{F}_2)|\mathcal{F}_1] = E(X|\mathcal{F}_1).$$

Specifically,  $EE(X|\mathcal{F}) = EX$ .

$$(h) \quad |E(X|\mathcal{F})| \leq E[|X||\mathcal{F}]$$

$$(i) \quad (\text{Markov}) \quad P(|X| \geq c|\mathcal{F}) \leq c^{-1}E[|X||\mathcal{F}] \text{ for } c > 0.$$

$$(j) \quad (\text{MCT}) \quad \text{If } X_n \geq 0, \quad X_n \nearrow X, \text{ then } E(X_n|\mathcal{F}) \nearrow E(X|\mathcal{F}).$$

$$(k) \quad (\text{DCT}) \quad \text{If } X_n \xrightarrow[n \rightarrow \infty]{a.s.} X \text{ and } |X_n| \leq Y \text{ for } E|Y| < \infty, \text{ then } E(X_n|\mathcal{F}) \xrightarrow[n \rightarrow \infty]{a.s.} E(X|\mathcal{F}).$$

$$(l) \quad (\text{Continuity}) \quad \text{Let } B_n \nearrow B \text{ be events. Then } P(B_n|\mathcal{F}) \nearrow P(B|\mathcal{F}).$$

$$(m) \quad P\left(\bigcup_{n=1}^{\infty} C_n|\mathcal{F}\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=1}^n C_k|\mathcal{F}\right) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(C_k|\mathcal{F}) \text{ holds. Last equality holds provided that } C_k \text{'s are disjoint.}$$

$$(n) \quad (\text{Jensen}) \quad \text{If } \varphi : \mathbb{R} \rightarrow \mathbb{R} \text{ is a convex function, and } E|\varphi(X)| < \infty, \text{ then } E[\varphi(X)|\mathcal{F}] \leq \varphi(E[X|\mathcal{F}]).$$

*Proof.* (a), (b), (c), (d). By definition.

(e) Note that if  $g$  is  $\mathcal{F}$ -mb function, then  $g = \sum_{i=1}^{\infty} a_i I_{\Omega_i}$  for some  $a_i$ . Then we get

$$E(X|\mathcal{F}) = \sum_{i=1}^{\infty} a_i I_{\Omega_i}.$$

Taking  $\int_{\Omega_i}$  on both sides, we get

$$P(\Omega_i)a_i = \int_{\Omega_i} X dP$$

and the assertion holds.

(f) Standard machine. If  $X = I_B$  for  $B \in \mathcal{F}$ , for any  $A \in \mathcal{F}$ , we get

$$\int_A E(XY|\mathcal{F})dP = \int_A XY dP = \int_{A \cap B} Y dP = \int_{A \cap B} E(Y|\mathcal{F})dP = \int_A X \cdot E(Y|\mathcal{F})dP$$

from  $A \cap B \in \mathcal{F}$ . If  $X$  is simple, i.e.,

$$X = \sum_{i=1}^m a_i I_{B_i} \text{ for } B_i \in \mathcal{F}, \quad a_i \in \mathbb{R},$$

then

$$E(XY|\mathcal{F}) = E\left[\sum_{i=1}^m a_i I_{B_i} Y \middle| \mathcal{F}\right] = \sum_{i=1}^m a_i E(I_{B_i} Y|\mathcal{F}) = \sum_{i=1}^m a_i I_{B_i} E(Y|\mathcal{F}) = X \cdot E(Y|\mathcal{F})$$

holds. If  $X \geq 0$ , there is a sequence of simple r.v.'s such that  $X_n \nearrow X$ , so  $|X_n Y| \leq |XY|$  holds.

Thus by DCT ((k)),

$$E[X_n Y | \mathcal{F}] \xrightarrow[n \rightarrow \infty]{} E[XY | \mathcal{F}],$$

and from  $E[X_n Y | \mathcal{F}] = X_n E[Y | \mathcal{F}] \xrightarrow[n \rightarrow \infty]{} X \cdot E[Y | \mathcal{F}]$ , we get the desired result. Finally, for general  $X$ , decomposition  $X = X^+ - X^-$  gives the conclusion. (For  $X \geq 0$  case, we can also prove it directly. For any  $A \in \mathcal{F}$ , we get

$$\int_A E[XY | \mathcal{F}] dP = \int_A XY dP \stackrel{DCT}{=} \lim_{n \rightarrow \infty} \int_A X_n Y dP = \lim_{n \rightarrow \infty} \int_A E[X_n Y | \mathcal{F}] dP \stackrel{DCT}{=} \int_A \lim_{n \rightarrow \infty} X_n E[Y | \mathcal{F}] dP$$

and hence

$$\int_A E[XY | \mathcal{F}] dP = \int_A X E[Y | \mathcal{F}] dP.$$

(g) First, since  $E[X | \mathcal{F}_1]$  is  $\mathcal{F}_1$ -mb, it is also  $\mathcal{F}_2$ -mb, and hence by (f),  $E[E[X | \mathcal{F}_1] | \mathcal{F}_2] = E[X | \mathcal{F}_1]$ .

Second, for any  $A \in \mathcal{F}_1$ ,

$$\int_A E[X | \mathcal{F}_2] dP \stackrel{A \in \mathcal{F}_2}{=} \int_A X dP \stackrel{A \in \mathcal{F}_1}{=} \int_A E[X | \mathcal{F}_1] dP$$

holds, and therefore  $E[E[X | \mathcal{F}_2] | \mathcal{F}_1] = E[X | \mathcal{F}_1]$ .

(h)  $-|X| \leq X \leq |X|$ .

(i) Clear.

(j) Since  $E(X_n | \mathcal{F})$  is monotone, we can define  $\lim_{n \rightarrow \infty} E(X_n | \mathcal{F})$ . Thus, for any  $A \in \mathcal{F}$ ,

$$\begin{aligned} \int_A \lim_{n \rightarrow \infty} E(X_n | \mathcal{F}) dP &\stackrel{MCT}{=} \lim_{n \rightarrow \infty} \int_A E(X_n | \mathcal{F}) dP \\ &= \lim_{n \rightarrow \infty} \int_A X_n dP \\ &\stackrel{MCT}{=} \int_A \lim_{n \rightarrow \infty} X_n dP \\ &= \int_A X dP = \int_A E(X | \mathcal{F}) dP. \end{aligned}$$

Also,  $\lim_{n \rightarrow \infty} E(X_n | \mathcal{F})$  is  $\mathcal{F}$ -mb.

(k) Let

$$Y_n := \sup_{k \geq n} |X_k - X|.$$

Then  $Y_n$  is monotone,  $Y_n \xrightarrow[n \rightarrow \infty]{a.s.} 0$ , and  $Y_n \leq 2Y$ . Then  $EY_n \xrightarrow[n \rightarrow \infty]{} 0$  by DCT. Note that since

$E(Y_n|\mathcal{F})$  is monotone,  $\exists Z \geq 0$  such that  $E(Y_n|\mathcal{F}) \searrow Z$ . Then by Fatou's lemma,

$$0 \leq EZ \leq \liminf_{n \rightarrow \infty} EE(Y_n|\mathcal{F}) = \liminf_{n \rightarrow \infty} EY_n = 0,$$

and hence

$$|E(X_n|\mathcal{F}) - E(X|\mathcal{F})| \leq E(|X_n - X||\mathcal{F}) \leq E(Y_n|\mathcal{F}) \xrightarrow[n \rightarrow \infty]{} 0.$$

(l) Clear by (k).

(m) Clear by (k) and (l).

(n) Note that

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

where

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

(By definition of  $S$ ,  $\varphi(x) \geq \sup\{ax + b : (a, b) \in S\}$ . Also, for any  $x$ , there is  $a$  and  $b$  such that  $\varphi(x) = ax + b$  and  $\varphi(y) \geq ay + b \ \forall y$ , so because of supremum, we get  $\varphi(x) \leq \sup\{ax + b : (a, b) \in S\}$ .) Therefore, from

$$E(\varphi(X)|\mathcal{F}) \geq a \cdot E(X|\mathcal{F}) + b,$$

we get

$$E(\varphi(X)|\mathcal{F}) \geq \sup_{a, b \in S} a \cdot E(X|\mathcal{F}) + b = \varphi(E(X|\mathcal{F})).$$

**Proposition 2.2.12.** *Let  $X, Y$  be integrable independent random variables with  $E|\varphi(X, Y)| < \infty$ , where  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  is Borel measurable. Also, define*

$$g(x) = E[\varphi(x, Y)].$$

*Then*

$$E[\varphi(X, Y)|X] = g(X).$$

*Proof.* By proof of Fubini theorem,  $g$  is Borel measurable, so  $g(X)$  is  $\sigma(X)$ -mb. Thus we may show

$$\int_A \varphi(X, Y) dP = \int_A g(X) dP \ \forall A \in \sigma(X).$$

Note that for  $A \in \sigma(X)$ ,  $\exists C \in \mathcal{B}$  such that  $A = (X \in C)$ . Also note that from independence,



we get  $P^{(X,Y)} = P^X \otimes P^Y$ . Therefore,

$$\begin{aligned}
 \int_A \varphi(X, Y) dP &= E[\varphi(X, Y) I_C(X)] \\
 &= \int \int \varphi(x, y) I_C(x) P^{(X,Y)}(dxdy) \\
 &= \int \left( \int \varphi(x, y) P^Y(dy) \right) I_C(x) P^X(dx) \quad (\because \text{Fubini}) \\
 &= \int E[\varphi(x, Y)] I_C(x) P^X(dx) \\
 &= \int g(x) I_C(x) P^X(dx) = \int_A g(X) dP.
 \end{aligned}$$

□

Note that conditional expectation can be interpreted as a *projection* in  $\mathcal{L}^2$ . In other words, our definition is coincident to the *temporary* definition in definition 2.2.4.

**Theorem 2.2.13.** Suppose that  $X$  is r.v. with  $EX^2 < \infty$ . Define

$$\mathcal{C} := \{Y : Y \in \mathcal{F} \text{ \& } EY^2 < \infty\}.$$

In here,  $Y \in \mathcal{F}$  means that  $Y$  is  $\mathcal{F}$ -mb. Then,

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

*Proof.* If  $Y \in \mathcal{C}$ ,

$$E(X - Y)^2 = E[(X - E(X|\mathcal{F}))^2] + E[(E(X|\mathcal{F}) - Y)^2] + 2E[(X - E(X|\mathcal{F}))(E(X|\mathcal{F}) - Y)]$$

and

$$\begin{aligned}
 E[(X - E(X|\mathcal{F}))(E(X|\mathcal{F}) - Y)] &= EE[(X - E(X|\mathcal{F}))(E(X|\mathcal{F}) - Y)|\mathcal{F}] \\
 &= E \left[ (E(X|\mathcal{F}) - Y) \underbrace{E[(X - E(X|\mathcal{F}))|\mathcal{F}]}_{=0} \right] = 0
 \end{aligned}$$

ends the proof. □

**Remark 2.2.14.** Note that  $E(X|\mathcal{F})$  is also  $\mathcal{L}^2$ , by Cauchy-Schwarz inequality,

$$[E(X|\mathcal{F})]^2 \leq E[X^2|\mathcal{F}].$$

Thus we can say that

$$E(X|\mathcal{F}) = \arg \min_{Y \in \mathcal{C}} E(X - Y)^2.$$

## 2.3 Martingales and Stopping Times

### 2.3.1 Definitions and Basic Theory

Fix a probability space  $(\Omega, \mathcal{F}, P)$ .

**Definition 2.3.1.** Let  $\{\mathcal{F}_n\}$  be a sequence of sub  $\sigma$ -fields of  $\mathcal{F}$ . Then  $\{\mathcal{F}_n\}_{n=0}^\infty$  is called a **filtration** if  $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \forall n$ .

**Definition 2.3.2.** Let  $\{\mathcal{F}_n\}_n$  be a filtration. A sequence of r.v.  $\{X_n\}_n$  is called  **$\mathcal{F}_n$ -adapted** if  $X_n \in \mathcal{F}_n$  for any  $n$ .

**Definition 2.3.3.** Let  $\{\mathcal{F}_n\}$  be a filtration and  $\{X_n\}$  be  $\mathcal{F}_n$ -adapted integrable r.v.'s. Then  $\{X_n\}$  or  $(X_n, \mathcal{F}_n)$  is called

**martingale** if  $E[X_n|\mathcal{F}_{n-1}] = X_{n-1} \forall n \geq 1$ .

**submartingale** if  $E[X_n|\mathcal{F}_{n-1}] \geq X_{n-1} \forall n \geq 1$ .

**supermartingale** if  $E[X_n|\mathcal{F}_{n-1}] \leq X_{n-1} \forall n \geq 1$ .

**Example 2.3.4.** Let  $\xi_1, \xi_2, \dots \stackrel{i.i.d.}{\sim} (0, \sigma^2)$ , and let

$$\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n), \quad X_n = \xi_1 + \dots + \xi_n = X_{n-1} + \xi_n.$$

Then  $\{\mathcal{F}_n\}$  is filtration  $\{X_n\}$  is  $\mathcal{F}_n$ -adapted, and  $\{X_n\}$  is a martingale.

**Example 2.3.5.** Let  $\eta_1, \eta_2, \dots \stackrel{i.i.d.}{\sim} (0, \sigma^2)$ , and let

$$X_{n+1} = X_n + h_n(X_1, \dots, X_n)\eta_{n+1}, \quad X_1 = \eta_1,$$

where  $h_n : \mathbb{R}^n \rightarrow \mathbb{R}$  is Borel. Assume that  $X_n$ 's are integrable. Then letting  $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$ , we get  $\{X_n\}$  is martingale.

Following is clear by Jensen.

**Proposition 2.3.6.** *Let  $\{\mathcal{F}_n\}$  be a filtration, and  $\{X_n\}$  be  $\mathcal{F}_n$ -adapted integrable random variables.*

- (a) *If  $\{X_n\}$  is a martinagle and  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is a convex function satisfying  $E|\varphi(X_n)| < \infty \forall n$ , then  $\{\varphi(X_n)\}$  is a submartingale.*
- (b) *If  $\{X_n\}$  is a submartingale and  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is an increasing, convex function satisfying  $E|\varphi(X_n)| < \infty \forall n$ , then  $\{\varphi(X_n)\}$  is a submartingale.*
- (c) *If  $\{X_n\}$  is a supermartingale and  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is an increasing, concave function satisfying  $E|\varphi(X_n)| < \infty \forall n$ , then  $\{\varphi(X_n)\}$  is a supermartingale.*

**Remark 2.3.7.** Consequence of previous proposition that we will use frequently is  $\varphi(x) = |x|$ ,  $x^+$ ,  $|x|^p$  ( $p \geq 1$ ),  $|x - a|$ ,  $(x - a)^+$ ,  $\dots$ .

**Definition 2.3.8.** *Let  $\{\mathcal{F}_n\}$  be a filtration. Then  $\{H_n\}$  is called **predictable** if  $H_n \in \mathcal{F}_{n-1} \forall n \geq 1$ . It means that,  $E(H_n | \mathcal{F}_{n-1}) = H_n$ .*

**Definition 2.3.9** (Martingale Transform). *Let  $X_n$  be a  $(\mathcal{F}_n)$ -martingale (sub- or super-), and  $H_n$  be predictable process, i.e.,  $H_n \in \mathcal{F}_{n-1}$ . Then  $\forall n \geq 1$ ,*

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

**Theorem 2.3.10.** *Let  $H_n$  be predictable process, and suppose that each  $H_n$  is bounded. Then*

- (a) *If  $X_n$  is  $(\mathcal{F}_n)$ -martingale, then  $(H \cdot X)_n$  is  $(\mathcal{F}_n)$ -martingale.*
- (b) *If  $X_n$  is  $(\mathcal{F}_n)$ -submartingale, then  $(H \cdot X)_n$  is  $(\mathcal{F}_n)$ -submartingale, “provided that  $H_n \geq 0$ .”*
- (c) *If  $X_n$  is  $(\mathcal{F}_n)$ -supermartingale, then  $(H \cdot X)_n$  is  $(\mathcal{F}_n)$ -supermartingale, “provided that  $H_n \geq 0$ .”*

*Proof.* Note that

$$\begin{aligned} E[(H \cdot X)_{n+1} | \mathcal{F}_n] &= E \left[ \sum_{m=1}^{n+1} H_m(X_m - X_{m-1}) \middle| \mathcal{F}_n \right] \\ &= \sum_{m=1}^n E[H_m(X_m - X_{m-1}) | \mathcal{F}_n] + E[H_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n] \\ &= \sum_{m=1}^n H_m(X_m - X_{m-1}) + H_{n+1}E[X_{n+1} - X_n | \mathcal{F}_n] \end{aligned}$$

$$= (H \cdot X)_n + \underbrace{H_{n+1}E[X_{n+1} - X_n | \mathcal{F}_n]}_{(*)}. \quad (2.1)$$

If  $X_n$  is martingale,  $(*)$  is equal to 0, so (2.1) becomes  $(H \cdot X)_n$ . If  $X_n$  is submartingale,  $(*) \geq 0$ , which implies  $(2.1) \geq (H \cdot X)_n$ .  $\square$

Now it's time to introduce a stopping time.

**Definition 2.3.11** (Stopping Time). *Let  $N$  be a r.v. taking values of nonnegative integers ( $\leq \infty$ ).  $N$  is called a **stopping time** if*

$$\forall n \geq 0, (N = n) \in \mathcal{F}_n.$$

Note that if  $N$  is a stopping time, then  $(N \leq n) \in \mathcal{F}_n$  and  $(N > n) \in \mathcal{F}_n$  also hold.

**Example 2.3.12** (Stopped process). Let  $X_n$  be a (sub-/super-) martingale, and  $N$  be a stopping time. Letting  $H_m = I(N \geq m)$ , it becomes predictable ( $H_m \in \mathcal{F}_{m-1}$ ). Thus,

$$\begin{aligned} (H \cdot X)_n &= \sum_{m=1}^n I(N \geq m)(X_m - X_{m-1}) \\ &= \sum_{m=1}^{\infty} I(m \leq n)I(N \geq m)(X_m - X_{m-1}) \\ &= \sum_{m=1}^{\infty} I(m \leq N \wedge n)(X_m - X_{m-1}) \\ &= \sum_{m=1}^{N \wedge n} (X_m - X_{m-1}) \\ &= X_{N \wedge n} - X_0 \end{aligned}$$

holds. It implies that a “stopped process”  $(X_{N \wedge n})_{n \geq 0}$  is  $(\mathcal{F}_n)$ -(sub-/super-) martingale.

Following “upcrossing process” is set-up for convergence theorem.

**Example 2.3.13.** Let  $X_n$  be  $(\mathcal{F}_n)$ -submartingale, and  $a < b$ . Define

$$N_1 = \inf\{m \geq 0 : X_m \leq a\}$$

$$N_2 = \inf\{m > N_1 : X_m \geq b\}$$

$$N_3 = \inf\{m > N_2 : X_m \leq a\}$$

$$N_4 = \inf\{m > N_3 : X_m \geq b\}$$

$$\vdots$$

See figure 2.1.

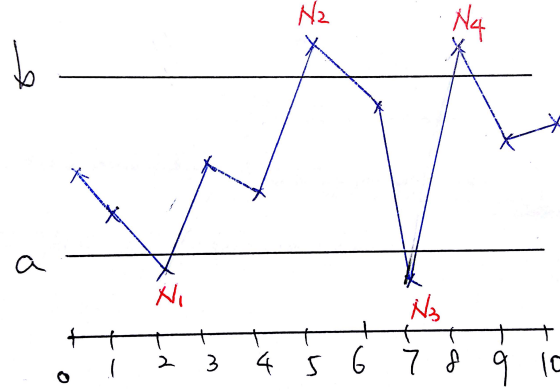


Figure 2.1:  $X_n$  and  $N_n$ 's. For example,  $N_4 = 8$ .

Then  $N_k$ 's become a stopping time. First,  $N_1$  is a stopping time, because

$$(N_1 = n) = (X_m > a \ \forall m \leq n-1, X_n \leq a) = \bigcap_{m=0}^{n-1} (X_m > a) \cap (X_n \leq a) \in \mathcal{F}_n.$$

Next,  $N_2$  is also a stopping time from

$$(N_2 = n) = \bigcup_{m=0}^{n-1} (N_1 = m) \cap (X_l < b \ \forall l \text{ s.t. } m < l \leq n-1) \cap (X_n \geq b) \in \mathcal{F}_n.$$

Then  $N_3$  is a stopping time, ..., and by induction, we get  $N_k$  is a stopping time.

Now define an “upcrossing process,”

$$U_n := \sup\{k : N_{2k} \leq n\} \text{ for } n \geq 1.$$

Then  $U_n$  is “the number of upcrossings (from  $a$  to  $b$ ) completely by time  $n$ .” Note that  $U_n \leq n$ .

Also note that,  $N_{2U_n} \leq n$ . See figure 2.2.

Now our assertion is:

**Theorem 2.3.14** (Upcrossing inequality).  $(b-a)EU_n \leq E(X_n - a)^+ - E(X_0 - a)^+.$

*Proof.* Let  $Y_n = (X_n - a)^+ + a = X_n \vee a$  (See figure 2.3). Then by Jensen’s inequality,  $Y_n$  is  $(\mathcal{F}_n)$ -submartingale, and the numbers of upcrossings of  $X_n$  and  $Y_n$  are the same. Thus, we may

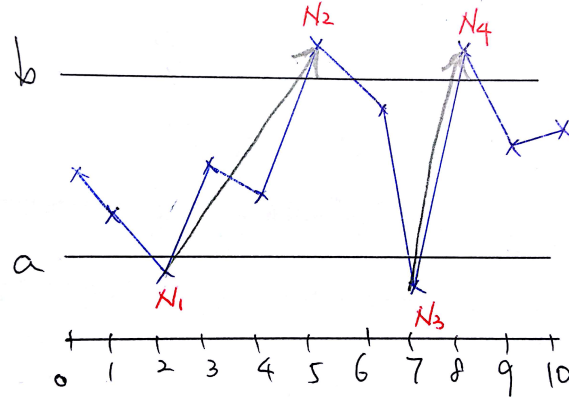


Figure 2.2: Upcrossing process. For example, in this figure,  $U_{10} = 2$ .

consider  $Y_n$  instead of  $X_n$  without loss of generality.

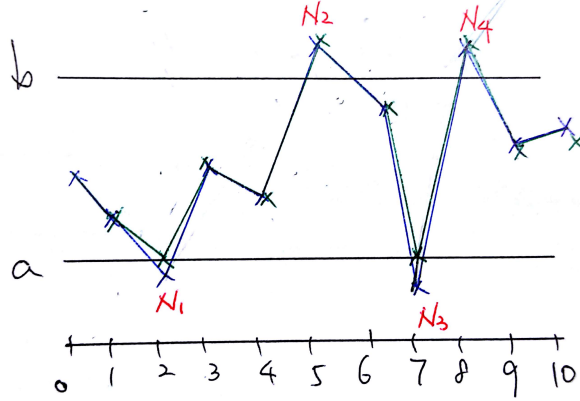


Figure 2.3: Upcrossing process and  $Y_n$ .

Note that from  $Y_{N_{2k}} - Y_{N_{2k-1}} \geq b - a$ , we get

$$(b - a)U_n \leq \sum_{k=1}^{U_n} (Y_{N_{2k}} - Y_{N_{2k-1}}).$$

Now letting  $J_k = \{N_{2k-1} + 1, \dots, N_{2k}\} = \{m : N_{2k-1} < m \leq N_{2k}\}$  and  $J = \bigcup_{k=1}^{U_n} J_k$ , we get

$$\begin{aligned} (b - a)U_n &\leq \sum_{k=1}^{U_n} (Y_{N_{2k}} - Y_{N_{2k-1}}) \\ &= \sum_{k=1}^{U_n} \sum_{i \in J_k} (Y_i - Y_{i-1}) \\ &= \sum_{m \in J} (Y_m - Y_{m-1}). \end{aligned}$$

Now define a predictable process

$$H_m = I(m \in J) = I(N_{2k-1} < m \leq N_{2k} \text{ for some } k = 1, 2, \dots, n).$$

(Note that  $N_{2U_n} \leq n$ ) Then

$$\sum_{m \in J} (Y_m - Y_{m-1}) = \sum_{m=1}^n H_m (Y_m - Y_{m-1}) = (H \cdot Y)_n$$

becomes a martingale transform. ( $H_m$  is predictable from  $(N_{2k-1} < m \leq N_{2k}) = (N_{2k-1} \leq m-1) \cap (N_{2k} \leq m-1)^c \in \mathcal{F}_{m-1}$ .) Hence,  $(H \cdot Y)_n$  is submartingale. Now, define  $\tilde{H}_m = 1 - H_m$ . Then  $(\tilde{H} \cdot Y)_n$  also becomes submartingale and

$$Y_n - Y_0 = \sum_{m=1}^n (H_m + \tilde{H}_m)(Y_m - Y_{m-1}) = (H \cdot Y)_n + (\tilde{H} \cdot Y)_n,$$

so we get  $E(\tilde{H} \cdot Y)_n \geq E(\tilde{H} \cdot Y)_1 \geq 0$  and hence

$$Y_n - Y_0 = (H \cdot Y)_n + (\tilde{H} \cdot Y)_n \geq (H \cdot Y)_n,$$

i.e.,

$$E(Y_n - Y_0) \geq E(H \cdot Y)_n.$$

Recall that  $Y_n = (X_n - a)^+ + a$ . Therefore, we get

$$(b - a)EU_n \leq E(H \cdot Y)_n \leq E(Y_n - Y_0) = E(X_n - a)^+ - E(X_0 - a)^+.$$

□

**Remark 2.3.15.** The key fact is that  $E(\tilde{H} \cdot Y)_n \geq 0$ , that is, *no matter how hard you try, you can't lose money betting on a submartingale.* (Note that  $(\tilde{H} \cdot Y)_n$  is “total profit resulted in downcrossing.”)

Indeed, our goal was following **Martingale convergence theorem**.

**Theorem 2.3.16** (Martingale convergence theorem). *If  $X_n$  is a  $((\mathcal{F}_n)$ -)submartingale with  $\sup_n EX_n^+ < \infty$ , then as  $n \rightarrow \infty$ ,  $X_n$  converges a.s. to a limit  $X$  with  $E|X| < \infty$ .*

*Proof.* Note that  $(x - a)^+ \leq x^+ + |a|$  (See figure 2.4). Then we get

$$EU_n \leq \frac{E(X_n - a)^+ - E(X_0 - a)^+}{b - a} \leq \frac{E(X_n - a)^+}{b - a} \leq \frac{EX_n^+ + |a|}{b - a} \leq \frac{\sup_n EX_n^+ + |a|}{b - a}.$$

Note that  $U_n$  is monotone, so  $\exists U$  s.t.  $U_n \nearrow U$ . Then from MCT (proposition 2.2.11)  $EU_n \nearrow EU$  and hence

$$EU \leq \frac{\sup_n EX_n^+ + |a|}{b - a} < \infty.$$

From this we get  $EU < \infty$ , which implies  $U < \infty$  a.s.. As  $U$  means “the number of whole upcrossings,” from  $U < \infty$ , we get

$$P\left(\liminf_{n \rightarrow \infty} X_n < a < b < \limsup_{n \rightarrow \infty} X_n\right) = 0.$$

(The number of whole upcrossing should not be infinite) Since it holds for any  $a, b \in \mathbb{Q}$  s.t.  $a < b$ , we get

$$P\left(\bigcup_{\substack{a, b \in \mathbb{Q} \\ a < b}} \left\{ \liminf_{n \rightarrow \infty} X_n < a < b < \limsup_{n \rightarrow \infty} X_n \right\}\right) = 0,$$

i.e.,  $\liminf X_n = \limsup X_n$   $P$ -a.s., which implies  $\exists \lim X_n =: X$   $P$ -a.s.. (For well-definedness, let  $X = 0$  if  $\liminf X_n \neq \limsup X_n$ ) Now by Fatou's lemma,

$$EX^+ \leq \liminf_{n \rightarrow \infty} EX_n^+ < \infty$$

holds, so  $EX^+ < \infty$  and  $X < \infty$   $P$ -a.s.. Since  $X_n$  is submartingale,  $EX_n \geq EX_0$ , so

$$EX_n^- = EX_n^+ - EX_n \leq EX_n^+ - EX_0$$

holds, and by Fatou again, we get

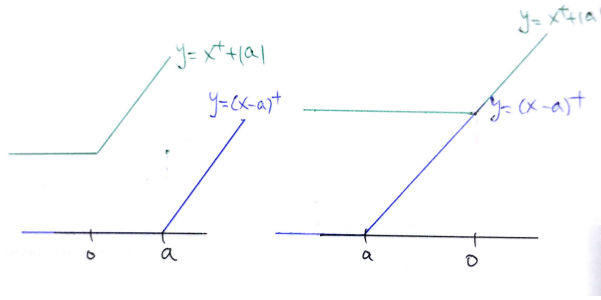
$$EX^- \leq \liminf_{n \rightarrow \infty} EX_n^- \leq \sup_n EX_n^+ - EX_0 < \infty.$$

Therefore,  $EX^- < \infty$ , which implies that (with  $EX^+ < \infty$ )  $X$  is finite almost surely, and integrable (i.e.,  $E|X| < \infty$ ).

□

**Corollary 2.3.17.** *If  $X_n \geq 0$  is a  $((\mathcal{F}_n))$ -supermartingale, then as  $n \rightarrow \infty$ ,  $\exists X$  s.t.  $X_n \rightarrow X$*



Figure 2.4:  $y = (x - a)^+$  and  $y = x^+ + |a|$ .

*a.s.* and  $EX \leq EX_0 < \infty$ .

*Proof.*  $Y_n = -X_n \leq 0$  is a submartingale with  $EY_n^+ = 0$ . Thus by previous theorem,  $Y_n$  has a limit  $Y$ , and  $X_n \xrightarrow[n \rightarrow \infty]{a.s.} -Y =: X$ . As  $X_n$  is a supermartingale, we get  $EX_0 \geq EX_n$ , and with Fatou's lemma, we obtain  $EX \leq EX_0$ .

**Example 2.3.18.** Let  $\xi_1, \xi_2, \dots$ , be i.i.d. r.v.'s with  $P(\xi_1 = 1) = P(\xi_1 = -1) = 1/2$ . Also define

$$S_0 = 1, S_n = S_{n-1} + \xi_n, n \geq 1,$$

and  $\mathcal{F}_0 = \{\phi, \Omega\}$ ,  $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$ . Then  $S_n$  is  $(\mathcal{F}_n)$ -martingale. Let  $N = \inf\{n : S_n = 0\}$ . Then from  $S_0 = 1$ ,  $N > 0$ . Also note that  $N$  becomes a stopping time. Let

$$X_n = S_{N \wedge n}.$$

Then by example 2.3.12,  $X_n$  is also a martingale. Now, note that by definition of  $N$ , and from  $S_0 = 1$ ,

$$m \leq N \Rightarrow S_m \geq 0,$$

which implies  $X_n \geq 0$ . Note that on  $(N = \infty)$ ,  $X_n = S_n$  holds  $(\star)$ . Also, it's known that

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n^{1/2}(\log \log n)^{1/2}} = \sigma\sqrt{2},$$

and with this we can obtain that

$$\limsup_{n \rightarrow \infty} S_n = \infty, \liminf_{n \rightarrow \infty} S_n = -\infty \quad P - a.s..$$

Thus,

$$P(N = \infty) = P\left(N = \infty, \limsup_{n \rightarrow \infty} S_n = \infty, \liminf_{n \rightarrow \infty} S_n = -\infty\right) \leq P\left(\limsup_{n \rightarrow \infty} X_n = \infty, \liminf_{n \rightarrow \infty} X_n = -\infty\right)$$

holds from  $(\star)$ . Note that by previous corollary, since  $X_n$  is martingale, it converges to some  $X$  almost surely, which implies that

$$P\left(\limsup_{n \rightarrow \infty} X_n = \infty, \liminf_{n \rightarrow \infty} X_n = -\infty\right) = 0.$$

This implies that  $N < \infty$  a.s.. Therefore,

$$\lim_{n \rightarrow \infty} X_n = \lim_{n \rightarrow \infty} S_{N \wedge n} = S_N = 0.$$

However, it means that  $X_n \xrightarrow{a.s.} 0$ , while  $EX_n = EX_0 = 1$  for any  $n$ . Therefore, even if  $X_n$  converges almost surely, we cannot say that  $X_n$  also converges in  $\mathcal{L}^1$ .  $\square$

**Example 2.3.19.** If  $X_n$  is  $(\mathcal{F}_n)_{n \geq 0}$ -submartingale s.t.  $X_n \leq 0$ , then we can define

$$X_\infty = \lim_{n \rightarrow \infty} X_n, \mathcal{F}_\infty = \sigma\left(\bigcup_{n=0}^{\infty} \mathcal{F}_n\right)$$

and it can be obtained that

$$(X_n)_{0 \leq n \leq \infty} \text{ is } (\mathcal{F}_n)_{0 \leq n \leq \infty}\text{-submartingale,}$$

i.e.,

$$E(X_\infty | \mathcal{F}_n) \geq X_n \text{ } P - a.s. \text{ } \forall n \geq 0.$$

In this situation, we say that  $X_n$  is “closable.” To show this, we need *Fatou’s lemma* in conditional context.

**Lemma 2.3.20** (Conditional Fatou lemma). *Suppose that  $X_n \geq 0$ ,  $X_n \xrightarrow{a.s.} X$ , and  $E|X| < \infty$ . Then for sub  $\sigma$ -field  $\mathcal{F}$ ,*

$$E(X | \mathcal{F}) \leq \liminf_{n \rightarrow \infty} E(X_n | \mathcal{F}).$$

*Proof.* Let  $M > 0$  be a constant. Then by DCT (proposition 2.2.11),

$$E(X \wedge M | \mathcal{F}) = \lim_{n \rightarrow \infty} E(X_n \wedge M | \mathcal{F})$$

holds.  $X_n \wedge M \leq X_n$  implies that  $\lim_{n \rightarrow \infty} E(X_n \wedge M | \mathcal{F}) \leq \liminf_{n \rightarrow \infty} E(X_n | \mathcal{F})$ , so we get

$$E(X \wedge M | \mathcal{F}) \leq \liminf_{n \rightarrow \infty} E(X_n | \mathcal{F}) \quad \forall M > 0.$$

Letting  $M \rightarrow \infty$ , we get  $E(X \wedge M | \mathcal{F}) \xrightarrow{n \rightarrow \infty} E(X | \mathcal{F})$  by MCT (proposition 2.2.11), and hence

$$E(X | \mathcal{F}) \leq \liminf_{n \rightarrow \infty} E(X_n | \mathcal{F}).$$

□

Now come back to our example. By martingale convergence theorem,  $\exists X_\infty = \lim_{n \rightarrow \infty} X_n \in \mathcal{F}_\infty$ , and  $X_\infty \leq 0$ , by negativity of  $X_n$ . By conditional Fatou,

$$E(-X_\infty | \mathcal{F}_n) \leq \liminf_{m \rightarrow \infty} E(-X_m | \mathcal{F}_n) \leq (-X_n)$$

for arbitrary given  $n$ . The last inequality holds because  $(-X_n)$  is supermartingale. Therefore, we get

$$E(X_\infty | \mathcal{F}_n) \geq X_n \quad P - a.s..$$

Following theorem is very useful in martingale theory.

**Theorem 2.3.21** (Doob decomposition theorem). *Any submartingale  $X_n$  can be expressed uniquely 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.* (Motivation: if it holds,  $E(X_n | \mathcal{F}_{n-1}) = E(M_n | \mathcal{F}_{n-1}) + E(A_n | \mathcal{F}_{n-1}) = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n$ .)

Let

$$A_n = A_{n-1} + E(X_n | \mathcal{F}_{n-1}) - X_{n-1}.$$

Then since  $X_n$  is submartingale,  $E(X_n | \mathcal{F}_{n-1}) - X_{n-1} \geq 0$ , and hence  $A_n$  is increasing. Further, by induction,  $A_n$  is predictable. Define

$$M_n = X_n - A_n,$$

and then we obtain

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

which implies that  $M_n$  is a martingale. In here, the second equality holds from the definition of  $A_n$  and predictability, while the third one comes from  $X_{n-1} \in \mathcal{F}_{n-1}$ .

Now for uniqueness, suppose that we have two decompositions,

$$X_n = M_n + A_n = M'_n + A'_n.$$

Then from

$$M_n - M'_n = A'_n - A_n,$$

$M_n - M'_n$  is predictable martingale, which implies that  $M_n - M'_n = M_0 - M'_0$ . Since  $A_0 = A'_0$ , it yields that  $M_n = M'_n$ .  $\square$

Note that Doob decomposition implies that, if  $X_n$  is a martingale,  $X_n^2$  is a submartingale, and therefore, there exists a unique predictable increasing process  $\langle X \rangle_n$  such that  $X_n^2 - \langle X \rangle_n$  becomes a martingale.  $\langle X \rangle$  is called a “quadratic variation.”

**Remark 2.3.22** (Annotation by compiler). In 1953, Doob published previous theorem, and conjectured a continuous time version of the theorem. In 1962 and 1963, Paul-André Meyer proved such a theorem, which became known as the *Doob-Meyer decomposition*. It implies following: For filtered probability space  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, P)$  and any right-continuous square-integrable  $(\mathcal{F}_t)$ -adapted martingale  $(X_t)_{t \geq 0}$ , there exists a unique continuous increasing predictable process  $\langle X \rangle$ ,  $\langle X \rangle_0 = 0$  and such that  $X^2 - \langle X \rangle$  is a martingale. For example, if  $(B_t)_{t \geq 0}$  is a standard Brownian motion, then  $\langle B \rangle_t = t$ .

One important application of Doob-Meyer decomposition in statistics is for survival analysis. Let  $N(t)$  be a counting process, which is defined as a stochastic process with the properties that  $N(0) = 0$ ,  $P(N(t) < \infty) = 1$ , and the sample paths of  $N(t)$  are right-continuous, piecewise constant with jumps of size +1. In survival analysis,  $N(t)$  often denotes “the number of event occurs,” i.e., the number of dead people at time  $t$ . Then there is a smooth predictable process  $\Lambda(t)$  which makes  $M(t) := N(t) - \Lambda(t)$  a martingale.  $M(t)$  is called a counting process martingale. Now, for quadratic variation  $\langle M \rangle$  of  $M^2$ , we have  $Var(dM(t)|\mathcal{F}_{t-}) = d\langle M \rangle(t)$ . Using this, we can construct a *stochastic integrals* of the basic martingale. For example, let  $Y(t)$  be “at risk process,” which denotes the number of individuals at risk at a given time. Then  $Y(t)$  becomes predictable, so we can define a stochastic integral

$$\int_0^t Y(s) dM(s),$$

which also becomes a martingale (Indeed, it is “generalization of martingale transform”), and quadratic variation becomes

$$\left\langle \int_0^t Y(s) dM(s) \right\rangle = \int_0^t Y^2(s) d\langle M \rangle(s).$$

### 2.3.2 Examples

#### Bounded increments

**Proposition 2.3.23** (Bounded increments). *Let  $X_n$  be a martingale with  $|X_{n+1} - X_n| \leq M < \infty$  for any  $n$ , and define*

$$C = \{\lim X_n \text{ exists and is finite}\}$$

$$D = \{\limsup X_n = \infty \text{ and } \liminf X_n = -\infty\}.$$

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

*Proof.* WLOG  $X_0 = 0$ . (Why “WLOG”? Let  $\tilde{X}_n = X_n - X_0$ . Then  $\tilde{X}_n$  is also a martingale, and it has bounded increments, i.e.,  $|\tilde{X}_{n+1} - \tilde{X}_n| \leq M$ . Further, for

$$\tilde{C} = \{\lim \tilde{X}_n \text{ exists and is finite}\}$$

$$\tilde{D} = \{\limsup \tilde{X}_n = \infty \text{ and } \liminf \tilde{X}_n = -\infty\},$$

$\tilde{C} = C$  and  $\tilde{D} = D$  holds.) For any  $K > 0$ , define

$$N_K = \inf\{n \geq 1 : X_n \leq -K\}.$$

Then

$$(N_K = n) = (\forall m < n \ X_m > -K, \ X_n \leq -K) \in \mathcal{F}_n$$

for any  $n$ , so  $N_K$  is a stopping time, and hence  $\{X_{n \wedge N_K} : n \geq 0\}$  is a martingale. Note that on  $(N_K < \infty)$ ,

$$X_k > -K \text{ for } k = 1, 2, \dots, N_K - 1,$$

and thus

$$X_{N_K} = X_{N_K-1} + \underbrace{(X_{N_K} - X_{N_K-1})}_{\geq -M} \geq -K - M,$$

and on  $(N_K = \infty)$ ,  $X_n > -K > -K - M$ , so for any cases  $X_{n \wedge N_K} + K + M \geq 0$ . Thus

( $X_{n \wedge N_K} + K + M$  is a nonnegative (super)martingale) by martingale convergence theorem, ( $X_{n \wedge N_K} + K + M$ , and consequently,)  $X_{n \wedge N_K}$  converges almost surely to some integrable random variable. In particular,  $X_n$  converges ( $P$ -)a.s. “on  $(N_K = \infty)$ .” (It means that,  $\exists E \subseteq (N_K = \infty)$  s.t.  $P((N_K = \infty) \setminus E) = 0$  and  $X_n$  converges pointwisely on  $E$ .) Since  $K > 0$  was arbitrary, so  $X_n$  converges  $P$ -a.s. on  $\bigcup_{K=1}^{\infty} (N_K = \infty)$ . Now, from

$$(\liminf X_n > -\infty) \subseteq \bigcup_{K=1}^{\infty} (N_K = \infty),$$

( $\because$  if  $\forall K (N_K < \infty)$ , then for any  $K$  we can find  $n$  s.t.  $X_n < -K$ , i.e.,  $\liminf X_n = -\infty$ ) we can obtain that  $X_n$  converges  $P$ -a.s. on  $(\liminf X_n > -\infty)$ . Applying such procedure to  $-X_n$  repeatedly, we can obtain that

$$-X_n \text{ converges on } (\liminf(-X_n) > -\infty) = (\limsup X_n < \infty).$$

Therefore,  $\underbrace{X_n \text{ converges } P\text{-a.s. on}}_{=C} \underbrace{(-\infty < \liminf X_n) \cup (\limsup X_n < \infty)}_{=D^c}$ , i.e.,  $C \supseteq D^c$  (except probability zero set). It implies that  $P(C \cup D) = 1$ .  $\square$

With this, we can find similar argument as “Borel-Cantelli Lemma” in filtered probability space. It can be also called as “conditional Borel-Cantelli lemma.”

**Theorem 2.3.24** (Second Borel Cantelli Lemma, “conditional”). *Let  $\mathcal{F}_0 = \{\phi, \Omega\}$  and  $(\mathcal{F}_n)_{n \geq 0}$  be a filtration. If  $A_n \in \mathcal{F}_n \forall n \geq 1$ , then*

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

**Remark 2.3.25.** If  $A_n$ ’s are independent set, letting  $\mathcal{F}_n = \sigma(A_1, \dots, A_n)$ , we get  $P(A_n | \mathcal{F}_{n-1}) = P(A_n)$ , and hence

$$(A_n \text{ i.o.}) = \left( \sum_{n=1}^{\infty} P(A_n) = \infty \right),$$

i.e.,

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Leftrightarrow A_n \text{ i.o.}$$

In other words, “conditional” version of Borel-Cantelli lemma contains ordinary one.

*Proof.* Let  $X_0 = 0$ , and define

$$X_n = \sum_{m=1}^n \underbrace{\{I_{A_m} - P(A_m|\mathcal{F}_{m-1})\}}_{\in \mathcal{F}_m}.$$

Then  $X_n$  is a martingale, because

$$E(X_{n+1}|\mathcal{F}_n) = X_n + E(I_{A_{n+1}} - P(A_{n+1}|\mathcal{F}_n)|\mathcal{F}_n) = X_n.$$

Also, note that

$$|X_{n+1} - X_n| = |I_{A_{n+1}} - P(A_{n+1}|\mathcal{F}_n)| \leq 1,$$

i.e.,  $\{X_n\}$  has bounded increments. Now on  $C$ ,  $X_n = \sum_{m=1}^n (I_{A_m} - P(A_m|\mathcal{F}_{m-1}))$  converges and is finite, so

$$\sum_{m=1}^{\infty} I_{A_m} = \infty \Leftrightarrow \sum_{m=1}^{\infty} P(A_m|\mathcal{F}_{m-1}) = \infty.$$

Note that  $\sum_{m=1}^{\infty} I_{A_m} = \infty$  means  $A_n$  occurs infinitely often. On the other hand, on  $D$ , from  $\sum_{m=1}^n I_{A_m} \geq X_n$ , we get

$$\sum_{m=1}^{\infty} I_{A_m} \geq \limsup_D X_n = \infty,$$

and from

$$\sum_{m=1}^n P(A_m|\mathcal{F}_{m-1}) = \sum_{m=1}^n I_{A_m} - X_n \geq -X_n,$$

we get

$$\sum_{m=1}^{\infty} P(A_m|\mathcal{F}_{m-1}) \geq \limsup(-X_n) = -\liminf X_n = \infty.$$

Therefore, on  $D$ ,

$$\sum_{m=1}^{\infty} I_{A_m} = \infty \text{ and } \sum_{m=1}^{\infty} P(A_m|\mathcal{F}_{m-1}) = \infty \text{ simultaneously.}$$

Now previous proposition ( $P(C \cup D) = 1$ ) ends the proof. More precisely, from

$$\left( \sum_{m=1}^{\infty} I_{A_m} = \infty \right) \cap C = \left( \sum_{m=1}^{\infty} P(A_m|\mathcal{F}_{m-1}) = \infty \right) \cap C$$

and

$$\left( \sum_{m=1}^{\infty} I_{A_m} = \infty \right) \cap D = \left( \sum_{m=1}^{\infty} P(A_m|\mathcal{F}_{m-1}) = \infty \right) \cap D,$$

we get

$$\left( \sum_{m=1}^{\infty} I_{A_m} = \infty \right) \cap (C \cup D) = \left( \sum_{m=1}^{\infty} P(A_m | \mathcal{F}_{m-1}) = \infty \right) \cap (C \cup D),$$

i.e.,

$$\left( \sum_{m=1}^{\infty} I_{A_m} = \infty \right) = \left( \sum_{m=1}^{\infty} P(A_m | \mathcal{F}_{m-1}) = \infty \right) \quad P - \text{a.s.}$$

□

## Branching Process

**Definition 2.3.26** (Branching process). Let  $\xi_i^n$ ,  $i, n \geq 0$  be i.i.d. nonnegative integer valued random variables, and  $Z_0 = 1$ . Now define

$$Z_{n+1} = \begin{cases} \sum_{k=1}^{Z_n} \xi_k^{n+1} & \text{if } Z_n > 0 \\ 0 & \text{if } Z_n = 0 \end{cases}.$$

$(Z_n)_{n \geq 0}$  is called a branching process.

**Remark 2.3.27.** In here,  $Z_n$  can be interpreted as “a population in generation  $n$ .” In  $n$ th generation, each  $(Z_n)$  individual produces some random number of individuals in  $(n+1)$ th generation. If  $Z_n$  becomes 0, it denotes “extinction.” In this model, our interest is “the probability of ultimate extinction.” It is known that for  $\mu = E\xi_i^n$ , if  $\mu < 1$ , then population ultimately extincts with probability 1, and if  $\mu > 1$ , then the probability of ultimate extinction is less than 1 (but not necessarily zero). In this lecture, we will see the case  $\mu < 1$ .

**Lemma 2.3.28.** Let  $\mathcal{F}_n = \sigma(\xi_i^m : i \geq 1, 1 \leq m \leq n)$ . ( $m$  denotes “generation”) Then under the assumption  $0 < \mu < \infty$ ,

$$\frac{Z_n}{\mu^n} \text{ is } (\mathcal{F}_n) - \text{martingale.}$$

*Proof.* First, it is clear that  $Z_n \in \mathcal{F}_n$ . Next,

$$\begin{aligned} E \left[ \frac{Z_{n+1}}{\mu^{n+1}} \middle| \mathcal{F}_n \right] &= \sum_{k=0}^{\infty} \frac{1}{\mu^{n+1}} E [ Z_{n+1} I(Z_n = k) | \mathcal{F}_n ] \quad (\text{conditional MCT}) \\ &= \sum_{k=0}^{\infty} \frac{1}{\mu^{n+1}} E \left[ \underbrace{(\xi_1^{n+1} + \dots + \xi_k^{n+1})}_{\text{independent of } \mathcal{F}_n} \underbrace{I(Z_n = k)}_{\in \mathcal{F}_n} \middle| \mathcal{F}_n \right] \\ &= \sum_{k=0}^{\infty} I(Z_n = k) \cdot \frac{E(\xi_1^{n+1} + \dots + \xi_k^{n+1})}{\mu^{n+1}} \end{aligned}$$



$$\begin{aligned}
&= \frac{1}{\mu^n} \sum_{k=0}^n I(Z_n = k)k \\
&= \frac{1}{\mu^n} \sum_{k=0}^n I(Z_n = k)Z_n \\
&= \frac{Z_n}{\mu^n}
\end{aligned}$$

holds. □

**Theorem 2.3.29.** *If  $0 < \mu < 1$ , then  $Z_n = 0 \forall$  large  $n$ ,  $P$ -a.s.*

*Proof.* Since  $Z_n$  is integer,  $P(Z_n > 0) = P(Z_n \geq 1) \leq E(Z_n I(Z_n \geq 1)) = E(Z_n I(Z_n > 0))$ , and so

$$P(Z_n > 0) \leq E(Z_n I(Z_n > 0)) = E(Z_n I(Z_n > 0) + Z_n I(Z_n = 0)) = EZ_n = \mu^n$$

holds. The last equality is from  $E(\mu^{-n} Z_n) = E(\mu^{-0} Z_0) = 1$  ( $\because \mathcal{F}_0 = \{\phi, \Omega\}$ ). Thus we get  $P(Z_n > 0) \leq \mu^n$ , and therefore, by Borel-Cantelli lemma,  $Z_n = 0$  holds for all but finite  $n$ . □

It also implies that,

$$\frac{Z_n}{\mu^n} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

### 2.3.3 Doob's inequality

**Proposition 2.3.30.** *If  $\{X_n\}$  is a submartingale, and  $N$  is a stopping time with  $P(N \leq k) = 1$  for some  $k \geq 0$ , then*

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

*Proof.* Note that  $X_{n \wedge N}$  is a submartingale. Thus,

$$EX_0 = EX_{0 \wedge N} \leq EX_{k \wedge N} = EX_N$$

holds. Thus our claim is:

**Claim.**  $EX_N \leq EX_k$ .

Let  $K_n = I(N \leq n - 1)$ . Then  $K_n \in \mathcal{F}_{n-1}$  so it is predictable, and hence we can define  $(K \cdot X)_n$ . Then since

$$I(N < m \leq n) = I(N \wedge n < m \leq n) = I(N \wedge n + 1 \leq m \leq n),$$

we get

$$\begin{aligned}
(K \cdot X)_n &= \sum_{m=1}^n I(N \leq m-1)(X_m - X_{m-1}) \\
&= \sum_{m=1}^n I(N < m \leq n)(X_m - X_{m-1}) \\
&= \sum_{m=1}^n I(N \wedge n + 1 \leq m \leq n)(X_m - X_{m-1}) \\
&= \sum_{N \wedge n + 1}^n (X_m - X_{m-1}) \\
&= X_n - X_{N \wedge n}.
\end{aligned}$$

Note that  $(K \cdot X)_n$  is also a submartingale; hence we get

$$E(K \cdot X)_k = EX_k - EX_{N \wedge k} \geq E(K \cdot X)_1 = E[I(N=0)(X_1 - X_0)] = E \left[ I(N=0) \underbrace{E(X_1 - X_0 | \mathcal{F}_0)}_{\geq 0} \right] \geq 0,$$

i.e.,

$$EX_k \geq EX_{N \wedge k}.$$

However,  $N \wedge k = N$ , so we get the conclusion.  $\square$

**Theorem 2.3.31** (Submartingale Inequality). *Let  $X_n$  be a submartingale. Then for  $\tilde{X}_n = \max_{0 \leq m \leq n} X_m$  and  $\lambda > 0$ ,*

$$\lambda P(\tilde{X}_n \geq \lambda) \leq EX_n I(\tilde{X}_n \geq \lambda) \leq EX_n^+ I(\tilde{X}_n \geq \lambda) \leq EX_n^+.$$

*Proof.* Let  $A = (\tilde{X}_n \geq \lambda)$ , and  $N = \inf\{m \leq n : X_m \geq \lambda\} \wedge n$ . Then  $N$  is a stopping time less than  $n$ . Note that

$$X_N I_A \geq \lambda I_A$$

holds (On  $A$ ,  $\exists m \leq n$  s.t.  $X_m \geq \lambda$ , so  $X_N \geq \lambda$ . On  $A^c$ , both sides are all zero). Therefore, we get

$$\lambda P(A) \leq EX_N I_A = EX_N - EX_N I_{A^c} \stackrel{(*)}{\leq} EX_n - EX_n I_{A^c} = EX_n I_A.$$

(\*) is obtained from  $EX_N \leq EX_n$  (previous proposition), and on  $A^c$ ,  $N = n$ , i.e.,  $X_N = X_n$ .  $\square$

If  $X_n$  is a submartingale, then  $-X_n$  is a supermartingale; thus, we can also get a similar result

for a supermartingale.

**Corollary 2.3.32** (Supermartingale inequality). *Let  $X_n$  be a supermartingale. Then*

$$\lambda P(\tilde{X}_n \geq \lambda) \leq EX_0 - EX_n I_{A^c} \leq EX_0 + EX_n^-.$$

*Proof.* Let  $N = \inf\{m \leq n : X_m \geq \lambda\} \wedge n$ . Then

$$EX_0 \geq EX_N = EX_N I_A + EX_N I_{A^c} \underset{(*)}{\geq} \lambda P(A) + EX_n I_{A^c}$$

holds.  $(*)$  holds from: On  $A^c$ ,  $N = n$ , and on  $A$ ,  $X_N \geq \lambda$ . □

Following is called **Doob's inequality**, or **Doob's maximal inequality**, which is very important result in martingale theory.

**Theorem 2.3.33** (Doob's maximal inequality). *If  $X_n$  is a nonnegative submartingale, then for  $1 < p < \infty$ ,*

$$E \max_{1 \leq m \leq n} X_m^p \leq \left( \frac{p}{p-1} \right)^p EX_n^p.$$

*We often use the case  $p = 2$ , i.e.,*

$$E \max_{1 \leq m \leq n} X_m^2 \leq 4EX_n^2.$$

*Proof.* If  $E\tilde{X}_n^p = 0$ , then  $X_n = 0$  almost surely for any  $n$ , so there is nothing to show. So we may assume that  $E\tilde{X}_n^p > 0$ . Let  $M > 0$ . Then

$$\begin{aligned} E(\tilde{X}_n \wedge M)^p &= \int_0^\infty P(\tilde{X}_n \wedge M > y) p y^{p-1} dy \quad (\because \text{Fubini}) \\ &= \int_0^M p y^{p-1} P(\tilde{X}_n \wedge M > y) dy \\ &\leq \int_0^M p y^{p-1} P(\tilde{X}_n > y) dy \\ &\leq \int_0^M p y^{p-1} \cdot \frac{1}{y} EX_n I(\tilde{X}_n \geq y) dy \quad (\because \lambda P(A) \leq EX_n I_A) \\ &= \int_0^M \int X_n I(\tilde{X}_n \geq y) d\mathbb{P} p y^{p-2} dy \\ &= \int X_n \left( \int_0^M I(\tilde{X}_n \geq y) p y^{p-2} dy \right) d\mathbb{P} \\ &= \int X_n \int_0^{M \wedge \tilde{X}_n} p y^{p-2} dy d\mathbb{P} \end{aligned}$$

$$\begin{aligned}
&= E \left[ X_n \cdot \frac{p}{p-1} \left( \tilde{X}_n \wedge M \right)^{p-1} \right] \\
&= \frac{p}{p-1} E \left[ X_n \cdot \left( \tilde{X}_n \wedge M \right)^{p-1} \right]
\end{aligned}$$

holds. Now let  $q$  be a Hölder conjugate of  $p$ , i.e.,  $q = \frac{p}{p-1}$ . Then by Hölder inequality,

$$\begin{aligned}
E \left[ X_n \cdot \left( \tilde{X}_n \wedge M \right)^{p-1} \right] &\leq (E(X_n^p))^{\frac{1}{p}} \left( E \left[ \left( \tilde{X}_n \wedge M \right)^{p-1} \right]^q \right)^{\frac{1}{q}} \\
&= (E(X_n^p))^{\frac{1}{p}} \left( E(\tilde{X}_n \wedge M)^p \right)^{\frac{1}{q}}
\end{aligned}$$

is obtained, and hence, we get

$$E(\tilde{X}_n \wedge M)^p \leq \frac{p}{p-1} (E(X_n^p))^{\frac{1}{p}} \left( E(\tilde{X}_n \wedge M)^p \right)^{\frac{1}{q}}.$$

It is equivalent to

$$\left( E(\tilde{X}_n \wedge M)^p \right)^{\frac{1}{p}} \leq \frac{p}{p-1} (E(X_n^p))^{\frac{1}{p}},$$

and therefore

$$E(\tilde{X}_n \wedge M)^p \leq \left( \frac{p}{p-1} \right)^p E(X_n^p).$$

As it holds for any  $M > 0$ , letting  $M \rightarrow \infty$ , we get

$$E\tilde{X}_n^p \leq \left( \frac{p}{p-1} \right)^p E(X_n^p)$$

with MCT. □

### 2.3.4 Stopping time and filtration

**Definition 2.3.34.** Let  $(\Omega, (\mathcal{F}_n)_{n \geq 0}, P)$  be a filtered probability space, and  $\tau$  be a stopping time. Then  $\mathcal{F}_\tau$  is defined as

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

**Remark 2.3.35.** Note that  $\mathcal{F}_\tau$  is a  $\sigma$ -field.

- (i)  $\phi \in \mathcal{F}_\tau$ , because for any  $n$ ,  $\phi \cap (\tau = n) = \phi \in \mathcal{F}_n$ .
- (ii) If  $A \in \mathcal{F}_\tau$ , for any  $n$ ,  $A^c \cap (\tau = n) = (\tau = n) \cap \{A \cap (\tau = n)\}^c \in \mathcal{F}_n$ , so  $A^c \in \mathcal{F}_\tau$ .
- (iii) If  $A_k \in \mathcal{F}_\tau$  for  $k = 1, 2, \dots$ , then  $(\cup_k A_k) \cap (\tau = n) = \cup_k (A_k \cap (\tau = n)) \in \mathcal{F}_n$ , so  $\cup_k A_k \in \mathcal{F}_\tau$ .

Also,  $\tau$  is  $\mathcal{F}_\tau$ -measurable, because for any  $k$  we get  $(\tau = k) \in \mathcal{F}_\tau$ , from

$$(\tau = k) \cap (\tau = n) = \begin{cases} (\tau = n) & n = k \\ \phi & n \neq k \end{cases} \in \mathcal{F}_n.$$

Following theorem is one version of **optional sampling theorem**, which is very important result. In here, we only see for bounded stopping times. We will deal with the general one later.

**Theorem 2.3.36** ((Bounded) Optional Sampling Theorem). *Let  $X_n$  be a submartingale and  $\sigma \leq \tau$  be bounded stopping times. Then,*

$$E(X_\tau | \mathcal{F}_\sigma) \geq X_\sigma \quad P - a.s.$$

*Especially, if  $X_n$  is a martingale, then*

$$E(X_\tau | \mathcal{F}_\sigma) = X_\sigma \quad P - a.s.$$

This statement seems very intuitive, by the definition of (sub)martingale.

*Proof.* From boundedness, we can find  $B > 0$  s.t.  $\sigma \leq \tau \leq B \in \mathbb{N}$ . First, our claim is:

**Claim.**  $E(X_\tau | \mathcal{F}_\sigma)I(\sigma = n) = E(X_\tau | \mathcal{F}_n)I(\sigma = n) \quad P\text{-a.s.}$

Proof of Claim.) For any  $a \in \mathbb{R}$  and  $k = 0, 1, 2, \dots$ , we get

$$\begin{aligned} (E(X_\tau | \mathcal{F}_n)I(\sigma = n) \leq a) \cap (\sigma = k) &= \{(E(X_\tau | \mathcal{F}_n) \leq a) \cap (\sigma = n) \cap (\sigma = k)\} \\ &\quad \cup \{(0 \leq a) \cap (\sigma \neq n) \cap (\sigma = k)\} \\ &= \begin{cases} (E(X_\tau | \mathcal{F}_n) \leq a) \cap (\sigma = n) & n = k \\ (0 \leq a) \cap (\sigma = k) & n \neq k \end{cases} \in \mathcal{F}_k \end{aligned}$$

and hence  $(E(X_\tau | \mathcal{F}_n)I(\sigma = n) \leq a) \in \mathcal{F}_\sigma \quad \forall a$ . It implies  $E(X_\tau | \mathcal{F}_n)I(\sigma = n) \in \mathcal{F}_\sigma \quad (\star)$ . Thus, for any  $A \in \mathcal{F}_\sigma$ ,

$$\begin{aligned} \int_A E(X_\tau | \mathcal{F}_\sigma)I(\sigma = n)dP &= \int_{\underbrace{A \cap (\sigma = n)}_{\in \mathcal{F}_\sigma}} E(X_\tau | \mathcal{F}_\sigma)dP \\ &= \int_{\underbrace{A \cap (\sigma = n)}_{\in \mathcal{F}_n}} X_\tau dP \quad (\text{def. of conditional expectation}) \end{aligned}$$

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

holds, which implies

$$\int_A \underbrace{(E(X_\tau | \mathcal{F}_\sigma) I(\sigma = n) - E(X_\tau | \mathcal{F}_n) I(\sigma = n))}_{\in \mathcal{F}_\sigma \text{ } (\because \star)} dP = 0 \quad \forall A \in \mathcal{F}_\sigma.$$

Hence we get

$$E(X_\tau | \mathcal{F}_\sigma) I(\sigma = n) - E(X_\tau | \mathcal{F}_n) I(\sigma = n) = 0.$$

(Recall that: if  $f$  is  $\mathcal{G}$ -mb, and  $\int_A f = 0$  for any  $A \in \mathcal{G}$ , then  $f = 0$  a.e.: Take  $A = (f > 0)$  and  $A = (f < 0)$ ! )

□ (Claim)

Back to our main theorem. To show  $E(X_\tau | \mathcal{F}_\sigma) \geq X_\sigma$ , it is sufficient to show that:

$$E(X_\tau | \mathcal{F}_\sigma) I(\sigma = n) \geq X_\sigma I(\sigma = n) \quad \forall n = 0, 1, \dots, B.$$

From  $X_\sigma I(\sigma = n) = X_n I(\sigma = n)$  and  $E(X_\tau | \mathcal{F}_\sigma) I(\sigma = n) = E(X_\tau | \mathcal{F}_n) I(\sigma = n)$  (Claim), for any  $A \in \mathcal{F}_n$ , we get

$$\begin{aligned}
\int_A X_\sigma I(\sigma = n) dP - \int_A E(X_\tau | \mathcal{F}_\sigma) I(\sigma = n) dP &= \int_A X_n I(\sigma = n) dP - \int_A E(X_\tau | \mathcal{F}_n) I(\sigma = n) dP \\
&= \int_{A \cap (\sigma=n)} X_n dP - \underbrace{\int_{A \cap (\sigma=n)} E(X_\tau | X_n) dP}_{\in \mathcal{F}_n} \\
&= \int_{A \cap (\sigma=n)} X_n dP - \int_{A \cap (\sigma=n)} X_\tau dP \\
&= \int_{A \cap (\sigma=n)} (X_n - X_\tau) dP \\
&\stackrel{\sigma \leq \tau}{=} \int_{A \cap (\sigma=n) \cap (\tau \geq n)} (X_n - X_\tau) dP \\
&= \underbrace{\int_{A \cap (\sigma=n) \cap (\tau=n)} (X_n - X_\tau) dP}_{=0 \text{ } (\because X_n = X_\tau \text{ on } (\tau=n))} \\
&\quad + \int_{A \cap (\sigma=n) \cap (\tau \geq n+1)} (X_n - X_\tau) dP \\
&= \int_{A \cap (\sigma=n) \cap (\tau \geq n+1)} (X_n - X_\tau) dP
\end{aligned}$$

$$\begin{aligned}
&\leq \int_{\underbrace{A \cap (\sigma=n) \cap (\tau \geq n+1)}_{\in \mathcal{F}_n \text{ } (\because (\tau \geq n+1) = (\tau \leq n)^c)}} (E(X_{n+1} | \mathcal{F}_n) - X_\tau) dP \\
&= \int_{A \cap (\sigma=n) \cap (\tau \geq n+1)} (X_{n+1} - X_\tau) dP \\
&\quad (\text{def. of conditional expectation}) \\
&\leq \int_{A \cap (\sigma=n) \cap (\tau \geq n+2)} (X_{n+2} - X_\tau) dP \text{ (Same way)} \\
&\leq \dots \\
&\leq \int_{A \cap (\sigma=n) \cap (\tau \geq B)} (X_B - X_\tau) dP \\
&= \int_{A \cap (\sigma=n) \cap (\tau=B)} (X_B - X_\tau) dP \text{ } (\because \tau \leq B) \\
&= 0,
\end{aligned}$$

i.e.,

$$\int_A \underbrace{(X_\sigma I(\sigma = n) - E(X_\tau | \mathcal{F}_\sigma) I(\sigma = n))}_{\in \mathcal{F}_n} dP \leq 0 \quad \forall A \in \mathcal{F}_n.$$

Therefore, we get

$$X_\sigma I(\sigma = n) \leq E(X_\tau | \mathcal{F}_\sigma) I(\sigma = n) \quad P - \text{a.s.},$$

which ends the proof. Also recall that: if  $f \in \mathcal{G}$  and  $\int_A f \geq 0 \quad \forall A \in \mathcal{G}$ , with taking  $A = (f < 0)$ , we get  $f \geq 0$ .  $\square$

## 2.4 Uniform Integrability

**Definition 2.4.1.** The family  $\{X_t : t \in T\}$  of random variables is said to be **uniformly integrable** if

$$\lim_{a \rightarrow \infty} \sup_{t \in T} \int_{|X_t| \geq a} |X_t| dP = 0.$$

**Example 2.4.2.** If  $\exists X \in \mathcal{L}^1$  s.t.  $|X_t| \leq X \quad \forall t \in T$ , then

$$\int_{|X_t| \geq a} |X_t| dP \leq \int_{|X_t| \geq a} |X| dP \leq \int_{|X| \geq a} |X| dP = aP(|X| \geq a) \xrightarrow{a \rightarrow \infty} 0,$$

so  $\{X_t : t \in T\}$  is uniformly integrable. Especially, the set of finite number of integrable r.v.'s is uniformly integrable.

Following proposition shows equivalent condition of uniform integrability. Such equivalence is

very useful.

**Proposition 2.4.3.**  $\{X_t : t \in T\}$  is uniformly integrable if and only if

$$(a) \sup_{t \in T} E|X_t| < \infty.$$

$$(b) \forall \epsilon > 0 \ \delta > 0 \text{ s.t.}$$

$$A \in \mathcal{F}, P(A) < \delta \Rightarrow \sup_{t \in T} \int_A |X_t| dP < \epsilon.$$

*Proof.*  $\Rightarrow$  (a) Let  $a$  be s.t.  $\sup_{t \in T} E[|X_t|I(|X_t| \geq a)] \leq 1$  (Such  $a$  exists because it converges to 0 as  $a \rightarrow \infty$ ). Then for any  $t \in T$

$$E|X_t| = \underbrace{E|X_t|I(|X_t| < a)}_{\leq a} + \underbrace{E|X_t|I(|X_t| \geq a)}_{\leq 1} \leq a + 1$$

holds, and hence,

$$\sup_{t \in T} E|X_t| \leq a + 1 < \infty.$$

(b) Let  $A \in \mathcal{F}$  and  $a > 0$ . Now note that

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

holds. Thus we get

$$\sup_{t \in T} \int_A |X_t| dP \leq \sup_{t \in T} \int_{|X_t| \geq a} |X_t| dP + aP(A).$$

Now choose  $a_0$  s.t.

$$\sup_{t \in T} \int_{|X_t| \geq a_0} |X_t| dP < \frac{\epsilon}{2}$$

and let  $\delta = \epsilon/2a_0$ . Then for measurable set  $A$  s.t.  $P(A) < \delta$ ,

$$\sup_{t \in T} \int_A |X_t| dP \leq \sup_{t \in T} \int_{|X_t| \geq a} |X_t| dP + aP(A) \leq \frac{\epsilon}{2} + a_0\delta = \epsilon$$

holds.

$\Leftarrow$ ) Let  $\epsilon > 0$  be arbitrarily given, and  $\delta > 0$  be the real number satisfying (b). Now put

$$M = \sup_{t \in T} E|X_t| \stackrel{(a)}{<} \infty$$



and let  $a_0 = M/\delta$ . Then

$$P(|X_t| \geq a_0) \leq \frac{E|X_t|}{a_0} \leq \frac{M}{a_0} = \delta,$$

so by (b),

$$\sup_{s \in T} \int_{|X_t| \geq a_0} |X_s| dP < \epsilon$$

holds for any  $t \in T$ . It implies that

$$\sup_{t \in T} \int_{|X_t| \geq a_0} |X_t| dP \leq \sup_{t \in T} \sup_{s \in T} \int_{|X_t| \geq a_0} |X_s| dP < \epsilon.$$

Now for any  $a \geq a_0$ ,

$$\sup_{t \in T} \int_{|X_t| \geq a} |X_t| dP < \epsilon$$

holds, i.e.,

$$\sup_{t \in T} \int_{|X_t| \geq a} |X_t| dP \xrightarrow{a \rightarrow \infty} 0.$$

□

Recall that, even if  $X_n \xrightarrow{a.s.} X$ , we cannot guarantee that  $X_n \xrightarrow{\mathcal{L}^1} X$ , or,  $EX_n \not\rightarrow EX$ . (See example 2.3.18) However, with uniform integrability, we can say that convergence in probability is equivalent to  $\mathcal{L}^1$ -convergence.

**Theorem 2.4.4** (Vitali's Lemma). *Suppose that  $X_n \xrightarrow{P} X$ , and  $X_n \in \mathcal{L}^r$  for  $r \geq 1$ . Then TFAE.*

- (i)  $\{|X_n|^r : n \geq 1\}$  is uniformly integrable.
- (ii)  $X_n \xrightarrow{\mathcal{L}^r} X$ , i.e.,  $E|X_n - X|^r \xrightarrow{n \rightarrow \infty} 0$ .
- (iii)  $E|X_n|^r \xrightarrow{n \rightarrow \infty} E|X|^r$ .

To show this, we need some basic properties of uniform integrable sequences.

**Lemma 2.4.5.** (a) *If  $\{X_n\}$  and  $\{Y_n\}$  are both uniformly integrable, then so is  $\{X_n + Y_n\}$ .*

(b) *If  $\{X_n\}$  is uniformly integrable and  $|Y_n| \leq |X_n|$ , then  $\{Y_n\}$  is also uniformly integrable.*

*Proof of lemma.* (a) We get the result from

$$\sup_n \int_{|X_n + Y_n| \geq a} |X_n + Y_n| dP \leq \sup_n \int_{|X_n| + |Y_n| \geq a} |X_n + Y_n| dP$$

$$\begin{aligned}
&\leq \sup_n \left( \int_{\substack{|X_n|+|Y_n|\geq a \\ |X_n|\geq|Y_n|}} (|X_n| + |Y_n|) dP + \int_{\substack{|X_n|+|Y_n|\geq a \\ |X_n|<|Y_n|}} (|X_n| + |Y_n|) dP \right) \\
&\leq \sup_n \left( \int_{2|X_n|\geq a} 2|X_n| dP + \int_{2|Y_n|\geq a} 2|Y_n| dP \right) \\
&\leq \sup_n \int_{|X_n|\geq a/2} 2|X_n| dP + \sup_n \int_{|Y_n|\geq a/2} 2|Y_n| dP \\
&\xrightarrow{a\rightarrow\infty} 0.
\end{aligned}$$

(b) Clear. □

*Proof.* (i) $\Rightarrow$ (ii): Since  $X_n \xrightarrow[n\rightarrow\infty]{P} X$ ,  $\exists\{n'\} \subseteq \{n\}$  s.t.  $X_{n'} \xrightarrow[n'\rightarrow\infty]{a.s.} X$ . Then by Fatou's lemma,

$$E|X|^r \leq \liminf_{n'\rightarrow\infty} E|X_{n'}|^r \leq \sup_n E|X_n|^r < \infty,$$

so  $X \in \mathcal{L}^r$ . Now from

$$|X_n - X|^r \leq 2^r(|X_n|^r + |X|^r),$$

( $\because |a+b|^r \leq 2^r|a|^r$  if  $|a| \geq |b|$ , and  $|a+b|^r \leq 2^r|b|^r$  otherwise, so  $|a+b|^r \leq 2^r(|a|^r + |b|^r)$ )  
 $\{|X_n - X|^r : n \geq 1\}$  is uniformly integrable. Thus,  $\forall \epsilon > 0 \exists \delta > 0$  s.t.

$$A \in \mathcal{F}, P(A) < \delta \Rightarrow \int_A |X_n - X|^r dP < \epsilon.$$

Now by assumption,  $|X_n - X|^r \xrightarrow[n\rightarrow\infty]{P} 0$ , and hence  $\exists N$  s.t.  $P(|X_n - X|^r > \epsilon) \leq \delta$  for any  $n > N$ .

Now

$$n \geq N \Rightarrow E|X_n - X|^r = \underbrace{E|X_n - X|^r I(|X_n - X|^r > \epsilon)}_{\leq \epsilon \text{ (:U.I.)}} + \underbrace{E|X_n - X|^r I(|X_n - X|^r \leq \epsilon)}_{\leq \epsilon} \leq 2\epsilon$$

holds, i.e.,

$$E|X_n - X|^r \xrightarrow[n\rightarrow\infty]{} 0.$$

(ii) $\Rightarrow$ (iii): Let  $\|X\|_r = (E|X|^r)^{1/r}$ . Then by Minkowski inequality,

$$|\|X\|_r - \|X_n\|_r| \leq \|X - X_n\|_r \xrightarrow[n\rightarrow\infty]{} 0$$

holds, i.e.,  $\|X_n\|_r \rightarrow \|X\|_r$ . It implies  $E|X_n|^r \xrightarrow[n\rightarrow\infty]{} E|X|^r$ .

(iii) $\Rightarrow$ (i): We can find infinitely many  $a > 0$  s.t.  $P(|X|^r = a) = 0$ . Since  $X_n \xrightarrow[n\rightarrow\infty]{P} X$ , if one

can show

$$I(|X_n|^r \leq a) \xrightarrow[n \rightarrow \infty]{P} I(|X|^r \leq a),$$

then we get

$$|X_n|^r I(|X_n|^r \leq a) \xrightarrow[n \rightarrow \infty]{P} |X|^r I(|X|^r \leq a).$$

**Claim.**  $I(|X_n|^r \leq a) \xrightarrow[n \rightarrow \infty]{P} I(|X|^r \leq a)$ .

Let  $a_n = P(|I(|X_n|^r \leq a) - I(|X|^r \leq a)| > \epsilon)$ . Then for small  $\epsilon$ ,

$$\begin{aligned} a_n &= P(|I(|X_n|^r \leq a) - I(|X|^r \leq a)| > \epsilon) \\ &= P(|X_n|^r \leq a, |X|^r > a) + P(|X_n|^r > a, |X|^r \leq a) \\ &= P(|X_n|^r \leq a, |X|^r > a + \delta) + P(|X_n|^r \leq a, a < |X|^r \leq a + \delta) \\ &\quad + P(|X_n|^r > a, |X|^r \leq a - \delta) + P(|X_n|^r > a, a - \delta < |X|^r \leq a) \\ &\leq P(|X_n|^r - |X|^r > \delta) + P(a < |X|^r \leq a + \delta) \\ &\quad + P(|X_n|^r - |X|^r > \delta) + P(a - \delta < |X|^r \leq a) \\ &= 2 \underbrace{P(|X_n|^r - |X|^r > \delta)}_{\xrightarrow[n \rightarrow \infty]{P} 0} + P(a - \delta < |X|^r \leq a + \delta) \end{aligned}$$

holds, for arbitrary  $\delta > 0$ . Thus we get

$$0 \leq \limsup_{n \rightarrow \infty} a_n \leq P(a - \delta < |X|^r \leq a + \delta),$$

and letting  $\delta \searrow 0$ , we get

$$0 \leq \limsup_{n \rightarrow \infty} a_n \leq P(|X|^r = a) = 0,$$

i.e.,  $a_n \xrightarrow[n \rightarrow \infty]{} 0$ .  $\square$  (Claim) Now,

i)  $\{|X_n|^r I(|X_n|^r \leq a) : n \geq 1\}$  is the collection of bounded random variables, so it is uniformly integrable.

ii)  $|X_n|^r I(|X_n|^r \leq a) \xrightarrow[n \rightarrow \infty]{P} |X|^r I(|X|^r \leq a)$ .

So by (i)  $\Rightarrow$  (iii) of this theorem, we get

$$E|X_n|^r I(|X_n|^r \leq a) \xrightarrow[n \rightarrow \infty]{} E|X|^r I(|X|^r \leq a)$$

holds. The assumption says  $E|X_n|^r \rightarrow E|X|^r$ , so

$$E|X_n|^r I(|X_n|^r > a) \xrightarrow{n \rightarrow \infty} E|X|^r I(|X|^r > a)$$

holds. Since such  $a$  is uncountably many, for any  $\epsilon > 0$ , we can choose  $a_0$  s.t.  $E|X|^r I(|X|^r > a_0) < \epsilon/2$ , and then we can find  $n > N$  s.t.

$$a \geq a_0 \Rightarrow E|X_n|^r I(|X_n|^r > a) \leq E|X_n|^r I(|X_n|^r > a_0) \leq \epsilon \quad \forall n > N.$$

Now let  $a_1, \dots, a_N$  be s.t.

$$E|X_n|^r I(|X_n|^r > a_n) \leq \epsilon \quad \text{for } n = 1, 2, \dots, N,$$

and  $a^* = \max(a_0, a_1, \dots, a_N)$ . Then,

$$a \geq a^* \Rightarrow E|X_n|^r I(|X_n|^r > a) \leq E|X_n|^r I(|X_n|^r > a^*) \leq \epsilon$$

holds for any  $n \geq 1$ , which implies

$$\sup_n E|X_n|^r I(|X_n|^r > a) \leq \epsilon \quad \forall a \geq a^*.$$

Since  $\epsilon > 0$  was arbitrary, we get

$$\sup_n E|X_n|^r I(|X_n|^r > a) \xrightarrow{a \rightarrow \infty} 0.$$

□

**Corollary 2.4.6.** *Let  $X_n \xrightarrow[n \rightarrow \infty]{d} X$  and  $\{X_n : n \geq 1\}$  be uniformly integrable. Then  $E|X_n| \rightarrow E|X|$  and  $EX_n \rightarrow EX$  as  $n \rightarrow \infty$ .*

*Proof.* By Skorohod theorem, we can find a probability space  $(\Omega', \mathcal{F}', P')$  and r.v.'s on this new probability space  $X'_n$  and  $X'$ , such that

$$X'_n \stackrel{d}{=} X_n, X' \stackrel{d}{=} X, \text{ and } X'_n \xrightarrow[n \rightarrow \infty]{} X' \quad P' - \text{a.s.}$$

Then

$$\sup_n E'|X'_n| I(|X'_n| \geq a) = \sup_n E|X_n| I(|X_n| \geq a)$$

holds, so  $\{X'_n : n \geq 1\}$  is uniformly integrable. Then we get

$$E'|X'_n| \rightarrow E'|X'|$$

and

$$E'|X'_n - X'| \xrightarrow{n \rightarrow \infty} 0,$$

which implies

$$E|X_n| \rightarrow E|X| \text{ and } EX_n \rightarrow EX \text{ as } n \rightarrow \infty.$$

□

### 2.4.1 Uniform integrable martingales

Now back to the martingale theory.

**Definition 2.4.7.** (1) A martingale  $(X_n, \mathcal{F}_n)_{n \geq 0}$  is said to be **regular** if  $\exists X \in \mathcal{L}^1$  s.t.  $X_n = E(X|\mathcal{F}_n)$  ( $P$ -a.s.).

(2) A martingale  $(X_n, (\mathcal{F}_n))_{n \geq 0}$  is said to be **closable** if  $\exists X_\infty \in \mathcal{L}^1$  s.t.  $X_\infty$  is  $\mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$ -mb and  $(X_n, \mathcal{F}_n)_{0 \leq n \leq \infty}$  becomes a martingale, i.e.,

$$E(X_\infty|\mathcal{F}_n) = X_n \quad \forall n \geq 0.$$

Note that, closable martingale is obviously regular. However, regular martingale may not be closable, because such  $X$  need not be  $\mathcal{F}_\infty$ -measurable. Nevertheless, under uniform integrability, we get equivalence of both conditions.

**Theorem 2.4.8.** Let  $\{X_n\}$  be a martingale. Then TFAE.

(i)  $\{X_n\}$  is regular.

(ii)  $\{X_n\}$  is uniformly integrable, and it converges a.s. to some  $X$ .

(iii)  $X_n$  converges in  $\mathcal{L}^1$ , i.e.,  $E|X_n - X| \rightarrow 0$ .

(iv)  $\{X_n\}$  is closable martingale, i.e.,  $E(X_\infty|\mathcal{F}_n) = X_n$  where  $X_\infty = \lim X_n$  a.s.

*Proof.* (ii)  $\Rightarrow$  (iii) : Vitali's lemma.

(iv)  $\Rightarrow$  (i) : Definition.

(i)  $\Rightarrow$  (ii) : Since  $X_n$  is regular, we can write  $X_n = E(X|\mathcal{F}_n)$  for some  $X \in \mathcal{L}^1$ . First, from

$$|X_n| = |E(X|\mathcal{F}_n)| \leq E(|X||\mathcal{F}_n),$$

we get

$$E|X_n| \leq E|X|,$$

and hence

$$\sup_n E|X_n| < \infty.$$

Next, since  $(|X_n| \geq a) \in \mathcal{F}_n$ , by the definition of conditional expectation,

$$\begin{aligned} \int_{|X_n| \geq a} |X_n| dP &\leq \int_{|X_n| \geq a} E(|X||\mathcal{F}_n) dP \\ &= \int_{|X_n| \geq a} |X| dP \\ &= \int_{|X_n| \geq a, |X| \leq b} |X| dP + \int_{|X_n| \geq a, |X| > b} |X| dP \\ &\leq bP(|X_n| \geq a) + \int_{|X| > b} |X| dP \\ &\leq \frac{b}{a} E|X_n| + \int_{|X| > b} |X| dP \\ &\leq \frac{b}{a} E|X| + \int_{|X| > b} |X| dP \end{aligned}$$

holds for any  $b > 0$ , and hence

$$\sup_n \int_{|X_n| \geq a} |X_n| dP \leq \frac{b}{a} E|X| + \int_{|X| > b} |X| dP$$

also holds. Letting  $a \rightarrow \infty$ , we get

$$\limsup_{a \rightarrow \infty} \sup_n \int_{|X_n| \geq a} |X_n| dP \leq \int_{|X| > b} |X| dP.$$

Since  $b > 0$  was arbitrary, letting  $b \rightarrow \infty$ , by integrability of  $X$ , we get

$$\limsup_{a \rightarrow \infty} \sup_n \int_{|X_n| \geq a} |X_n| dP = 0.$$

Therefore  $\{X_n\}$  is uniformly integrable. An a.s. convergence comes from martingale convergence theorem, since  $\sup_n E|X_n| < \infty$ .

(iii)  $\Rightarrow$  (iv) : Suppose that  $E|X_n - X| \rightarrow 0$  as  $n \rightarrow \infty$ , for some  $X$ . It means that

$$\forall \epsilon > 0, \exists N \text{ s.t. } n \geq N \Rightarrow E|X_n - X| \leq \epsilon,$$

and hence

$$E|X_n| \leq E|X| + \epsilon,$$

i.e.,  $\sup_n E|X_n| < \infty$ . Then by martingale convergence theorem,  $\exists X_\infty$  s.t.

$$X_n \xrightarrow[n \rightarrow \infty]{a.s.} X_\infty.$$

Now,

$$\text{i) } X_n \xrightarrow[n \rightarrow \infty]{P} X_\infty.$$

$$\text{ii) } X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}^1} X \text{ implies } X_n \xrightarrow[n \rightarrow \infty]{P} X.$$

Two things imply that  $X_\infty = X$  a.s., and thus

$$E|X_n - X_\infty| \rightarrow 0.$$

Now, for any  $m \geq n$ ,

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

holds, and letting  $m \rightarrow \infty$ , we get

$$E|E(X_\infty|\mathcal{F}_n) - X_n| \leq \lim_{m \rightarrow \infty} E|X_\infty - X_m| = 0.$$

The last equality is from  $\mathcal{L}^1$ -convergence. Therefore,

$$E|E(X_\infty|\mathcal{F}_n) - X_n| = 0,$$

which implies

$$E(X_\infty|\mathcal{F}_n) = X_n \text{ } P - \text{a.s..}$$

Note that  $X_\infty \in \mathcal{F}_\infty$  comes from a.s.-convergence. □

**Corollary 2.4.9** (Lévy). *If  $X \in \mathcal{L}^1$  and  $(\mathcal{F}_n)_{n \geq 0}$  is a filtration, then*

$$E(X|\mathcal{F}_n) \xrightarrow{n \rightarrow \infty} E(X|\mathcal{F}_\infty) \text{ } P\text{-a.s., and in } \mathcal{L}^1,$$

where

$$\mathcal{F}_\infty = \sigma \left( \bigcup_{n=0}^{\infty} \mathcal{F}_n \right).$$

**Remark 2.4.10.** Now we will denote

$$\bigvee_{n=0}^{\infty} \mathcal{F}_n := \sigma \left( \bigcup_{n=0}^{\infty} \mathcal{F}_n \right).$$

*Proof.* Let  $X_n = E(X|\mathcal{F}_n)$ . Then  $\{X_n\}$  is regular, and so by theorem 2.4.8,  $\exists X_\infty$  s.t.

$$X_n \xrightarrow{n \rightarrow \infty} X_\infty \text{ } P\text{-a.s., and in } \mathcal{L}^1,$$

and  $X_n = E(X_\infty|\mathcal{F}_n)$  almost surely. Now, for any  $A \in \mathcal{F}_n$ ,

$$\int_A X_\infty dP = \int_A E(X_\infty|\mathcal{F}_n) dP = \int_A X_n dP = \int_A E(X|\mathcal{F}_n) dP = \int_A X dP$$

holds, for arbitrarily given  $n$ . Thus, we get

$$\underbrace{\bigcup_{n=0}^{\infty} \mathcal{F}_n}_{\pi\text{-sys}} \subseteq \underbrace{\left\{ A : \int_A X_\infty dP = \int_A X dP \right\}}_{\lambda\text{-sys}}.$$

Note that

$$\bigcup_{n=0}^{\infty} \mathcal{F}_n$$

is a  $\pi$ -system, and using

$$EX = EX_\infty, \text{ i.e., } \Omega \in \left\{ A : \int_A X_\infty dP = \int_A X dP \right\},$$

we can easily get that

$$\left\{ A : \int_A X_\infty dP = \int_A X dP \right\}$$



is a  $\lambda$ -system. Thus by Dynkin's theorem,

$$\bigvee_{n=0}^{\infty} \mathcal{F}_n \subseteq \left\{ A : \int_A X_{\infty} dP = \int_A X dP \right\},$$

so

$$\forall A \in \mathcal{F}_{\infty} \quad \int_A X_{\infty} dP = \int_A X dP.$$

Now, by  $X_{\infty} \in \mathcal{F}_{\infty}$ , we get

$$X_{\infty} = E(X|\mathcal{F}_{\infty}),$$

by definition of conditional expectation. □

Following is the another version of “dominated convergence theorem.”

**Theorem 2.4.11.** *If  $Y_n \xrightarrow[n \rightarrow \infty]{a.s.} Y$ , and  $\exists Z \in \mathcal{L}^1$  s.t.  $|Y_n| \leq Z \forall n$ , then*

$$E(Y_n|\mathcal{F}_n) \xrightarrow[n \rightarrow \infty]{a.s.} E(Y|\mathcal{F}_{\infty}).$$

*Proof.* Let  $W_n = \sup_{k,l \geq n} |Y_k - Y_l|$ . Then,

i)  $0 \leq W_n \leq 2Z$ .

ii)  $W_n$  “monotonely” (sup) “converges to 0” ( $\{Y_n\}$  is pathwise Cauchy)

Thus  $W_n \searrow 0$  as  $n \nearrow \infty$ . Now note that,

$$|Y_n - Y| \leq |Y_n - Y_m| + |Y_m - Y| \leq W_m + |Y_m - Y|$$

for any  $m \leq n$ , and letting  $n \rightarrow \infty$ ,

$$\limsup_{n \rightarrow \infty} E(|Y_n - Y||\mathcal{F}_n) \leq \lim_{n \rightarrow \infty} E(W_m|\mathcal{F}_n) + \lim_{n \rightarrow \infty} E(|Y_m - Y||\mathcal{F}_n) \stackrel{\text{Lévy}}{=} E(W_m|\mathcal{F}_{\infty}) + E(|Y_m - Y||\mathcal{F}_{\infty})$$

for any  $m$ . Note that,

$$0 \leq E(W_m|\mathcal{F}_{\infty}) + E(|Y_m - Y||\mathcal{F}_{\infty}) \leq 4E(Z|\mathcal{F}_{\infty}),$$

and  $E(Z|\mathcal{F}_{\infty})$  is integrable. Therefore, by DCT, we get

$$\lim_{m \rightarrow \infty} (E(W_m|\mathcal{F}_{\infty}) + E(|Y_m - Y||\mathcal{F}_{\infty})) = 0,$$

i.e.,

$$\limsup_{n \rightarrow \infty} E(|Y_n - Y| | \mathcal{F}_n) = 0.$$

It implies that

$$E(Y_n - Y | \mathcal{F}_n) \xrightarrow[n \rightarrow \infty]{a.s.} 0,$$

i.e.,

$$\lim_{n \rightarrow \infty} E(Y_n | \mathcal{F}_n) = \lim_{n \rightarrow \infty} E(Y | \mathcal{F}_n) \stackrel{\text{Lévy}}{=} E(Y | \mathcal{F}_\infty),$$

which is the desired result.  $\square$

### 2.4.2 Riesz Decomposition

**Definition 2.4.12.** A nonnegative supermartingale  $X_n$  is **potential** if  $EX_n \rightarrow 0$ .

**Remark 2.4.13.** (i) A potential supermartingale  $(X_n)$ , indeed, converges to 0 a.s.. By martingale convergence theorem,

$$X_n \xrightarrow[n \rightarrow \infty]{a.s.} X_\infty$$

for some  $X_\infty$ , and then by Fatou's lemma,

$$EX_\infty \leq \liminf_{n \rightarrow \infty} EX_n = 0$$

holds. Nonnegativity yields  $X_\infty = 0$  a.s.

(ii) Further,  $\{X_n\}$  is uniformly integrable. By potentiality,  $\forall \epsilon > 0, \exists N$  s.t.

$$n > N \Rightarrow EX_n \leq \epsilon.$$

Since  $N$  is finite,  $\exists a_0$  s.t.

$$a \geq a_0 \Rightarrow \sup_{n \leq N} EX_n I(|X_n| \geq a) \leq \epsilon,$$

and

$$\sup_{n > N} EX_n I(|X_n| \geq a) \leq \sup_{n > N} EX_n \leq \epsilon.$$

Therefore, we get

$$\sup_n EX_n I(|X_n| \geq a) \leq \epsilon \quad \forall a \geq a_0,$$

which yields

$$\sup_n EX_n I(|X_n| \geq a) \xrightarrow{a \rightarrow \infty} 0.$$

Following theorem shows “Doob-like” decomposition for nonnegative supermartingales, which is called **Riesz decomposition**.

**Theorem 2.4.14** (Riesz Decomposition). *Let  $X_n$  be a nonnegative supermartingale. Then,  $\exists a$  “unique” decomposition*

$$X_n = M_n + V_n,$$

where

i)  $M_n$  is uniformly integrable martingale.

ii)  $V_n$  is a nonnegative supermartingale satisfying  $V_n \xrightarrow[n \rightarrow \infty]{a.s.} 0$ .

*Proof.* (Existence) Note that  $\exists X_\infty$  s.t.  $X_n \xrightarrow[n \rightarrow \infty]{a.s.} X_\infty$ . Put

$$M_n = E(X_\infty | \mathcal{F}_n) \text{ and } V_n = X_n - M_n.$$

Then  $M_n$  is a regular martingale, and hence by theorem 2.4.8,  $\{M_n\}$  is uniformly integrable. Furthermore,

i)  $V_n$  is a supermartingale from

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

ii)  $V_n = X_n - E(X_\infty | \mathcal{F}_n)$  is nonnegative from

$$E(X_\infty | \mathcal{F}_n) \leq \liminf_{m \rightarrow \infty} E(X_m | \mathcal{F}_n) \leq X_n.$$

First inequality is from conditional Fatou (lemma 2.3.20), and second one is from that  $X_n$  is supermartingale.

iii) By Lévy’s theorem,

$$\lim_{n \rightarrow \infty} V_n = X_\infty - E(X_\infty | \mathcal{F}_\infty) = 0.$$

Thus the assertion holds.

(Uniqueness) Let

$$X_n = M_n + V_n = M'_n + V'_n.$$

Then since  $M_n$  is uniformly integrable converging to  $X_\infty$ , by theorem 2.4.8, it is regular, i.e.,

$\exists \eta, \eta'$  s.t.

$$M_n = E(\eta|\mathcal{F}_n), \quad M'_n = E(\eta'|\mathcal{F}_n).$$

Now since  $V_n - V'_n \xrightarrow[n \rightarrow \infty]{a.s.} 0$ ,

$$M'_n - M_n = V_n - V'_n = E(\eta' - \eta|\mathcal{F}_n) \xrightarrow[n \rightarrow \infty]{\text{Lévy}} E(\eta' - \eta|\mathcal{F}_\infty) = 0,$$

and hence  $E(\eta|\mathcal{F}_\infty) = E(\eta'|\mathcal{F}_\infty)$ , i.e.,

$$M_n = E(\eta|\mathcal{F}_n) = E(E(\eta|\mathcal{F}_\infty)|\mathcal{F}_n) = E(E(\eta'|\mathcal{F}_\infty)|\mathcal{F}_n) = E(\eta'|\mathcal{F}_n) = M'_n$$

holds. □

### 2.4.3 Optional Sampling Theorem

**Theorem 2.4.15.** *If  $\{X_n\}$  is uniformly integrable submartingale, and  $N$  is a stopping time, then  $\{X_{N \wedge n}\}$  is also a uniformly integrable submartingale.*

*Proof.*  $(X_{N \wedge n})$  is submartingale from example 2.3.12, and hence uniform integrability is left. Proof will be given step by step.

i) From uniform integrability, we get  $\sup_n E|X_n| < \infty$ , and so by martingale convergence theorem,  $\exists X_\infty$  s.t.

$$X_n \xrightarrow[n \rightarrow \infty]{a.s.} X_\infty.$$

ii) Since  $(X_n)$  is a submartingale, we get

$$(X_n^+): \text{submartingale}, \quad (X_n^-): \text{supermartingale}.$$

Therefore, so are  $(X_{N \wedge n}^+)$  and  $(X_{N \wedge n}^-)$ , respectively, and so by martingale convergence theorem,

$$\sup_n EX_{N \wedge n}^+ \leq \sup_n EX_n^+ < \infty$$

$$\sup_n EX_{N \wedge n}^- \leq EX_0^- < \infty.$$

iii)  $X_{N \wedge n} \xrightarrow[n \rightarrow \infty]{a.s.} X_N$ . It comes from:

$$\text{On } (N < \infty), \quad X_{N \wedge n} \xrightarrow[n \rightarrow \infty]{} X_N.$$

$$\text{On } (N = \infty), X_{N \wedge n} \xrightarrow[n \rightarrow \infty]{} X_\infty = X_N.$$

iv) Then by Fatou's lemma,

$$EX_N^+ \leq \liminf_{n \rightarrow \infty} EX_{N \wedge n}^+ \leq \sup_{n \rightarrow \infty} EX_{N \wedge n}^+ < \infty$$

$$EX_N^- \leq \liminf_{n \rightarrow \infty} EX_{N \wedge n}^- \leq \sup_{n \rightarrow \infty} EX_{N \wedge n}^- < \infty$$

holds, and hence

$$E|X_N| = EX_N^+ + EX_N^- < \infty,$$

i.e.,  $X_N$  is integrable.

v) Therefore, we get uniform integrability, from

$$\begin{aligned} E|X_{N \wedge n}|I(|X_{N \wedge n}| \geq a) &= E|X_{N \wedge n}|I(|X_{N \wedge n}| \geq a, N \leq n) + E|X_{N \wedge n}|I(|X_{N \wedge n}| \geq a, N > n) \\ &= E|X_N|I(|X_N| \geq a, N \leq n) + E|X_n|I(|X_n| \geq a, N > n) \\ &\leq E|X_N|I(|X_N| \geq a) + E|X_n|I(|X_n| \geq a) \end{aligned}$$

and consequently

$$\sup_n E|X_{N \wedge n}|I(|X_{N \wedge n}| \geq a) \leq E|X_N|I(|X_N| \geq a) + \sup_n E|X_n|I(|X_n| \geq a) \xrightarrow[a \rightarrow \infty]{} 0.$$

□

**Theorem 2.4.16.** *If  $X_n$  is a uniformly integrable submartingale, then for any stopping time  $N$ ,*

$$EX_0 \leq EX_N \leq EX_\infty,$$

where

$$X_\infty = \lim_{n \rightarrow \infty} X_n \text{ a.s.}$$

**Remark 2.4.17.** Note that, since  $X_n$  is uniformly integrable, it satisfies  $\sup_n E|X_n| < \infty$ , and hence by martingale convergence theorem, we can define  $X_\infty$ .

*Proof.* We know that  $X_{N \wedge n}$  is uniformly integrable submartingale. so  $X_{N \wedge n}$  converges to  $X_N$ ;

$$X_{N \wedge n} \xrightarrow[n \rightarrow \infty]{a.s.} X_N \text{ if } N < \infty$$

$$X_{N \wedge n} \xrightarrow[n \rightarrow \infty]{a.s.} X_\infty \text{ if } N = \infty,$$

Thus,  $X_{N \wedge n}$  converges  $P$ -a.s. to  $X_N$ . Note that  $X_{N \wedge n}$  converges to  $X_N$  in  $\mathcal{L}^1$ . Since  $N \wedge n$  is bounded stopping time, we get

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

By Vitali lemma, we get

$$EX_{N \wedge n} \xrightarrow[n \rightarrow \infty]{} EX_N, \quad EX_n \xrightarrow[n \rightarrow \infty]{} EX_\infty$$

( $\because |EX_{N \wedge n} - EX_N| \leq E|X_{N \wedge n} - X_N| \rightarrow 0$ ) and therefore

$$EX_0 \leq EX_N \leq EX_\infty.$$

□

Now we reach to our goal.

**Theorem 2.4.18** (Optional Sampling Theorem). *If  $L \leq M$  are stopping times and  $Y_{M \wedge n}$  is a uniformly integrable submartingale (assume that  $Y_\infty$  is well defined), then  $EY_L \leq EY_M$ , and further,*

$$Y_L \leq E(Y_M | \mathcal{F}_L) \text{ } P\text{-a.s.}$$

**Remark 2.4.19.** Note that if  $Y_n$  is uniformly integrable submartingale, then  $Y_{M \wedge n}$  is also a uniformly integrable submartingale, so we can apply this theorem.

*Proof.* Let  $X_n = Y_{M \wedge n}$  be a submartingale. Then by previous theorem, we get

$$EX_L \leq EX_\infty.$$

Note that  $X_L = Y_{M \wedge L} = Y_L$  and  $X_\infty = Y_M$ , and hence

$$EY_L \leq EY_M. \tag{2.2}$$

Now, fix  $A \in \mathcal{F}_L$ , and let

$$N = \begin{cases} L & \text{on } A \\ M & \text{on } A^c. \end{cases}$$

Then  $N = LI_A + MI_{A^c}$  is a stopping time ( $\because (N = n) = ((L = n) \cap A) \cup \underbrace{((M = n) \cap A^c)}_{=(M=n) \cap (L \leq n) \cap A^c} \in \mathcal{F}_n$ ) from  $(L \leq n) \cap A^c \in \mathcal{F}_n$ , and  $L \leq N \leq M$  holds. Thus we get

$$EY_N \leq EY_M$$

by (2.2), and it implies

$$E[Y_N] = E[Y_L I_A + Y_M I_{A^c}] \leq E[Y_M] = E[Y_M I_A + Y_M I_{A^c}],$$

i.e.,

$$EY_L I_A \leq EY_M I_A.$$

Since it holds for any  $A \in \mathcal{F}_L$ , we get

$$\int_A E[Y_M | \mathcal{F}_L] dP = \int_A Y_M dP \geq \int_A Y_L dP \quad \forall A \in \mathcal{F}_L,$$

i.e.,

$$E[Y_M | \mathcal{F}_L] \geq Y_L \text{ a.s..}$$

□

Optional sampling theorem has many applications. In here we see some corollaries, and one example, which is related to random walk.

**Corollary 2.4.20.** *Suppose that  $X_n$  is a submartingale and  $E[|X_{n+1} - X_n| | \mathcal{F}_n] \leq B$   $P$ -a.s.. Then if  $EN < \infty$ ,  $X_{N \wedge n}$  is uniformly integrable and  $EX_N \geq EX_0$ .*

*Proof.* Recall that

$$X_{N \wedge n} = X_0 + \sum_{m=1}^n (X_m - X_{m-1}) I(N \geq m).$$

Thus we get

$$|X_{N \wedge n}| \leq |X_0| + \sum_{m=1}^n |X_m - X_{m-1}| I(N \geq m) =: Z.$$

Note that

$$EZ \leq E|X_0| + E \sum_{m=1}^{\infty} |X_m - X_{m-1}| I(N \geq m)$$

$$\begin{aligned}
&= E|X_0| + \sum_{m=1}^{\infty} E|X_m - X_{m-1}|I(N \geq m) \quad (\text{MCT}) \\
&= E|X_0| + \sum_{m=1}^{\infty} EE[|X_m - X_{m-1}|I(N \geq m)|\mathcal{F}_{m-1}] \\
&= E|X_0| + \sum_{m=1}^{\infty} EE[|X_m - X_{m-1}||\mathcal{F}_{m-1}] I(N \geq m) \\
&\quad (\because I(N \geq m) = 1 - I(N \leq m-1) \in \mathcal{F}_{m-1}) \\
&\leq E|X_0| + \sum_{m=1}^{\infty} BP(N \geq m) \\
&= E|X_0| + BEN \quad (\because EN < \infty)
\end{aligned}$$

holds, so  $EZ < \infty$ , i.e.,  $\{|X_{N \wedge n}| : n \geq 1\}$  is dominated by integrable r.v.  $Z$ . Therefore, we get  $\{X_{N \wedge n}\}$  is uniformly integrable. Now optional sampling theorem gives  $EX_N \geq EX_0$ .  $\square$

**Corollary 2.4.21.** *If  $X_n \geq 0$  is nonnegative supermartingale and  $N$  is a stopping time, then  $EX_0 \geq EX_N$ .*

**Remark 2.4.22.** Note that, by martingale convergence theorem,  $\exists X_\infty \stackrel{a.s.}{=} \lim_n X_n$ .

*Proof.* By bounded optional sampling theorem, we get

$$EX_0 \geq EX_{N \wedge n}.$$

Now using Fatou's lemma, we obtain

$$EX_N \leq \liminf_{n \rightarrow \infty} EX_{N \wedge n} \leq EX_0.$$

$\square$

**Example 2.4.23** (Asymmetric simple random walk.). Let  $\xi_1, \xi_2, \dots$  be i.i.d. random variables s.t.

$$P(\xi_i = 1) = p, \quad P(\xi_i = -1) = q = 1 - p.$$

Define

$$S_n = \xi_1 + \dots + \xi_n, \quad S_0 = 0$$

and

$$\mathcal{F}_n = \sigma(\xi_1, \xi_2, \dots, \xi_n), \quad \mathcal{F}_0 = \{\phi, \Omega\}.$$



(a) If  $0 < p < 1$ , then for  $\varphi(x) = \left(\frac{1-p}{p}\right)^x$ ,  $\varphi(S_n)$  is a martingale.

(b) Let  $T_x = \inf\{n : S_n = x\}$  be “the first time touching  $x$ .” ( $x \in \mathbb{Z}$ ) Then for  $a < 0 < b$ ,

$$P(T_a < T_b) = \frac{\varphi(b) - \varphi(0)}{\varphi(b) - \varphi(a)}.$$

(c) Now further assume that  $1/2 < p < 1$ . If  $a < 0 < b$ , then  $T_b < \infty$   $P$ -a.s., and  $P(\inf_n S_n \leq a) = P(T_a < \infty) = \left(\frac{1-p}{p}\right)^{-a}$ .

(d)  $ET_b = \frac{b}{2p-1}$ .

*Proof.* (a) It comes from

$$\begin{aligned} E[\varphi(S_{n+1})|\mathcal{F}_n] &= E\left[\left(\frac{1-p}{p}\right)^{S_n} \left(\frac{1-p}{p}\right)^{\xi_{n+1}} \middle| \mathcal{F}_n\right] \\ &= \left(\frac{1-p}{p}\right)^{S_n} E\left[\left(\frac{1-p}{p}\right)^{\xi_{n+1}} \middle| \mathcal{F}_n\right] \\ &= \left(\frac{1-p}{p}\right)^{S_n} E\left[\left(\frac{1-p}{p}\right)^{\xi_{n+1}}\right] \\ &= \left(\frac{1-p}{p}\right)^{S_n} \left[\left(\frac{1-p}{p}\right)^{-1} (1-p) + \left(\frac{1-p}{p}\right)^p\right] \\ &= \left(\frac{1-p}{p}\right)^{S_n} = \varphi(S_n). \end{aligned}$$

(b) Let  $N = T_a \wedge T_b$ . For any  $x \in (a, b)$ , we get

$$P(x + S_{b-a} \notin (a, b)) \geq p^{b-a},$$

because  $b - a$  steps of size  $+1$  in a row will take us out of the interval. Similarly

$$P(x + S_{b-a} \notin (a, b)) \geq q^{b-a}.$$

Now, note that  $N = \inf\{n : S_n \notin (a, b)\}$ . Thus we get

$$\begin{aligned} P(N > n(b-a)) &= P(S_{b-a} \in (a, b))P(S_{b-a} + (S_{2(b-a)} - S_{b-a}) \in (a, b)) \cdots \\ &\geq (1 - p^{b-a})(1 - p^{b-a}) \cdots (1 - p^{b-a}) \\ &= (1 - p^{b-a})^n \end{aligned}$$

and hence  $EN < \infty$ , i.e.,  $N < \infty$  a.s.. (Or, you can use the approximation

$$S_n \approx n(p - q) \pm \sigma \sqrt{2n \log \log n}$$

from

$$\limsup_{n \rightarrow \infty} \frac{S_n - n(p - q)}{\sigma \sqrt{2n \log \log n}} = 1$$

and

$$\liminf_{n \rightarrow \infty} \frac{S_n - n(p - q)}{\sigma \sqrt{2n \log \log n}} = -1,$$

where  $\sigma^2 = \text{Var}(\xi_1)$ . Note that  $\lim S_n = \infty$  if  $p > q$ , and  $\lim S_n = -\infty$  if  $p < q$ , so for each case,  $T_b < \infty$  and  $T_a > \infty$  respectively.) Note that  $S_{N \wedge n}$  is between  $a$  and  $b$ , and so  $\varphi(S_{N \wedge n})$  is bounded martingale. Thus,  $\varphi(S_{N \wedge n})$  is uniformly integrable, and hence it is closable. Now also note that

$$\varphi(S_{N \wedge n}) \xrightarrow[n \rightarrow \infty]{a.s.} \varphi(S_N).$$

Note that  $S_N = a$  or  $b$ , and  $T_a = T_b$  is impossible. Thus we get

$$S_N = NI(T_a < T_b) + NI(T_a > T_b) = aI(T_a < T_b) + bI(T_a > T_b).$$

It implies that

$$\begin{aligned} 1 &= \varphi(0) = E\varphi(S_0) = E\varphi(S_N) \\ &= E[\varphi(a)I(T_a < T_b)] + E[\varphi(b)I(T_a > T_b)] \\ &= \varphi(a)P(T_a < T_b) + \varphi(b)P(T_a > T_b) \\ &= (\varphi(a) - \varphi(b))P(T_a < T_b) + \varphi(b) \end{aligned}$$

and hence

$$P(T_a < T_b) = \frac{\varphi(b) - \varphi(0)}{\varphi(b) - \varphi(a)}.$$

(c) We have already shown that

$$T_b < \infty \text{ } P\text{-a.s..}$$

Furthermore, since

$$\forall b < b' \Rightarrow T_b < T_{b'},$$

we get

$$\lim_{b \rightarrow \infty} T_b = \infty,$$

and hence,

$$P(T_a < \infty) = \lim_{b \rightarrow \infty} P(T_a < T_b) = \left( \frac{1-p}{p} \right)^{-a}.$$

Note that we used  $1/2 < p < 1$  to get  $\varphi(b) \rightarrow 0$  as  $b \rightarrow \infty$ .

(d) Note that for any  $a \in \mathbb{Z}$ , we get

$$\left( \inf_n S_n \leq a \right) = (T_a < \infty).$$

Thus, we get

$$P\left(\inf_n S_n \leq a\right) = \begin{cases} \left(\frac{1-p}{p}\right)^{-a} & a < 0 \\ 1 & a \geq 0 \end{cases}$$

and thus

$$\begin{aligned} E \left| \inf_n S_n \right| &= \sum_{a=-\infty}^{\infty} P\left(\inf_n S_n = a\right) |a| \\ &= \sum_{a=-\infty}^0 \left[ \left(\frac{1-p}{p}\right)^{-a} - \left(\frac{1-p}{p}\right)^{-(a-1)} \right] |a| \\ &= \sum_{a=-\infty}^0 \underbrace{\left(\frac{1-p}{p}\right)^{-a}}_{<1} \left(1 - \frac{1-p}{p}\right) |a| < \infty \end{aligned}$$

holds (“power series”). In other words,  $\inf_n S_n$  is integrable. Now put

$$X_n = S_n - (p-q)n,$$

and then  $X_n$  is a martingale ( $\because E\xi_1 = p-q$ ). Thus so is  $X_{T_b \wedge n}$ , and hence by optional sampling theorem,

$$EX_0 = EX_{T_b \wedge n} = E(S_{T_b \wedge n} - (p-q)(T_b \wedge n)).$$

Now note that

$$\inf_n S_n \leq S_{T_b \wedge n} \leq b,$$

so by DCT, we get

$$ES_{T_b \wedge n} \xrightarrow{n \rightarrow \infty} S_{T_b}$$

from

$$S_{T_b \wedge n} \xrightarrow[n \rightarrow \infty]{a.s.} S_{T_b} \quad (\because T_b < \infty).$$

However, by  $T_b < \infty$  again, we get

$$S_{T_b} = b \text{ and } T_b \wedge n \xrightarrow[n \rightarrow \infty]{a.s.} T_b,$$

and therefore,

$$0 = EX_0 = EX_{T_b \wedge n} ES_{T_b \wedge n} - (p - q)E(T_b \wedge n) \xrightarrow[n \rightarrow \infty]{} b - (p - q)ET_b$$

by MCT, which implies that

$$0 = b - (p - q)ET_b,$$

i.e.,

$$ET_b = \frac{b}{p - q}.$$

□

## 2.5 Square integrable Martingales

In this section, we see some special properties of square integrable martingales. In this section, let  $X_n$  be a martingale with  $X_0 = 0$ , and  $EX_n^2 < \infty$  for all  $n$ . Put

$$A_0 \equiv 0, \quad A_n = \sum_{m=1}^n \{E(X_m^2 | \mathcal{F}_{m-1}) - X_{m-1}^2\}.$$

Then since  $E(X_m^2 | \mathcal{F}_{m-1}) - X_{m-1}^2 \geq 0$ ,  $A_n$  is increasing, and it is also predictable. Now let

$$M_n = X_n^2 - A_n.$$

Then

$$\begin{aligned} E(M_n | \mathcal{F}_{n-1}) &= E(X_n^2 | \mathcal{F}_{n-1}) - A_n \\ &= E(X_n^2 | \mathcal{F}_{n-1}) - (A_{n-1} + E(X_n^2 | \mathcal{F}_{n-1}) - X_{n-1}^2) \\ &= X_{n-1}^2 - A_{n-1} \\ &= M_{n-1}, \end{aligned}$$

so  $M_n$  is a martingale. It gives a Doob's decomposition of  $X_n^2$ ,

$$X_n^2 = M_n + A_n.$$

Also note that

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

Now let  $A_\infty = \lim_{n \rightarrow \infty} A_n$  (Note that it is well-defined, because  $A_n$  is increasing). Then we get an upper bound for  $E \sup_m |X_m|^2$ :

**Theorem 2.5.1.**  $E \sup_m |X_m|^2 \leq 4EA_\infty$ .

*Proof.* By Doob's maximal inequality, we get

$$E \left( \sup_{0 \leq m \leq n} |X_m|^2 \right) \leq 4E|X_n|^2 = 4EA_n \quad (\because EM_n = EM_0 = 0),$$

and therefore, by MCT,

$$E \left( \sup_m |X_m|^2 \right) \leq 4EA_\infty.$$

□

Actually, we can obtain more sharp bound.

**Theorem 2.5.2.**  $E \sup_m |X_m| \leq 3EA_\infty^{1/2}$ .

*Proof.* Let

$$N_a = \inf \{n : A_{n+1} > a^2\}$$

be a stopping time  $((N = k) = (A_1 \leq a^2) \cap \dots \cap (A_k \leq a^2) \cap (A_{k+1} > a^2) \in \mathcal{F}_k)$ . Then

$$\begin{aligned} P \left( \sup_{0 \leq m \leq n} |X_m| > a \right) &\leq P(N < \infty) + P \left( \sup_{0 \leq m \leq n} |X_{N \wedge m}| > a, N = \infty \right) \\ &\leq P(A_\infty > a^2) + P \left( \sup_{0 \leq m} |X_{N \wedge m}| > a \right) \\ &(\because (N < \infty) = (\exists n \text{ s.t. } A_{n+1} > a^2) \subseteq (A_\infty > a^2)) \\ &\leq P(A_\infty > a^2) + \lim_{n \rightarrow \infty} P \left( \sup_{0 \leq m \leq n} |X_{N \wedge m}| > a \right) \\ &\quad \underbrace{= P(\sup_{0 \leq m \leq n} |X_{N \wedge m}|^2 > a^2)} \\ &\leq P(A_\infty > a^2) + \lim_{n \rightarrow \infty} a^{-2} EX_{N \wedge n}^2 \quad (\because \text{submtg ineq}) \end{aligned}$$

$$\begin{aligned}
&\leq P(A_\infty > a^2) + \lim_{n \rightarrow \infty} a^{-2} E A_{N \wedge n} \quad (\because EM_{N \wedge n} = 0) \\
&\leq P(A_\infty > a^2) + a^{-2} E(A_\infty \wedge a^2) \quad (\because A_{n \wedge N} \leq A_\infty, A_{n \wedge N} \leq a^2)
\end{aligned}$$

holds. (Check where we used submartingale inequality, optional sampling theorem, and Doob's decomposition!) Therefore, we get

$$\begin{aligned}
E \sup_{0 \leq m \leq n} |X_m| &= \int_0^\infty P \left( \sup_{0 \leq m \leq n} |X_m| > a \right) da \\
&\leq \int_0^\infty \underbrace{P(A_\infty > a^2)}_{=P(A_\infty^{1/2} > a)} da + \int_0^\infty [a^{-2} E(A_\infty \wedge a^2)] da \\
&= EA_\infty^{1/2} + \int_0^\infty a^{-2} \left( \int_0^\infty P(A_\infty \wedge a^2 > b) db \right) da \\
&= EA_\infty^{1/2} + \int_0^\infty a^{-2} \int_0^{a^2} P(A_\infty > b) db da \\
&\stackrel{\text{Fubini}}{=} EA_\infty^{1/2} + \int_0^\infty \int_{\sqrt{b}}^\infty a^{-2} P(A_\infty > b) da db \\
&= EA_\infty^{1/2} + \int_0^\infty b^{-1/2} P(A_\infty > b) db \\
&\stackrel{(*)}{=} EA_\infty^{1/2} + 2EA_\infty^{1/2} = 3EA_\infty^{1/2}.
\end{aligned}$$

Note that in (\*), we used

$$EX^p = \int_0^\infty p y^{p-1} P(X > y) dy$$

when  $X \geq 0$ . Now MCT gives the conclusion,

$$E \sup_m |X_m| \leq 3EA_\infty^{1/2}.$$

□

**Theorem 2.5.3.**  $\lim_{n \rightarrow \infty} X_n$  exists, and it is finite  $P$ -a.s. on  $(A_\infty < \infty)$ .

*Proof.* Let  $a > 0$  and

$$N_a = \inf \{n : A_{n+1} > a^2\}$$

be a stopping time. Then  $X_{N_a \wedge n}$  is a martingale, so by previous theorem, we get

$$E \left[ \sup_n E |X_{N_a \wedge n}|^2 \right] \leq 4E \lim_{n \rightarrow \infty} A_{N_a \wedge n} \leq 4a^2,$$

from  $A_{N_a \wedge n} \leq a^2$ . (Check:  $X_{N_a \wedge n}^2 = M_{N_a \wedge n} + A_{N_a \wedge n}$  is also a Doob's decomposition for

$X_{N_a \wedge n}^2$ ) Thus by (sub)martingale convergence theorem,  $X_{N_a \wedge n}$  converges  $P$ -a.s., and by DCT,  $EX_{N_a \wedge n}^2 \rightarrow EX_{N_a}^2$ . Now, for  $k = 1, 2, \dots$ , let

$$C_k = (X_{N_k \wedge n} \text{ converges}),$$

and

$$C = \bigcap_{k=1}^{\infty} C_k.$$

Then clearly  $P(C_k) = 1$ , and hence,  $P(C) = 1$ . Now, if  $\omega \in C \cap (A_{\infty} < \infty)$ , then from  $A_{\infty}(\omega) < \infty$ ,  $\exists k$  s.t.  $N_k(\omega) = \infty$ . Thus, for such  $k$ ,

$$(X_{N_k \wedge n})(\omega) = X_{N_k(\omega) \wedge n}(\omega) = X_n(\omega)$$

holds, and for almost all  $\omega \in C$ , it should converges. Therefore,

$$X_n(\omega) \text{ converges for almost all } \omega \in C \cap (A_{\infty} < \infty),$$

i.e.,

$$X_n \text{ converges } P\text{-a.s. on } (A_{\infty} < \infty).$$

□

**Theorem 2.5.4.** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be increasing function with  $f \geq 1$  and  $\int_0^{\infty} f(t)^{-2} dt < \infty$ . Then

$$\frac{X_n}{f(A_n)} \xrightarrow[n \rightarrow \infty]{a.s.} 0 \text{ on } (A_{\infty} = \infty).$$

*Proof.* Let  $H_n := 1/f(A_n) \in \mathcal{F}_{n-1}$ . Then  $H_n$  is nonnegative bounded ( $\because 1/f \leq 1$ ) predictable process, so

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

is a martingale, and  $Y_n$  is also square integrable. Thus  $Y_n^2$  is a submartingale, and hence we can find its Doob decomposition

$$Y_n^2 = M_n^y + B_n.$$

Note that

$$B_0 \equiv 0 \text{ and } B_n = \sum_{m=1}^n E[(Y_m - Y_{m-1})^2 | \mathcal{F}_{m-1}]$$

and hence  $B_{n+1} - B_n = E[(Y_{n+1} - Y_n)^2 | \mathcal{F}_n]$ . Since  $B_n$  is increasing, we can define  $B_\infty := \lim_{n \rightarrow \infty} B_n$  and it is obtained as

$$\begin{aligned}
 B_\infty &= \sum_{n=0}^{\infty} (B_{n+1} - B_n) \\
 &= \sum_{n=0}^{\infty} E \left[ \frac{(X_{n+1} - X_n)^2}{f(A_{n+1})^2} \middle| \mathcal{F}_n \right] \\
 &= \sum_{n=0}^{\infty} \frac{E[(X_{n+1} - X_n)^2 | \mathcal{F}_n]}{f(A_{n+1})^2} \\
 &\stackrel{(*)}{=} \sum_{n=0}^{\infty} \frac{A_{n+1} - A_n}{f(A_{n+1})^2} \\
 &= \sum_{n=0}^{\infty} \int_{A_n}^{A_{n+1}} \frac{1}{f(A_{n+1})^2} dt \\
 &\leq \sum_{n=0}^{\infty} \int_{A_n}^{A_{n+1}} \frac{1}{f(t)^2} dt \\
 &= \int_0^\infty \frac{1}{f(t)^2} dt < \infty,
 \end{aligned}$$

so  $B_\infty < \infty$  a.s..  $(*)$  is obtained from  $E[(X_{n+1} - X_n)^2 | \mathcal{F}_n] = E(X_{n+1}^2 | \mathcal{F}_n) - X_n^2 = A_{n+1} - A_n$ . By previous theorem,  $Y_n$  converges and is finite on  $(B_\infty < \infty)$ , so we get

$$Y_n \xrightarrow[n \rightarrow \infty]{a.s.} Y_\infty < \infty.$$

Therefore, on  $(A_\infty = \infty)$ ,  $f(A_n) \nearrow \infty$ , so by Kronecker's lemma, we get

$$\frac{1}{f(A_n)} \sum_{m=1}^n (X_m - X_{m-1}) = \frac{X_n}{f(A_n)} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

□

With this, we can further extend second Borel-Cantelli lemma.

**Corollary 2.5.5** (Second Borel Cantelli Lemma). *Let  $B_n \in \mathcal{F}_n$  be adapted events and  $p_n = P(B_n | \mathcal{F}_{n-1})$  for  $n \geq 1$ . Then*

$$\frac{\sum_{m=1}^n I_{B_m}}{\sum_{m=1}^n p_m} \xrightarrow[n \rightarrow \infty]{a.s.} 1 \text{ on } \left( \sum_{m=1}^{\infty} p_m = \infty \right).$$



*Proof.* Let  $X_0 = 0$  and  $X_n - X_{n-1} = I_{B_n} - P(B_n|\mathcal{F}_{n-1})$ , i.e.,

$$X_n = \sum_{m=1}^n (I_{B_m} - P(B_m|\mathcal{F}_{m-1})).$$

Then  $X_n$  becomes a square-integrable martingale. Note that

$$\frac{\sum_{m=1}^n I_{B_m}}{\sum_{m=1}^n p_m} - 1 = \frac{\sum_{m=1}^n (I_{B_m} - p_m)}{\sum_{m=1}^n p_m} = \frac{\sum_{m=1}^n (I_{B_m} - P(B_m|\mathcal{F}_{m-1}))}{\sum_{m=1}^n p_m} = \frac{X_n}{\sum_{m=1}^n p_m}.$$

Let  $X_n^2 = M_n + A_n$  be a Doob decomposition of  $X_n^2$ . Then we get

$$A_n - A_{n-1} = E[(X_n - X_{n-1})^2|\mathcal{F}_{n-1}] = E[(I_{B_n} - p_n)^2|\mathcal{F}_{n-1}] = p_n - p_n^2 \leq p_n,$$

and hence

$$A_n \leq \sum_{m=1}^n p_m.$$

On  $(A_\infty < \infty)$ ,  $X_n$  converges to a finite limit  $X_\infty$ , and so

$$\text{On } (A_\infty < \infty) \cap \left( \sum_{m=1}^{\infty} p_m = \infty \right), \frac{X_n}{\sum_{m=1}^n p_m} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Meanwhile, on  $(A_\infty = \infty)$ ,  $A_n - A_{n-1} \leq p_n$  implies  $\sum_{n=1}^{\infty} p_n = \infty$ , i.e.,

$$(A_\infty = \infty) = (A_\infty = \infty) \cap \left( \sum_{m=1}^{\infty} p_m = \infty \right),$$

and applying previous theorem to  $f(t) = t \vee 1$ , we get

$$\text{On } (A_\infty = \infty) \cap \left( \sum_{m=1}^{\infty} p_m = \infty \right), \frac{X_n}{\sum_{m=1}^n p_m} \underset{(\star)}{\leq} \frac{X_n}{f(A_n)} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

(( $\star$ ) holds for sufficiently large  $n$  from  $A_n \rightarrow \infty$ ) Therefore, we get the conclusion.  $\square$

## 2.6 Further Topics

### 2.6.1 Square Function Inequalities

Our first goal is to show following *Burkholder's inequality*:

**Theorem 2.6.1** (Burkholder). *If  $S_i$ ,  $i = 1, 2, \dots, n$  is a martingale, then for  $1 < p < \infty$ ,  $\exists C_1$  and  $C_2$ , depending only upon  $p$ , such that*

$$C_1 \cdot E \left| \sum_{i=1}^n X_i^2 \right|^{p/2} \leq E|S_n|^p \leq C_2 \cdot E \left| \sum_{i=1}^n X_i^2 \right|^{p/2},$$

where  $X_i = S_i - S_{i-1}$  and  $S_0 \equiv 0$  (In fact,  $C_1 = (18p^{1/2}q)^{-p}$  and  $C_2 = (18pq^{1/2})^p$ , where  $q$  is a Hölder conjugate of  $p$ ).

**Remark 2.6.2.** In here,  $X_t$  is called “martingale difference.” With this inequality, we can handle martingale with squared sum of martingale difference sequences. For the proof, we need some lemmas.

**Lemma 2.6.3.** *Suppose that  $S_i$ ,  $i = 1, 2, \dots, n$  is a martingale or nonnegative submartingale. Then for  $\lambda > 0$ , defining the stopping time  $\tau$  by*

$$\tau = \inf\{i : 1 \leq i \leq n, |S_i| > \lambda\} \wedge (n+1),$$

we get

$$E \left( \sum_{i=1}^{\tau-1} X_i^2 \right) + E(S_{\tau-1}^2) \leq 2\lambda E|S_n|.$$

In here, we defined as  $X_i = S_i - S_{i-1}$ ,  $S_0 \equiv 0$ ,  $S_{n+1} = S_n$ , and  $X_{n+1} = 0$ .

*Proof.* For any  $m = 1, 2, \dots, n+1$  we get

$$S_{m-1}^2 = (X_1 + \dots + X_{m-1})^2 = \sum_{i=1}^{m-1} X_i^2 + 2 \sum_{1 \leq i < j \leq m-1} X_i X_j$$

and hence

$$\begin{aligned} \sum_{i=1}^{m-1} X_i^2 + S_{m-1}^2 &= 2S_{m-1}^2 - 2 \sum_{1 \leq i < j \leq m-1} X_i X_j \\ &= 2S_{m-1}^2 - 2 \sum_{j=2}^{m-1} \sum_{i=1}^{j-1} X_i X_j \end{aligned}$$

$$\begin{aligned}
&= 2S_{m-1}^2 - 2 \sum_{j=2}^{m-1} S_{j-1}X_j \\
&= 2(S_m - X_m)S_{m-1} - 2 \sum_{j=2}^{m-1} S_{j-1}X_j \\
&= 2S_mS_{m-1} - 2 \sum_{j=2}^m S_{j-1}X_j
\end{aligned}$$

holds, which implies

$$\sum_{i=1}^{\tau-1} X_i^2 + S_{\tau-1}^2 = 2S_\tau S_{\tau-1} - 2 \sum_{j=2}^{\tau} S_{j-1}X_j.$$

Note that, letting  $\mathcal{F}_{n+1} = \mathcal{F}_n$ ,  $(S_i, \mathcal{F}_i)_{i=1,2,\dots,n+1}$  also becomes a martingale or nonnegative submartingale. Thus we get

$$\begin{aligned}
E \left( \sum_{i=2}^{\tau} S_{i-1}X_i \right) &= E \left( \sum_{i=2}^{n+1} I(\tau \geq i) S_{i-1}X_i \right) \\
&= \sum_{i=2}^{n+1} E S_{i-1}X_i I(\tau \geq i) \\
&= \sum_{i=2}^{n+1} E [E(S_{i-1}X_i I(\tau \geq i) | \mathcal{F}_{i-1})] \\
&= \sum_{i=2}^{n+1} E [S_{i-1} E(X_i | \mathcal{F}_{i-1}) I(\tau \geq i)] \\
&\geq 0.
\end{aligned}$$

The last inequality holds from: if  $S_i$  is a martingale, then  $E(X_i | \mathcal{F}_{i-1}) = 0$ , and if  $S_i$  is a nonnegative submartingale, then  $S_{i-1} \geq 0$ ,  $E(X_i | \mathcal{F}_{i-1}) \geq 0$ . From this, we obtain

$$E \left( \sum_{i=1}^{\tau-1} X_i^2 \right) + E(S_{\tau-1}^2) = E(2S_\tau S_{\tau-1}) - \underbrace{2 E \left( \sum_{j=2}^{\tau} S_{j-1}X_j \right)}_{\geq 0} \leq 2E(S_\tau S_{\tau-1}).$$

Note that, by definition,  $|S_{\tau-1}| \leq \lambda$ . This yields that

$$E(S_\tau S_{\tau-1}) \leq E|S_\tau S_{\tau-1}| \leq \lambda E|S_\tau|.$$

Note that  $|S_n|$  is a submartingale in any case ( $S_n$  is martingale or nonnegative submartingale), and by optional sampling theorem,  $E|S_\tau| \leq E|S_{n+1}| = E|S_n|$  (Recall that  $S_{n+1} = S_n$ ). There-

fore, we get

$$E \left( \sum_{i=1}^{\tau-1} X_i^2 \right) + E(S_{\tau-1}^2) \leq 2E(S_{\tau}S_{\tau-1}) \leq 2\lambda E|S_n|.$$

□

**Lemma 2.6.4.** *Let  $S_i$ ,  $i = 1, 2, \dots, n$  be a nonnegative submartingale and define*

$$Y = \max \left( \theta \left( \sum_{i=1}^n X_i^2 \right)^{1/2}, \max_{1 \leq i \leq n} S_i \right) \text{ for } \theta > 0.$$

Then  $\forall \lambda > 0$ ,

$$\lambda P(Y > \beta\lambda) \leq 3ES_n I(Y > \lambda), \quad (2.3)$$

where  $\beta = (1 + 2\theta^2)^{1/2}$ , and for each  $1 < p < \infty$ ,

$$\left\| \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right\|_p \leq 9p^{1/2} q \|S_n\|_p, \quad (2.4)$$

where  $q$  is the Hölder conjugate of  $p$ .

*Proof.* First note that

$$\begin{aligned} \lambda P(Y > \beta\lambda) &= \lambda P \left( \max_{1 \leq i \leq n} S_i > \lambda, Y > \beta\lambda \right) + \lambda P \left( \max_{1 \leq i \leq n} S_i \leq \lambda, Y > \beta\lambda \right) \\ &\leq \underbrace{\lambda P \left( \max_{1 \leq i \leq n} S_i > \lambda \right)}_{:=F_n} + \underbrace{\lambda P \left( \max_{1 \leq i \leq n} S_i \leq \lambda, Y > \beta\lambda \right)}_{:=G_n}. \end{aligned}$$

Letting  $A_i = (S_1 \leq \lambda, \dots, S_{i-1} \leq \lambda, S_i > \lambda)$ , we get

$$\left( \max_{1 \leq i \leq n} S_i > \lambda \right) = \bigcup_{i=1}^n A_i,$$

and so

$$\begin{aligned} F_n &= \lambda P \left( \bigcup_{i=1}^n A_i \right) \\ &= \lambda \sum_{i=1}^n P(A_i) = \sum_{i=1}^n E\lambda I_{A_i} \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^n ES_i I_{A_i} \quad (\because \text{on } A_i, \lambda < S_i) \\
&\leq \sum_{i=1}^n E(E(S_n | \mathcal{F}_i) I_{A_i}) \quad (\because S_i \leq E(S_n | \mathcal{F}_i)) \\
&= \sum_{i=1}^n E(E(S_n I_{A_i} | \mathcal{F}_i)) \\
&= \sum_{i=1}^n ES_n I_{A_i} = ES_n \sum_{i=1}^n I_{A_i} \\
&= ES_n I\left(\max_{1 \leq i \leq n} S_i > \lambda\right) \\
&\leq ES_n I(Y > \lambda) \quad (\because \max S_i \leq Y).
\end{aligned}$$

On the other hand, let's see  $G_n$ . If  $Y > \beta\lambda$ , from  $\beta > 1$ ,  $Y > \lambda$  holds, and by definition of  $Y$ , if further  $\max_{1 \leq i \leq n} S_i \leq \lambda$ , we get  $Y = \theta(\sum X_i^2)^{1/2} > \beta\lambda$ . Thus,

$$G_n = \lambda P\left(\max_{1 \leq i \leq n} S_i \leq \lambda, Y > \beta\lambda\right) \leq \lambda P\left(\max_{1 \leq i \leq n} S_i \leq \lambda, \theta\left(\sum_{i=1}^n X_i^2\right)^{1/2} > \beta\lambda\right).$$

Our goal is to show that  $G_n \leq 2ES_n I(Y > \lambda)$ , so that (2.3) is proven. Let  $T_m$  be a “truncated  $S_m$ ,” which is defined as

$$T_m = S_m I\left(\theta\left(\sum_{i=1}^m X_i^2\right)^{1/2} > \lambda\right).$$

Then from

$$\begin{aligned}
E(T_m | \mathcal{F}_{m-1}) &\geq E\left(S_m I\left(\theta\left(\sum_{i=1}^{m-1} X_i^2\right)^{1/2} > \lambda\right) \middle| \mathcal{F}_{m-1}\right) \\
&= E(S_m | \mathcal{F}_{m-1}) I\left(\theta\left(\sum_{i=1}^{m-1} X_i^2\right)^{1/2} > \lambda\right) \\
&= S_{m-1} I\left(\theta\left(\sum_{i=1}^{m-1} X_i^2\right)^{1/2} > \lambda\right) \\
&= T_{m-1}
\end{aligned}$$

$T_m$  is a nonnegative submartingale. Now define  $Y_1 = T_1$ ,  $Y_2 = T_2 - T_1$ ,  $\dots$ ,  $Y_n = T_n - T_{n-1}$ . Put

$$\mathcal{E}_1 = \left\{ \theta\left(\sum_{i=1}^n X_i^2\right)^{1/2} > \beta\lambda, \max_{1 \leq i \leq n} S_i \leq \lambda \right\}$$

and

$$\mathcal{E}_2 = \left\{ \left( \sum_{i=1}^n Y_i^2 \right)^{1/2} > \lambda, \max_{1 \leq i \leq n} T_i \leq \lambda \right\}.$$

We will show that  $\mathcal{E}_1 \subseteq \mathcal{E}_2$ . Let

$$\nu = \inf \left\{ i : 1 \leq i \leq n, \theta \left( \sum_{j=1}^i X_j^2 \right)^{1/2} > \lambda \right\} \wedge n$$

be a stopping time. On  $\mathcal{E}_1$ , from the definition of  $T_i$ ,  $T_i \leq S_i$  holds, which implies  $\max_i T_i \leq \max_i S_i \leq \lambda$ . Further, on  $\mathcal{E}_1$ , we get

$$\begin{aligned} \beta^2 \lambda^2 &< \theta^2 \sum_{i=1}^n X_i^2 \\ &= \underbrace{\theta^2 \sum_{i=1}^{\nu-1} X_i^2}_{\leq \lambda^2 \text{ (def of } \nu)} + \underbrace{\theta^2 X_\nu^2}_{\leq \lambda^2 (*)} + \underbrace{\theta^2 \sum_{i=\nu+1}^n X_i^2}_{=\sum Y_i^2 (**)} \\ &\leq \lambda^2 + \theta^2 \lambda^2 + \theta^2 \sum_{i=\nu+1}^n Y_i^2 \\ &\leq \lambda^2 + \theta^2 \lambda^2 + \theta^2 \sum_{i=1}^n Y_i^2 \end{aligned}$$

holds ((\*): since  $S_n$  is nonnegative,  $|X_\nu| = |S_\nu - S_{\nu-1}| \leq \max(S_{\nu-1}, S_\nu) \leq \lambda$  on  $\mathcal{E}_1$ ; (\*\*): if  $i \geq \nu$ ,  $S_i = T_i$ , so  $X_i = Y_i$  if  $i \geq \nu + 1$ ). Thus we get

$$\lambda^2 + \theta^2 \lambda^2 + \theta^2 \sum_{i=1}^n Y_i^2 > \beta^2 \lambda^2,$$

i.e.,

$$(1 + 2\theta^2)\lambda^2 + \theta^2 \sum_{i=1}^n Y_i^2 > \beta^2 \lambda^2 + \theta^2 \lambda^2,$$

which implies

$$\sum_{i=1}^n Y_i^2 > \lambda^2.$$

Therefore  $\mathcal{E} \subseteq \mathcal{E}_2$ . Now, under this, we get

$$G_n \leq \lambda P(\mathcal{E}_1) \leq \lambda P(\mathcal{E}_2)$$

$$\begin{aligned}
&= \lambda P \left( \left( \sum_{i=1}^n Y_i^2 \right)^{1/2} > \lambda, \max_{1 \leq i \leq n} T_i \leq \lambda \right) \\
&= \lambda P \left( \left( \sum_{i=1}^n Y_i^2 \right)^{1/2} I \left( \max_{1 \leq i \leq n} T_i \leq \lambda \right) > \lambda \right) \\
&\leq \lambda^{-1} E \left[ \left( \sum_{i=1}^n Y_i^2 \right) I \left( \max_{1 \leq i \leq n} T_i \leq \lambda \right) \right] \\
&\leq \lambda^{-1} E \left[ \left( \sum_{i=1}^n Y_i^2 \right) I \left( \max_{1 \leq j \leq i} T_j \leq \lambda \right) \right].
\end{aligned}$$

Now, let

$$\tau = \inf \{i : 1 \leq i \leq n, |T_i| > \lambda\} \wedge (n+1)$$

be a stopping time. Then if  $\tau \leq n$ , then  $i \geq \tau \Rightarrow I(\max_{1 \leq j \leq i} T_j \leq \lambda) = 0$ , and if  $\tau = n+1$ , then  $\tau - 1 = n$ , so we get

$$E \left[ \left( \sum_{i=1}^n Y_i^2 \right) I \left( \max_{1 \leq j \leq i} T_j \leq \lambda \right) \right] = E \left[ \left( \sum_{i=1}^{\tau-1} Y_i^2 \right) I \left( \max_{1 \leq j \leq i} T_j \leq \lambda \right) \right].$$

Consequently, we get

$$\begin{aligned}
G_n &\leq \lambda^{-1} E \left[ \left( \sum_{i=1}^n Y_i^2 \right) I \left( \max_{1 \leq j \leq i} T_j \leq \lambda \right) \right] \\
&\leq \lambda^{-1} E \left[ \left( \sum_{i=1}^{\tau-1} Y_i^2 \right) I \left( \max_{1 \leq j \leq i} T_j \leq \lambda \right) \right] \\
&\leq \lambda^{-1} E \left[ \left( \sum_{i=1}^{\tau-1} Y_i^2 \right) \right] \\
&\stackrel{(\star)}{\leq} 2ET_n \leq 2ES_n I(Y > \lambda),
\end{aligned}$$

and hence get (2.3). Note that  $(\star)$  comes from lemma 2.6.3;

$$E \left[ \left( \sum_{i=1}^{\tau-1} Y_i^2 \right) \right] \leq E \left[ \left( \sum_{i=1}^{\tau-1} Y_i^2 \right) \right] + E(T_{\tau-1}^2) \leq 2\lambda E|T_n|.$$

For (2.4), note that

$$\begin{aligned}
EY^p &= \int_0^\infty P(Y^p > z) dz \\
&= p\beta^p \int_0^\infty y^{p-1} P(Y > \beta y) dy \quad (z = (\beta y)^p)
\end{aligned}$$

$$\begin{aligned}
& \stackrel{(2.3)}{\leq} 3p\beta^p \int_0^\infty y^{p-2} E S_n I(Y > y) dy \\
& \stackrel{\text{Fubini}}{=} 3p\beta^p E \left[ S_n \int_0^\infty y^{p-2} I(Y > y) dy \right] \\
& = 3p\beta^p E \left[ S_n \int_0^Y y^{p-2} dy \right] \\
& = 3p\beta^p E \left[ S_n \cdot \frac{1}{p-1} Y^{p-1} \right] \\
& = 3q\beta^p E[S_n Y^{p-1}] \quad (q = p/(p-1)) \\
& \stackrel{\text{H\"older}}{\leq} 3q\beta^p (E S_n^p)^{1/p} (E Y^{(p-1)q})^{1/q} \\
& = 3q\beta^p (E S_n^p)^{1/p} (E Y^p)^{1/q}
\end{aligned}$$

holds. Thus we get

$$(E Y^p)^{1/p} \leq 3q\beta^p (E S_n^p)^{1/p},$$

and hence,

$$\theta \left\| \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right\|_p \leq \|Y\|_p \leq 3q\beta^p \|S_n\|_p.$$

Letting  $\theta = p^{-1/2}$ , we get

$$\beta^p = \left( 1 + \frac{2}{p} \right)^{p/2} < e < 3,$$

and hence

$$\left\| \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right\|_p \leq \theta^{-1} \cdot 3q\beta^p \|S_n\|_p < 9p^{1/2} q \|S_n\|_p.$$

□

*Proof of theorem 2.6.1.* Let  $T_i := E(S_n^+ | \mathcal{F}_i)$  and  $U_i := E(S_n^- | \mathcal{F}_i)$  be nonnegative martingales. Also define  $T_0 = U_0 = 0$ . Then for martingale differences  $Y_i := T_i - T_{i-1}$  and  $Z_i := U_i - U_{i-1}$  ( $i \geq 1$ ), we get

$$X_i = E(S_n | \mathcal{F}_i) - E(S_n | \mathcal{F}_{i-1}) = (T_i - U_i) - (T_{i-1} - U_{i-1}) = Y_i - Z_i$$

and hence

$$\left( \sum_{i=1}^n X_i^2 \right)^{1/2} \leq \left( \sum_{i=1}^n Y_i^2 \right)^{1/2} + \left( \sum_{i=1}^n Z_i^2 \right)^{1/2}$$



holds, applying Minkowski inequality to vectors  $(Y_1, \dots, Y_n)^\top$  and  $(Z_1, \dots, Z_n)^\top$ . Hence we get

$$\begin{aligned} \left\| \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right\|_p &\leq \left\| \left( \sum_{i=1}^n Y_i^2 \right)^{1/2} + \left( \sum_{i=1}^n Z_i^2 \right)^{1/2} \right\|_p \\ &\leq \left\| \left( \sum_{i=1}^n Y_i^2 \right)^{1/2} \right\|_p + \left\| \left( \sum_{i=1}^n Z_i^2 \right)^{1/2} \right\|_p \\ &\leq 9p^{1/2}q(\|T_n\|_p + \|U_n\|_p). \end{aligned}$$

In the last inequality, lemma 2.6.4 is used. Now note that

$$0 \leq T_n = S_n^+ \leq |S_n| \text{ and } 0 \leq U_n = S_n^- \leq |S_n|,$$

and hence

$$\|T_n\|_p = (ET_n^p)^{1/p} \leq (E|S_n|^p)^{1/p} = \|S_n\|_p,$$

and similarly

$$\|U_n\|_p \leq \|S_n\|_p.$$

Thus we get

$$\begin{aligned} \left\| \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right\|_p &\leq 9p^{1/2}q(\|T_n\|_p + \|U_n\|_p) \\ &\leq 18p^{1/2}q\|S_n\|_p. \end{aligned} \tag{2.5}$$

This implies

$$\left( 18p^{1/2}q \right)^{-1} \left\| \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right\|_p \leq \|S_n\|_p,$$

i.e.,

$$\left( 18p^{1/2}q \right)^{-p} E \left| \left( \sum_{i=1}^n X_i^2 \right)^{p/2} \right| \leq E|S_n|^p.$$

This is the left hand side. For the right one, consider the *conjugate function* (cf. Royden, p.157)

$$R_n = \operatorname{sgn}(S_n) \frac{|S_n|^{p-1}}{\|S_n\|_p^{p-1}}.$$

Then  $R_i := E(R_n | \mathcal{F}_i)$  is a martingale. Let  $W_1 = R_1$  and  $W_i = R_i - R_{i-1}$  be martingale

differences. Note that

$$\|S_n\|_p = \frac{E|S_n|^p}{\|S_n\|_p^{p-1}} = \frac{E[|S_n|^{p-1} \text{sgn}(S_n) S_n]}{\|S_n\|_p^{p-1}} = E(R_n S_n)$$

holds. Also note that if  $i \neq j$ , WLOG  $i < j$ ,

$$E(W_i X_j) = E[E(W_i X_j | \mathcal{F}_i)] = E \left[ W_i \underbrace{E(X_j | \mathcal{F}_i)}_{=E[S_j - S_{j-1} | \mathcal{F}_i] = 0} \right] = 0.$$

So we get

$$\begin{aligned} \|S_n\|_p &= E(R_n S_n) \\ &= E \left( \sum_{i=1}^n \sum_{j=1}^n W_i X_j \right) \\ &= E \left( \sum_{i=1}^n W_i X_i \right) \\ &\leq E \left[ \left( \sum_{i=1}^n W_i^2 \right)^{1/2} \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right] \quad (\text{Cauchy-Schwarz}) \\ &\leq \left\| \left( \sum_{i=1}^n W_i^2 \right)^{1/2} \right\|_q \left\| \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right\|_p \quad (\text{Hölder}) \\ &\stackrel{(2.5)}{\leq} 18q^{1/2} p \|R_n\|_q \left\| \left( \sum_{i=1}^n X_i^2 \right)^{1/2} \right\|_p, \end{aligned}$$

and consequently, we get

$$\begin{aligned} E|S_n|^p &\leq (18q^{1/2} p)^p (E|R_n|^q)^{p/q} E \left[ \left( \sum_{i=1}^n X_i^2 \right)^{p/2} \right] \\ &= (18q^{1/2} p)^p E \left[ \left( \sum_{i=1}^n X_i^2 \right)^{p/2} \right]. \end{aligned}$$

Note that in the last equality,

$$E|R_n|^q = \frac{E|S_n|^{(p-1)q}}{\|S_n\|_p^{(p-1)q}} = \frac{E|S_n|^p}{\|S_n\|_p^p} = 1$$

is used. □

**Corollary 2.6.5** (Marcinkiewicz-Zygmund). *Suppose that  $X_i$  are i.i.d. r.v's with*

$$E|X_1|^p < \infty, \quad p > 1, \quad \text{and let } S_n = X_1 + \cdots + X_n.$$

*Then*

$$\|S_n\|_p \leq 18q^{1/2}p\sqrt{n}\|X_1\|_p.$$

*Proof.*

$$\begin{aligned} E \left[ \left| \sum_{i=1}^n X_i^2 \right|^{p/2} \right] &= \left\| \sum_{i=1}^n X_i^2 \right\|_{p/2}^{p/2} \\ &\leq \left( \sum_{i=1}^n \|X_i^2\|_{p/2} \right)^{p/2} \\ &= (n\|X_1^2\|_{p/2})^{p/2} \\ &= \sqrt{n}^p E|X_1^2|^{p/2} \\ &= \sqrt{n}^p E|X_1|^p \end{aligned}$$

so by Burkholder inequality,

$$\|S_n\|_p \leq (18q^{1/2}p) \left( E \left[ \left| \sum_{i=1}^n X_i^2 \right|^{p/2} \right] \right)^{1/p} \leq 18q^{1/2}p\sqrt{n}\|X_1\|_p.$$

□

Now we see next result.

**Theorem 2.6.6.** *If  $S_i$ ,  $i = 1, 2, \dots, n$  is a martingale with  $EX_1^2 < \infty$  and  $p > 0$ , then  $\exists C > 0$ , depending only on  $p$ , s.t.*

$$E \left( \max_{1 \leq i \leq n} |S_i|^p \right) \leq C \left\{ E \left[ \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_{i-1}) \right)^{p/2} \right] + E \left( \max_{1 \leq i \leq n} |X_i|^p \right) \right\},$$

*provided that every expected value exists (e.g.  $\max |S_i|^p$  is integrable).*

For proof, we need following lemma.

**Lemma 2.6.7.** *Let  $X, Y$  be nonnegative r.v.'s and suppose that  $\beta > 1$ ,  $\delta > 0$ , and  $\epsilon > 0$  satisfy*

$$\forall \lambda > 0 \quad P(X > \beta\lambda, Y \leq \delta\lambda) \leq \epsilon P(X > \lambda).$$

*Then if  $0 < p < \infty$  and  $\epsilon < \beta^{-p}$ ,*

$$EX^p \leq \beta^p \delta^{-p} (1 - \epsilon \beta^p)^{-1} EY^p.$$

*Proof.* Observe

$$\begin{aligned} P(X > \beta\lambda) &= P(X > \beta\lambda, Y \leq \delta\lambda) + P(X > \beta\lambda, Y > \delta\lambda) \\ &\leq \epsilon P(X > \lambda) + P(Y > \delta\lambda). \end{aligned}$$

Thus we get

$$\begin{aligned} EX^p &= \int_0^\infty p\beta^p \lambda^{p-1} P(X > \beta\lambda) d\lambda \\ &\leq \int_0^\infty p\beta^p \lambda^{p-1} (\epsilon P(X > \lambda) + P(Y > \delta\lambda)) d\lambda \\ &= \epsilon\beta^p \int_0^\infty p\lambda^{p-1} P(X > \lambda) d\lambda + \beta^p \int_0^\infty p\lambda^{p-1} P(Y > \delta\lambda) d\lambda \\ &= \epsilon\beta^p EX^p + \beta^p \delta^{-p} EY^p, \end{aligned}$$

and we get the desired result. □

*Proof of theorem 2.6.6.* Let

$$X = \max_{1 \leq i \leq n} |S_i|$$

and

$$Y = \max \left\{ \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_i) \right)^{1/2}, \max_{1 \leq i \leq n} |X_i| \right\}.$$

Also let

$$\epsilon = \frac{\delta^2}{(\beta - \delta - 1)^2} \quad (\beta > 1, \quad 0 < \delta < \beta - 1)$$

and  $S_0 \equiv 0$ . Also let

$$I_k = I \left( \lambda < \max_{1 \leq i \leq k-1} |S_i| \leq \beta\lambda, \max_{1 \leq i \leq k-1} |X_i| \leq \delta\lambda, \sum_{i=1}^k E(X_i^2 | \mathcal{F}_{i-1}) \leq \delta^2 \lambda^2 \right)$$

and

$$J_k = I \left( \lambda < \max_{1 \leq i \leq k-1} |S_i| \leq \beta \lambda \right)$$

be indicators for  $k = 1, 2, \dots, n$ , and define  $I_0 = J_0 = \phi$ . Then  $I_k, J_k \in \mathcal{F}_{k-1}$ , so

$$T_i = \sum_{k=1}^i X_k I_k \quad (1 \leq i \leq n)$$

and

$$T_i^* = \sum_{k=1}^i X_k J_k \quad (1 \leq i \leq n)$$

are martingales with martingale difference  $X_i I_i$  and  $X_i J_i$ . Next, let

$$\sigma = \inf \{ i : 1 \leq i \leq n, \max_{1 \leq j \leq i} |S_j| > \lambda \} \wedge (n+1)$$

and

$$\tau = \inf \{ i : 1 \leq i \leq n, \max_{1 \leq j \leq i} |S_j| > \beta \lambda \} \wedge (n+1)$$

be stopping times.  $\beta > 1$  implies  $\tau \geq \sigma$ . Note that

$$X > \beta \lambda, Y \leq \delta \lambda \Rightarrow \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_i) \right)^{1/2} \leq \delta \lambda, \max_{1 \leq i \leq n} |X_i| \leq \delta \lambda \Rightarrow I_k = J_k \quad \forall k = 1, 2, \dots, n \Rightarrow T_i = T_i^*.$$

Also note that,

$$X > \beta \lambda, Y \leq \delta \lambda \Rightarrow \sigma, \tau \leq n.$$

Thus we get

$$\begin{aligned} P(X > \beta \lambda, Y \leq \delta \lambda) &\leq P \left( I_k = J_k \quad \forall k = 1, 2, \dots, n, \max_{1 \leq i \leq n} |X_i| \leq \delta \lambda, 1 \leq \sigma \leq \tau \leq n \right) \\ &\leq P \left( \max_{1 \leq i \leq n} |T_i| \geq T_\tau^*, \max_{1 \leq i \leq n} |X_i| \leq \delta \lambda, 1 \leq \sigma \leq \tau \leq n \right) \\ &\leq P \left( \max_{1 \leq i \leq n} |T_i| \geq \left| \sum_{k=1}^{\tau} X_k I_k \left( \max_{1 \leq i \leq k-1} |S_i| \leq \beta \lambda \right) \right| - \left| \sum_{k=1}^{\tau} X_k I_k \left( \max_{1 \leq i \leq k-1} |S_i| \leq \lambda \right) \right|, \right. \\ &\quad \left. \max_{1 \leq i \leq n} |X_i| \leq \delta \lambda, 1 \leq \sigma \leq \tau \leq n \right) \\ &(\because J_k = I(\max |S_i| \leq \beta \lambda) - I(\max |S_i| \leq \lambda), |a - b| \geq |a| - |b|) \\ &\leq P \left( \max_{1 \leq i \leq n} |T_i| \geq |S_\tau| - |S_\sigma|, \max_{1 \leq i \leq n} |X_i| \leq \delta \lambda, 1 \leq \sigma \leq \tau \leq n \right) \end{aligned}$$

$$\begin{aligned}
& \left( \because k \leq \tau \Leftrightarrow I \left( \max_{1 \leq i \leq k-1} |S_i| \leq \beta\lambda \right) = 1 \text{ so } \sum_{k=1}^{\tau} X_k I \left( \max_{1 \leq i \leq k-1} |S_i| \leq \beta\lambda \right) \right. \\
& = \sum_{k=1}^{\tau} X_k = S_{\tau}, \text{ similar for the second term, remark that } \sigma \leq \tau \left. \right) \\
& \leq P \left( \max_{1 \leq i \leq n} |T_i| \geq [(\beta - \delta - 1)\lambda]^2 \right) (\because |S_{\tau}| - |S_{\sigma}| \geq |S_{\tau}| - |S_{\sigma-1}| - |X_{\sigma}|, \\
& |S_{\tau}| > \beta\lambda, |S_{\sigma-1}| \leq \lambda, |X_{\sigma}| \leq \delta\lambda \text{ provided that } \max |X_i| \leq \delta\lambda, \sigma \leq \tau \leq n) \\
& \leq \frac{1}{(\beta - \delta - 1)^2 \lambda^2} ET_n^2 \text{ (submartingale inequality)}
\end{aligned}$$

Also note that, since  $T_i$  is martingale, and  $X_i I_i$  is its martingale difference, we get

$$ET_n^2 = E \sum_{k=1}^n (X_k I_k)^2 = E \sum_{k=1}^n X_k^2 I_k,$$

and hence

$$\begin{aligned}
ET_n^2 &= E \sum_{k=1}^n E(X_k^2 I_k | \mathcal{F}_{k-1}) \\
&= E \sum_{k=1}^n I_k E(X_k^2 | \mathcal{F}_{k-1}).
\end{aligned}$$

Now, note that

$$\begin{aligned}
I_k &= I \left( \lambda < \max_{1 \leq i \leq k-1} |S_i| \leq \beta\lambda, \max_{1 \leq i \leq k-1} |X_i| \leq \delta\lambda, \sum_{i=1}^k E(X_i^2 | \mathcal{F}_{i-1}) \leq \delta^2 \lambda^2 \right) \\
&\leq I \left( \lambda < \max_{1 \leq i \leq k-1} |S_i|, \sum_{i=1}^k E(X_i^2 | \mathcal{F}_{i-1}) \leq \delta^2 \lambda^2 \right) \\
&= I \left( \lambda < \max_{1 \leq i \leq k-1} |S_i| \right) I \left( \sum_{i=1}^k E(X_i^2 | \mathcal{F}_{i-1}) \leq \delta^2 \lambda^2 \right),
\end{aligned}$$

and it yields

$$\begin{aligned}
ET_n^2 &= E \sum_{k=1}^n I_k E(X_k^2 | \mathcal{F}_{k-1}) \\
&\leq E \left[ \sum_{k=1}^n I \left( \lambda < \max_{1 \leq i \leq k-1} |S_i| \right) I \left( \sum_{i=1}^k E(X_i^2 | \mathcal{F}_{i-1}) \leq \delta^2 \lambda^2 \right) E(X_k^2 | \mathcal{F}_{k-1}) \right] \\
&\leq E \left[ I \left( \lambda < \max_{1 \leq i \leq n} |S_i| \right) \sum_{k=1}^n E(X_k^2 | \mathcal{F}_{k-1}) I \left( \sum_{i=1}^k E(X_i^2 | \mathcal{F}_{i-1}) \leq \delta^2 \lambda^2 \right) \right]
\end{aligned}$$

$$\begin{aligned}
&\leq P\left(\max_{1 \leq i \leq n} |S_i| > \lambda\right) \cdot \delta^2 \lambda^2 \\
&(\because \text{let } k_0 \text{ be the maximum } k \text{ s.t. } \sum_{i=1}^k E(X_i^2 | \mathcal{F}_{i-1}) \leq \delta^2 \lambda^2, \text{ then}) \\
&\sum_{k=1}^n E(X_k^2 | \mathcal{F}_{k-1}) I\left(\sum_{i=1}^k E(X_i^2 | \mathcal{F}_{i-1}) \leq \delta^2 \lambda^2\right) = \sum_{k=1}^{k_0} E(X_k^2 | \mathcal{F}_{k-1}) \leq \delta^2 \lambda^2.
\end{aligned}$$

Therefore, we get

$$P(X > \beta\lambda, Y \leq \delta\lambda) \leq \frac{\delta^2 \lambda^2}{(\beta - \delta - 1)^2 \lambda^2} P(X > \lambda),$$

so by lemma 2.6.7, we obtain

$$EX^p \leq CEY^p \leq C \left\{ E \left[ \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_{i-1}) \right)^{p/2} \right] + E \left( \max_{1 \leq i \leq n} |X_i|^p \right) \right\},$$

where

$$C = \beta^p \delta^{-p} (1 - \epsilon \beta^p)^{-1}, \quad \epsilon = \frac{\delta^2}{(\beta - 1 - \delta)^2}.$$

(Recall that  $\max(a, b) \leq a + b$  if  $a, b \geq 0$ ) □

**Theorem 2.6.8** (Rosenthal Inequality). *Let  $S_i$  ( $1 \leq i \leq n$ ) be a martingale and  $2 \leq p < \infty$ .*

*Then  $\exists C_1, C_2$ , which are constants depending only on  $p$ , s.t.*

$$C_1 \left\{ E \left[ \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_{i-1}) \right)^{p/2} \right] + \sum_{i=1}^n E|X_i|^p \right\} \leq E|S_n|^p \leq C_2 \left\{ E \left[ \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_{i-1}) \right)^{p/2} \right] + \sum_{i=1}^n E|X_i|^p \right\},$$

*provided that every expected value exists (e.g.  $|S_n|^p$  is integrable).*

**Lemma 2.6.9.** (a) *For any  $x, y \geq 0$  and  $p \geq 1$ ,*

$$pxy^{p-1} \leq p^p x^p + \left(1 - \frac{1}{p}\right) y^p.$$

(b) *For any  $x_i \in \mathbb{R}$ ,  $i = 1, 2, \dots, n$  and  $p \geq 2$ ,*

$$\sum_{i=1}^n |x_i|^p \leq \left( \sum_{i=1}^n x_i^2 \right)^{p/2}.$$

*Proof.* (a) If  $p = 1$ , it is trivial. If  $p > 1$ , applying Hölder's inequality to  $px$  and  $y^{p-1}$ , we get

$$pxy^{p-1} \leq \frac{(px)^p}{p} + \frac{(y^{p-1})^q}{q} = p^{p-1}x^p + \left(1 - \frac{1}{p}\right)y^p \leq p^p x^p + \left(1 - \frac{1}{p}\right)y^p,$$

using  $q^{-1} = 1 - p^{-1}$  and  $q = p/(p-1)$ .

(b) First, note that  $x^p + y^p \leq (x+y)^p$  if  $x, y \geq 0$  and  $p \geq 1$  (WLOG  $y = 1$ .  $f(x) = (x+1)^p - x^p$  is nondecreasing function on  $x \geq 0$ , so it achieves its minimum at  $x = 0$ , and the minimum is 1). Then by mathematical induction, we get

$$\sum_{i=1}^n x_i^p \leq \left(\sum_{i=1}^n x_i\right)^p,$$

and applying this to  $x_i^2$  and  $p/2$  instead of  $x_i$  and  $p$  ends the proof.  $\square$

**Lemma 2.6.10.** Let  $Z_1, Z_2, \dots$  be nonnegative r.v.'s, and  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_{n-1}$  be sub- $\sigma$ -fields.

Then for any  $p \geq 1$ ,

$$E\left(\sum_{i=1}^n E(Z_i|\mathcal{F}_{i-1})\right)^p \leq p^{p+1} E\left(\sum_{i=1}^n Z_i\right)^p.$$

*Proof.* Let  $A_0 \equiv 0$  and

$$A_k = \sum_{i=1}^k E(Z_i|\mathcal{F}_{i-1}), \quad 1 \leq k \leq n.$$

Then  $A_k \in \mathcal{F}_{i-1}$ . Also note that  $A_k$  is nondecreasing, since  $Z_i \geq 0$ . Thus we get

$$\begin{aligned} A_n^p &= \int_0^{A_n} px^{p-1} dx \\ &= \sum_{k=1}^n \int_{A_{k-1}}^{A_k} px^{p-1} dx \\ &\leq \sum_{k=1}^n \int_{A_{k-1}}^{A_k} pA_k^{p-1} dx \\ &= \sum_{k=1}^n pA_k^{p-1}(A_k - A_{k-1}). \end{aligned}$$

Let  $B_k = A_k^{p-1} - A_{k-1}^{p-1}$ ,  $k = 1, 2, \dots, n$ . Then  $B_k \geq 0$  and  $A_k^{p-1} = \sum_{s=1}^k B_s$ . So we get

$$\begin{aligned} A_n^p &\leq \sum_{k=1}^n pA_k^{p-1}(A_k - A_{k-1}) \\ &= \sum_{k=1}^n p \sum_{s=1}^k B_s (A_k - A_{k-1}) \end{aligned}$$



$$\begin{aligned}
&= p \sum_{s=1}^n B_s \sum_{k=s}^n (A_k - A_{k-1}) \\
&= p \sum_{s=1}^n B_s (A_n - A_{s-1}) \\
&= p \sum_{s=1}^n B_s \left( \sum_{i=s}^n E(Z_i | \mathcal{F}_{i-1}) \right).
\end{aligned}$$

Taking expectation on both hand sides, and using  $B_k \in \mathcal{F}_{k-1}$ , we get

$$\begin{aligned}
EA_n^p &\leq p \sum_{s=1}^n E \left( B_s \sum_{i=s}^n E(Z_i | \mathcal{F}_{i-1}) \right) \\
&= p \sum_{s=1}^n E \left( \sum_{i=s}^n E(B_s Z_i | \mathcal{F}_{i-1}) \right) \\
&= p \sum_{s=1}^n \sum_{i=s}^n E(B_s Z_i) \\
&\leq p E \sum_{s=1}^n \sum_{i=1}^n B_s Z_i \quad (\cdot : B_s, Z_i \geq 0) \\
&= p E \left( \sum_{s=1}^n B_s \right) \left( \sum_{i=1}^n Z_i \right) \\
&= E \left( p A_n^{p-1} \sum_{i=1}^n Z_i \right) \\
&\leq E \left( p^p \left( \sum_{i=1}^n Z_i \right)^p + \left( 1 - \frac{1}{p} \right) A_n^p \right) \quad (\text{lemma 2.6.9}),
\end{aligned}$$

and therefore,

$$\frac{1}{p} EA_n^p \leq p^p E \left( \sum_{i=1}^n Z_i \right)^p$$

holds, which is the desired result.  $\square$

*Proof of theorem 2.6.8.* The right hand side is immediate from theorem 2.6.6, so we only show left hand side. Apply lemma 2.6.10 ( $Z_i = X_i^2$ ,  $p \leftarrow p/2$ ), and we obtain

$$E \left[ \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_{i-1}) \right)^{p/2} \right] \leq K E \left( \sum_{i=1}^n X_i^2 \right)^{p/2},$$

with  $K = (p/2)^{p/2+1}$ . Thus by lemma 2.6.9(b),

$$\begin{aligned} E \left[ \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_{i-1}) \right)^{p/2} \right] + E \left( \sum_{i=1}^n |X_i|^p \right) &\leq K E \left( \sum_{i=1}^n X_i^2 \right)^{p/2} + E \left( \sum_{i=1}^n X_i^2 \right)^{p/2} \\ &= (K+1) E \left[ \left( \sum_{i=1}^n X_i^2 \right)^{p/2} \right] \end{aligned}$$

holds, and hence by Burkholder inequality,

$$E \left[ \left( \sum_{i=1}^n E(X_i^2 | \mathcal{F}_{i-1}) \right)^{p/2} \right] + E \left( \sum_{i=1}^n |X_i|^p \right) \leq C^{-1}(K+1) E \left[ \left( \sum_{i=1}^n X_i^2 \right)^{p/2} \right]$$

is obtained, where  $C^{-1} = (18p^{1/2}q)^p$ . Thus the assertion holds, with desired constant  $C_1 = C/(K+1)$ .  $\square$

### 2.6.2 Application: Uniform integrability of moments of sample mean

**Theorem 2.6.11.** *If  $X_1, X_2, \dots$  are i.i.d. r.v.'s with mean 0 and variance 1, and  $E|X_1|^r < \infty$  for some  $r \geq 2$ , then*

$$\left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right|^r : n \geq 1 \right\}$$

*is uniformly integrable.*

*Proof.* Let's define some notations. Let  $S_n = X_1 + X_2 + \dots + X_n$ . For  $\epsilon > 0$ , let  $M > 0$  be s.t.

$$E|X_1|^r I(|X_1| > M) < \epsilon.$$

Also, let  $X'_k$  be “centered truncated  $X_k$ ,” i.e.,

$$X'_k = X_k I(|X_k| \leq M) - EX_k I(|X_k| \leq M),$$

and  $X''_k = X_k - X'_k$ . Since  $EX_k = 0$ , we can easily obtain that

$$X''_k = X_k I(|X_k| > M) - EX_k I(|X_k| > M).$$

Also note that  $X'_k$  and  $X''_k$  are i.i.d. random variables with mean 0, respectively, and in addition,

$X'_k$  is bounded. Finally, let

$$S'_n = X'_1 + \cdots + X'_n,$$

$$S''_n = X''_1 + \cdots + X''_n.$$

Recall that by Marcinkiewicz-Zygmund inequality, if  $X_1, \dots$  are i.i.d. with  $EX_1 = 0$ ,  $E|X_1|^r < \infty$  for some  $r \geq 2$ , then

$$E|X_1 + \cdots + X_n|^r \leq B_r n^{-r/2} E|X_1|^r$$

for some  $B_r$ . Let's denote such bound  $B_r$  continuously.

**STEP 1.**  $E \left| \frac{S'_n}{\sqrt{n}} \right|^r I \left( \left| \frac{S'_n}{\sqrt{n}} \right| > a \right) \leq \frac{B_{2r}(2M)^{2r}}{a^r}.$

Note that for any r.v.  $U \geq 0$ ,

$$\begin{aligned} EU^r I(U > a) &= \int_a^\infty u^r dF_U(u) \\ &\leq \int_a^\infty \left( \frac{u}{a} \right)^r u^r dF_U(u) \\ &\leq a^{-r} \int_0^\infty u^{2r} dF_U(u) \\ &= a^{-r} EU^{2r} \end{aligned}$$

holds. Thus we get

$$\begin{aligned} E \left| \frac{S'_n}{\sqrt{n}} \right|^r I \left( \left| \frac{S'_n}{\sqrt{n}} \right| > a \right) &\leq a^{-r} E \left| \frac{S'_n}{\sqrt{n}} \right|^{2r} \\ &= (an)^{-r} E|S'_n|^{2r} \\ &\leq (an)^{-r} B_{2r} n^r E|X'_1|^{2r} \text{ (Marcinkiewicz-Zygmund)} \\ &\leq a^{-r} B_{2r} (2M)^{2r} \text{ } (\because |X'_1| \leq 2M). \end{aligned}$$

**STEP 2.**  $E \left| \frac{S''_n}{\sqrt{n}} \right|^r I \left( \left| \frac{S''_n}{\sqrt{n}} \right| > a \right) \leq B_r \cdot 2^r \cdot \epsilon.$

By Marcinkiewicz-Zygmund inequality,

$$\begin{aligned} E \left| \frac{S''_n}{\sqrt{n}} \right|^r &\leq B_r E|X''_1|^r \\ &= B_r E [|X_1 I(|X_1| > M) - EX_1 I(|X_1| > M)|^r] \\ &\leq B_r \cdot E [2^{r-1} \{(|X_1 I(|X_1| > M)|^r + |EX_1 I(|X_1| > M)|^r)\}] \\ &(\because r \geq 1 \Rightarrow (a+b)^r \leq 2^{r-1}(a^r + b^r) \text{ by Jensen}) \end{aligned}$$

$$\begin{aligned}
&= B_r \cdot 2^{r-1} \{E|X_1|^r I(|X_1| > M) + |EX_1 I(|X_1| > M)|^r\} \\
&\leq B_r \cdot 2^r E|X_1|^r I(|X_1| > M) \text{ (Jensen)} \\
&\leq B_r \cdot 2^r \cdot \epsilon \text{ (def of } M\text{)}.
\end{aligned}$$

**STEP 3.** Note that if  $x, y \geq 0$ , then

$$(x + y)I(x + y > a) \leq 2 \max(x, y)I(2 \max(x, y) > a) \leq 2xI(2x > a) + 2yI(2y > a).$$

Using this fact and

$$\left| \frac{S_n}{\sqrt{n}} \right| \leq \left| \frac{S'_n}{\sqrt{n}} \right| + \left| \frac{S''_n}{\sqrt{n}} \right|,$$

we get

$$\begin{aligned}
E \left| \frac{S_n}{\sqrt{n}} \right|^r I \left( \left| \frac{S_n}{\sqrt{n}} \right|^r > a \right) &\leq E 2^{r-1} \left( \left| \frac{S'_n}{\sqrt{n}} \right|^r + \left| \frac{S''_n}{\sqrt{n}} \right|^r \right) I \left( \left| \frac{S_n}{\sqrt{n}} \right|^r > a \right) \\
&\leq 2^r \left\{ E \left| \frac{S'_n}{\sqrt{n}} \right|^r I \left( 2 \left| \frac{S_n}{\sqrt{n}} \right| > a^{1/r} \right) + E \left| \frac{S''_n}{\sqrt{n}} \right|^r I \left( 2 \left| \frac{S_n}{\sqrt{n}} \right| > a^{1/r} \right) \right\} \\
&\leq 2^r \left\{ \frac{B_{2r}(2M)^{2r}}{(a^{1/r}/2)^r} + B_r 2^r \epsilon \right\} \text{ (STEP 1 \& 2)} \\
&= 2^{2r} (a^{-1} B_{2r} (2M)^{2r} + B_r \epsilon).
\end{aligned}$$

Thus, we obtain

$$\limsup_{a \rightarrow \infty} \sup_n E \left| \frac{S_n}{\sqrt{n}} \right|^r I \left( \left| \frac{S_n}{\sqrt{n}} \right|^r > a \right) \leq \lim_{a \rightarrow \infty} 2^{2r} (a^{-1} B_{2r} (2M)^{2r} + B_r \epsilon) = 2^{2r} B_r \epsilon.$$

Since  $\epsilon > 0$  was arbitrary, we get

$$\sup_n E \left| \frac{S_n}{\sqrt{n}} \right|^r I \left( \left| \frac{S_n}{\sqrt{n}} \right|^r > a \right) \xrightarrow{a \rightarrow \infty} 0,$$

which is the desired result. □

**Remark 2.6.12.** Using this theorem and CLT, by Vitali convergence theorem, we can obtain that if  $E|X_1|^r < \infty$ , then

$$E \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right|^r \xrightarrow{n \rightarrow \infty} E|Z|^r,$$

where  $Z \sim N(0, 1)$  (cf. Skorokhod representation theorem).

## Chapter 3

# Infinitely Divisible Distribution

### 3.1 Preliminaries

**Definition 3.1.1.** A distribution  $F$  is called *infinitely divisible* if, for each  $n \geq 1$ ,  $\exists$  a distribution  $F_n$  s.t.

$$F = \underbrace{F_n * F_n * \cdots * F_n}_{n \text{ } F_n \text{'s}},$$

where  $*$  denotes the convolution

$$F * G(t) = \int G(t - \xi) dF(\xi).$$

**Remark 3.1.2.** There are some equivalent definitions.

(a)  $F$  is infinitely divisible if and only if  $\exists(\Omega, \mathcal{F}, P)$  and  $X, X_{n1}, X_{n2}, \dots, X_{nn}$  for any  $n \geq 1$ , s.t.

$$X \sim F, X_{ni} \sim \text{i.i.d. and } X \stackrel{d}{=} X_{n1} + \cdots + X_{nn}.$$

(b)  $F$  is infinitely divisible if and only if for each  $n \geq 1 \exists F_n$  s.t.

$$\varphi_F = (\varphi_{F_n})^n,$$

where  $\varphi_G$  denotes the characteristic function of the distribution  $G$ .

To see some properties of infinitely divisible distributions, we need some preliminaries.

**Definition 3.1.3** (Review of weak convergence). Let  $\mu_n, \mu$  be finite measures on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

$\{\mu_n\}$  is said to **converge to  $\mu$  weakly** if  $\forall a, b$  s.t.  $\mu\{a\} = \mu\{b\} = 0$ ,

$$\mu_n(a, b] \xrightarrow{n \rightarrow \infty} \mu(a, b],$$

and denote it as

$$\mu_n \xrightarrow[n \rightarrow \infty]{w} \mu.$$

**Theorem 3.1.4.** If  $\mu_n \xrightarrow[n \rightarrow \infty]{w} \mu$  and  $\sup_n \mu_n(\mathbb{R}) < \infty$ , then for any continuous function  $f$  with

$$\lim_{|x| \rightarrow \infty} |f(x)| = 0,$$

$$\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu.$$

*Proof.* Note that such  $f$  is always bounded. Choose  $M > 0$  s.t.  $\mu(\mathbb{R}) \leq M$  and  $\mu_n(\mathbb{R}) \leq M$  for any  $n$ . Also, given  $\epsilon > 0$ , choose  $a < b$  s.t.

$$\mu\{a\} = \mu\{b\} = 0 \text{ and } |f(x)| \leq \frac{\epsilon}{M} \quad \forall x \in (a, b]^c.$$

Let  $A := (a, b]$ . Then we get

$$\left| \int_{A^c} f d\mu \right| \leq \int_{A^c} |f| d\mu \leq \frac{\epsilon}{M} \mu(A^c) \leq \frac{\epsilon}{M} \mu(\mathbb{R}) \leq \epsilon,$$

and similarly,

$$\left| \int_{A^c} f d\mu_n \right| \leq \epsilon.$$

CASE 1. If  $\mu(A) = 0$ , then using

$$\left| \int_A f d\mu_n \right| \leq \sup_{x \in A} |f(x)| \mu_n(A),$$

we get

$$\begin{aligned} \left| \int f d\mu_n - \int f d\mu \right| &\leq \left| \int_A f d\mu_n - \int_A f d\mu \right| + \left| \int_{A^c} f d\mu \right| + \left| \int_{A^c} f d\mu_n \right| \\ &= \left| \int_A f d\mu_n \right| + \left| \int_{A^c} f d\mu \right| + \left| \int_{A^c} f d\mu_n \right| \quad (\because \int_A f d\mu = 0) \\ &\leq \sup_{x \in A} |f(x)| \mu_n(A) + \epsilon + \epsilon, \end{aligned}$$

and from  $\mu_n(A) \xrightarrow{n \rightarrow \infty} \mu(A)$ , we get

$$\limsup_{n \rightarrow \infty} \left| \int f d\mu_n - \int f d\mu \right| \leq 2\epsilon.$$

CASE 2. If  $\mu(A) > 0$ , then define probability measures  $\nu_n$  and  $\nu$  as

$$\nu(B) = \frac{\mu(B \cap A)}{\mu(A)}, \quad \nu_n(B) = \frac{\mu_n(B \cap A)}{\mu_n(A)}.$$

(Only consider sufficiently large  $n$ 's, so that  $\nu_n$  can be defined) Then it can be easily shown that

$$\nu_n \xrightarrow[n \rightarrow \infty]{w} \nu,$$

from the fact that  $A$  is an interval. Thus by Portmanteau lemma, for any bounded and (a.e.) continuous function  $g$ ,

$$\int g d\nu_n \xrightarrow{n \rightarrow \infty} \int g d\nu$$

holds (We can say this because  $\nu_n$  and  $\nu$  are probability measure!). So we have

$$\begin{aligned} \int_A f d\mu_n &= \mu_n(A) \int f d\nu_n \text{ (Standard Machine)} \\ &\xrightarrow[n \rightarrow \infty]{} \mu(A) \int f d\nu \text{ (}\because \text{Portmanteau)} \\ &= \int_A f d\mu, \end{aligned}$$

and hence

$$\limsup_{n \rightarrow \infty} \left| \int f d\mu_n - \int f d\mu \right| \leq \limsup_{n \rightarrow \infty} \left| \int_A f d\mu_n - \int_A f d\mu \right| + \left| \int_{A^c} f d\mu \right| + \left| \int_{A^c} f d\mu_n \right| \leq 2\epsilon.$$

No matter which case  $A$  belongs to, we get

$$\limsup_{n \rightarrow \infty} \left| \int f d\mu_n - \int f d\mu \right| \leq 2\epsilon.$$

Since  $\epsilon > 0$  was arbitrarily chosen, we get the conclusion,

$$\lim_{n \rightarrow \infty} \left| \int f d\mu_n - \int f d\mu \right| = 0.$$

□

**Remark 3.1.5. BE CAUTIOUS!** Since the interval  $A$  depends on  $\epsilon$ , we should use the logic such as “arbitrary  $\epsilon$ ” later!

**Proposition 3.1.6.** *For any ch.f.  $\varphi$ , we have*

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

*Proof.* Note that, for a distribution  $F$  with  $\varphi = \varphi_F$ , we have

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

Since it holds for all ch.f.  $\varphi$ , it also holds for ch.f.  $|\varphi|^2 = \varphi \bar{\varphi}$ . Therefore we get

$$\frac{1}{4} \operatorname{Re}(1 - |\varphi(2t)|^2) \leq \operatorname{Re}(1 - |\varphi(t)|^2).$$

Note that both hand sides are all real. □

**Remark 3.1.7.** Let  $\varphi(t) = Ee^{itX}$ . Then  $\bar{\varphi}(t) = Ee^{-itX}$ , so if  $Y$  be a r.v. which is independent of  $X$  and  $-X \stackrel{d}{=} Y$ , then  $|\varphi|^2$  is a ch.f of  $X + Y$ .

**Proposition 3.1.8.** *If  $\varphi$  is infinitely divisible (it means that corresponding distribution is i.d.), then  $\varphi(t) \neq 0$  for any  $t$ .*

*Proof.* Let  $\varphi_n$  be a ch.f s.t.

$$\varphi(t) = (\varphi_n(t))^n.$$

Since  $\varphi(0) = 1$  and  $\varphi$  is continuous,  $\exists a > 0$  s.t.

$$\inf_{|t| \leq a} |\varphi(t)| > 0.$$



It implies that,

$$|t| \leq a \Rightarrow |\varphi_n(t)| = |\varphi(t)|^{1/n} \geq \left( \inf_{|t| \leq a} \varphi(t) \right)^{1/n} \xrightarrow[n \rightarrow \infty]{} 1,$$

so

$$\liminf_{n \rightarrow \infty} |\varphi_n(t)| \geq 1.$$

However,  $|\varphi_n(t)| \leq 1$ , so we get

$$\limsup_{n \rightarrow \infty} |\varphi_n(t)| \leq 1,$$

which yields

$$\lim_{n \rightarrow \infty} |\varphi_n(t)| = 1.$$

Thus for any  $\epsilon \in (0, 1)$ ,  $\exists N = N(\epsilon)$  s.t.

$$n > N \Rightarrow |\varphi_n(t)| > 1 - \epsilon.$$

By proposition 3.1.6, for any  $t$  s.t.  $|t| \leq a$ ,

$$\begin{aligned} 1 - |\varphi_n(2t)|^2 &\leq 4(1 - |\varphi_n(t)|^2) \\ &\leq 4(1 - (1 - \epsilon)^2) \\ &\leq 8\epsilon \end{aligned}$$

for large  $n$  ( $n$  larger than  $N$ ). Thus if we let  $\epsilon < 1/8$ , we get

$$|\varphi_n(2t)|^2 \geq 1 - 8\epsilon > 0$$

for  $|t| \leq a$  and large  $n$ . It means that,

$$|\varphi_n(t)|^2 > 0 \quad \forall |t| \leq 2a$$

for sufficiently large  $n$ . Note that for such large  $n$ , we get

$$|\varphi(t)| = |\varphi_n(t)|^n > 0 \quad \forall |t| \leq 2a.$$

Sine  $[-2a, 2a]$  is compact, we get

$$\inf_{|t| \leq 2a} |\varphi(t)| > 0.$$

Repeat the similar logic, and we get the conclusion.  $\square$

**Remark 3.1.9.** Key point of the proof is that  $\varphi$  is “infinitely” divisible; No matter how many we repeat such procedure, we can always find “very large  $n$ ” s.t.  $\varphi_n(t) \neq 0$  in proper interval and  $\varphi = (\varphi_n)^n$ .

## 3.2 Canonical Representation

**Theorem 3.2.1.** Let  $\mu$  be a finite measure in  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Define

$$\varphi(t) = \exp \left( \int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} d\mu(x) \right). \quad (3.1)$$

Then  $\varphi$  is a ch.f of an infinite divisible distribution with mean 0 and variance  $\mu(\mathbb{R})$ .

**Definition 3.2.2.** (3.1) is called “canonical representation of  $\varphi$ ,” and  $\mu$  is called a “canonical measure.”

*Proof.* CASE 1.  $\mu$  has a mass only at 0, i.e.,  $\mu(\mathbb{R}) = \mu\{0\} \stackrel{\text{let}}{=} \sigma^2 > 0$ .

Then

$$\varphi(t) = \exp \left( -\frac{t^2}{2} \mu\{0\} \right) = \exp \left( -\frac{t^2}{2} \sigma^2 \right)$$

is a ch.f of  $N(0, \sigma^2)$ , which is infinite divisible.

CASE 2.  $\mu$  has a mass only at  $x \neq 0$ , i.e.,  $\mu(\mathbb{R}) = \mu\{x\} \stackrel{\text{let}}{=} \lambda x^2, \lambda > 0$ .

Then

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

is a ch.f of  $x(Poi(\lambda) - \lambda)$ , which is an infinite divisible distribution with mean 0 and variance  $\lambda x^2$ .

CASE 3.  $\mu$  has masses at  $x_1, x_2, \dots, x_k$  s.t.  $\mu\{x_j\} = \delta_j > 0$ .

Let  $\mu_j$  be a measure s.t.  $\mu_j(\mathbb{R}) = \mu_j\{x_j\} = \delta_j$ . Then

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

is a ch.f of infinite divisible distribution with mean 0 and variance  $\delta_j$ , and  $\mu = \mu_1 + \dots + \mu_k$ . So

$$\varphi(t) = \prod_{j=1}^k \varphi_j(t)$$

is also a ch.f of infinite divisible distribution. Let  $X \sim \varphi$ ,  $X_j \stackrel{indep}{\sim} \varphi_j$ . Then

$$X \stackrel{d}{=} X_1 + \cdots + X_k,$$

$EX_j = 0$  and  $var(X_j) = \delta_j$ , so  $EX = 0$  and  $var(X) = \delta_1 + \cdots + \delta_k = \mu(\mathbb{R})$ .

CASE 4. General case.

Define  $\mu_k$  as following:

$$\mu_k\{j \cdot 2^{-k}\} = \mu(j \cdot 2^{-k}, (j+1)2^{-k}], \quad j \in \mathcal{J} := \{0, \pm 1, \pm 2, \dots, \pm 2^{2k}\}.$$

Then

$$\mu_k(\mathbb{R}) = \sum_{j \in \mathcal{J}} \mu_k\{j \cdot 2^{-k}\} = \mu(2^{-k}, 2^k + 2^{-k}] \leq \mu(\mathbb{R}) < \infty$$

so  $\mu_k$  is a finite measure. Now assume that  $f : \mathbb{R} \rightarrow \mathbb{R}$  is continuous and  $\lim_{|x| \rightarrow \infty} |f(x)| = 0$ .

Then

$$\begin{aligned} \int f d\mu_k &= \sum_{j \in \mathcal{J}} f(j \cdot 2^{-k}) \mu_k\{j \cdot 2^{-k}\} \\ &= \sum_{j \in \mathcal{J}} f(j \cdot 2^{-k}) \mu(j \cdot 2^{-k}, (j+1)2^{-k}] \\ &= \int f_k d\mu, \end{aligned}$$

where

$$f_k(x) = \sum_{j \in \mathcal{J}} f(j \cdot 2^{-k}) I_{(j \cdot 2^{-k}, (j+1)2^{-k}]}(x).$$

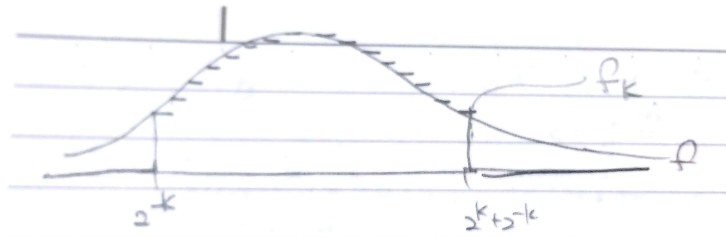


Figure 3.1:  $f$  and  $f_k$

Since  $f$  is bounded, for  $\sup_x |f(x)| = B$ , we get  $|f_k(x)| \leq B < \infty$ , so by BCT,

$$\int f d\mu_k = \int f_k d\mu \xrightarrow{k \rightarrow \infty} \int f d\mu.$$

Now let

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

Since

$$\frac{e^{itx} - 1 - itx}{x^2}$$

is continuous function on  $x$  and decays to 0 as  $|x| \rightarrow \infty$ , so letting  $f(x) = (e^{itx} - 1 - itx)/x^2$ , we get

$$\varphi_k(t) \xrightarrow[k \rightarrow \infty]{} \exp \left( \int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} d\mu(x) \right) =: \varphi(t).$$

Since  $\varphi(0) = 1$ , and  $\varphi$  is continuous at  $t = 0$  ( $\because (e^{itx} - 1 - itx)/x^2 \rightarrow 0$  as  $t \rightarrow 0$ ,  $|(e^{itx} - 1 - itx)/x^2| \leq Kt^2 \leq K$  for some  $K > 0$  if  $t$  is small, so by BCT,

$$\int \frac{e^{itx} - 1 - itx}{x^2} d\mu(x) \rightarrow 0 \text{ as } t \rightarrow 0$$

holds), and thus by Lévy's continuity theorem,  $\varphi$  is a ch.f, and for  $X_k \sim \varphi_k$  and  $X \sim \varphi$ ,

$$X_k \xrightarrow[k \rightarrow \infty]{d} X$$

holds. With Skorokhod theorem, we may assume a.s. convergence. Then we get

$$EX^2 \leq \liminf EX_k^2 \stackrel{\text{Case 3}}{=} \liminf \mu_k(\mathbb{R}) \leq \mu(\mathbb{R}) < \infty,$$

so second moment exists. Note that

$$\varphi'(0) = iEX \text{ and } \varphi''(0) = -EX^2.$$

From

$$\varphi'(t) = \varphi(t) \int \frac{\partial}{\partial t} \frac{e^{itx} - 1 - itx}{x^2} d\mu(x) = \varphi(t) \int \frac{i(e^{itx} - 1)}{x} d\mu(x)$$

and

$$\varphi''(t) = \varphi'(t) \int \frac{i(e^{itx} - 1)}{x} d\mu(x) + \varphi''(t) \int -e^{itx} d\mu(x),$$

we get

$$EX = 0 \text{ and } EX^2 = \text{var}(X) = \mu(\mathbb{R}).$$

Considering  $\mu/n$  instead of  $\mu$ , we get

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

is also a ch.f., and

$$\varphi = (\varphi_n)^n,$$

which implies that  $\varphi$  is infinitely divisible. □

Now our interest is that “when” does i.d. have such canonical representation.

**Definition 3.2.3.** A row-wise independent double array  $\{X_{nk} : n \geq 1, 1 \leq k \leq r_n\}$  is said to satisfy **condition  $\mathcal{R}$**  if

$$(i) \quad EX_{nk} = 0, EX_{nk}^2 = \sigma_{nk}^2 < \infty, s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2 > 0.$$

$$(ii) \quad \sup_n s_n^2 < \infty$$

$$(iii) \quad \max_{1 \leq k \leq r_n} \sigma_{nk}^2 \xrightarrow{n \rightarrow \infty} 0.$$

Also, let  $S_n := X_{n1} + X_{n2} + \cdots + X_{nr_n}$ .

**Lemma 3.2.4.** If  $X, Y$  are independent and  $E(X + Y)^2 < \infty$ , then  $EX^2 + EY^2 < \infty$ .

*Proof.* Note that

$$E|X + Y| = \int \int |x + y| dF_Y(y) dF_X(x) = \int E|Y + x| dF_X(x).$$

If  $E|Y| = \infty$ , then from  $|Y| \leq |Y + x| + |x|$ , we get  $E|Y + x| = \infty \forall x \in \mathbb{R}$ , and hence

$$E|X + Y| = \int E|Y + x| dF_X(x) = \infty,$$

which yields contradiction. Therefore  $E|Y| < \infty$ . By symmetry,  $E|X| < \infty$ . □

**Theorem 3.2.5.** Every infinite divisible distribution (let  $F$ ) with mean 0 and finite variance  $\sigma^2$  is the limit law of  $S_n = X_{n1} + \cdots + X_{nr_n}$  for some row-wise independent double array satisfying “condition  $\mathcal{R}$ .”

*Proof.* For any  $n$ , let  $X_{ni}$  be s.t.

$$X_{n1}, \dots, X_{nn} \sim i.i.d., \quad X_{n1} + \cdots + X_{nn} \sim F.$$

(Such  $\{X_{nk}\}$  exists because  $F$  is infinitely divisible) Then from

$$E(X_{n1} + \cdots + X_{nn})^2 = \sigma^2 < \infty,$$

by lemma,  $EX_{ni}^2 < \infty \forall i = 1, 2, \dots, n$ , and since they are i.i.d.,

$$EX_{nk} = 0, \quad EX_{nk}^2 = \sigma_{nk}^2 = \frac{\sigma^2}{n}.$$

Thus double array  $\{X_{nk} : k = 1, 2, \dots, n\}$  satisfies condition  $\mathcal{R}$ , and

$$S_n \xrightarrow[n \rightarrow \infty]{d} F,$$

from  $S_n \stackrel{d}{=} F \forall n$ . □

**Theorem 3.2.6.** *If  $F$  is the limit law of  $S_n = X_{n1} + \cdots + X_{nr_n}$  for some (row-wise independent) double array  $\{X_{nk} : n \geq 1, 1 \leq k \leq r_n\}$  satisfying condition  $\mathcal{R}$ , then  $F$  has the ch.f with canonical representation*

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

**Remark 3.2.7.** If  $F$  is infinitely divisible, then  $F$  is the limit law of  $S_n$ , so  $F$  has a canonical representation!

*Proof.* First we need following fact.

**Fact.** *If  $\{\mu_n\}$  is a sequence of finite measures with  $\sup_n \mu_n(\mathbb{R}) < \infty$ , then  $\exists \{n'\} \subseteq \{n\}$  and finite measure  $\mu$  s.t. for every continuous function  $h : \mathbb{R} \rightarrow \mathbb{R}$  with  $\lim_{|x| \rightarrow \infty} h(x) = 0$ ,*

$$\int h d\mu_{n'} \xrightarrow[n' \rightarrow \infty]{} \int h d\mu.$$

Note that

$$\left| \prod_{k=1}^{r_n} \varphi_{nk}(t) - e^{\sum_{k=1}^{r_n} (\varphi_{nk}(t) - 1)} \right| \leq \sum_{k=1}^{r_n} |\varphi_{nk}(t) - e^{\varphi_{nk}(t) - 1}|$$

$$(\because |\varphi_{nk}(t)| \leq 1, |e^{z-1}| = e^{-1}|e^z| \leq e^{|z|-1} \text{ if } |z| \leq 1)$$

$$\begin{aligned}
&= \sum_{k=1}^{r_n} \left| e^{\varphi_{nk}(t)-1} - 1 - (\varphi_{nk}(t) - 1) \right| \\
&\leq \sum_{k=1}^{r_n} K |\varphi_{nk}(t) - 1|^2 \quad (\because |e^z - 1 - z| \leq K|z|^2 \text{ for some } K) \\
&\leq \sum_{k=1}^{r_n} K \left( \frac{t^2 EX_{nk}^2}{2} \right)^2 \quad \left( \because |\varphi_{nk}(t) - 1 - EX_{nk}(t)| \leq \frac{t^2 EX_{nk}^2}{2} \right) \\
&= Ct^4 \sum_{k=1}^{r_n} \sigma_{nk}^4 \\
&\leq Ct^4 \cdot \left( \max_{1 \leq k \leq r_n} \sigma_{nk}^2 \right) \cdot s_n^2 \\
&\xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

for some constant  $C > 0$ . If  $F_{nk}$  is a distribution function of  $X_{nk}$ . Then

$$\begin{aligned}
\sum_{k=1}^{r_n} (\varphi_{nk}(t) - 1) &= \sum_{k=1}^{r_n} \int (e^{itx} - 1) dF_{nk}(x) \\
&= \sum_{k=1}^{r_n} \int e^{itx} - 1 - itx dF_{nk}(x) \quad (EX_{nk} = 0) \\
&= \sum_{k=1}^{r_n} \int \frac{e^{itx} - 1 - itx}{x^2} x^2 dF_{nk}(x) \\
&= \int \frac{e^{itx} - 1 - itx}{x^2} \left( \sum_{k=1}^{r_n} x^2 dF_{nk}(x) \right) \\
&= \int \frac{e^{itx} - 1 - itx}{x^2} d\mu_n(x)
\end{aligned}$$

holds for

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

For such  $\mu_n$ , we get

$$\mu_n(\mathbb{R}) = \sum_{k=1}^{r_n} \int y^2 dF_{nk}(y) = \sum_{k=1}^{r_n} \sigma_{nk}^2 = s_n^2,$$

and hence

$$\sup_n \mu_n(\mathbb{R}) < \infty.$$

Thus by the fact mentioned above,  $\exists \{n'\} \subseteq \{n\}$  and finite measure  $\mu$  s.t.

$$\int h d\mu_{n'} \xrightarrow{n' \rightarrow \infty} \int h d\mu$$

for any continuous function  $h$  decays to 0. Applying this to the function

$$h(x) = \frac{e^{itx} - 1 - itx}{x^2},$$

we get

$$\int \frac{e^{itx} - 1 - itx}{x^2} d\mu_{n'}(x) \xrightarrow{n' \rightarrow \infty} \int \frac{e^{itx} - 1 - itx}{x^2} d\mu$$

for some finite measure  $\mu$ . In other words, letting

$$\varphi_n(t) = \exp \left( \int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} d\mu_n(x) \right) = \exp \left( \sum_{k=1}^{r_n} (\varphi_{nk}(t) - 1) \right),$$

we obtain

$$\varphi_{n'} \xrightarrow{n' \rightarrow \infty} \exp \left( \int \frac{e^{itx} - 1 - itx}{x^2} d\mu(x) \right).$$

Meanwhile, we obtained

$$\left| e^{\sum_{k=1}^{r_n} (\varphi_{nk}(t) - 1)} - \prod_{k=1}^{r_n} \varphi_{nk}(t) \right| = \left| \varphi_n(t) - \prod_{k=1}^{r_n} \varphi_{nk}(t) \right| \xrightarrow{n \rightarrow \infty} 0,$$

where

$$\prod_{k=1}^{r_n} \varphi_{nk}(t)$$

is a ch.f of  $S_n$ . Since  $S_n \xrightarrow[n \rightarrow \infty]{d} F$ ,  $\varphi_n(t) \rightarrow \varphi_F(t)$ , i.e.,

$$\varphi_{n'} \xrightarrow{n' \rightarrow \infty} \varphi_F(t).$$

Therefore, we get

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

□

**Exercise 3.2.8.** One can also show that such canonical representation is *unique*, if one uses the inversion formula (or distribution determinant property) of characteristic function. It is left to readers.

**Remark 3.2.9.** Actually, more general canonical representation is possible. Following statement is known as *Lévy-Khintchine Theorem*: *A distribution  $F$  is infinitely divisible if and only if its*



*ch.f*  $\varphi(t)$  has the form of

$$\log \varphi(t) = ict - \frac{\sigma^2 t^2}{2} + \int \left( e^{itx} - 1 - \frac{itx}{1+x^2} \right) d\mu(x),$$

where  $\mu$  is a measure with

$$\mu\{0\} = 0 \text{ and } \int \frac{x^2}{1+x^2} d\mu(x) < \infty.$$

It gives a representation of infinite divisible distributions “with nonzero means.”