Probability Theory II (Fall 2016)

J.P.Kim

Dept. of Statistics

Finally modified at September 27, 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`

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*

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

Proof. Note that, by Taylor's theorem, there exists ξ 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

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

Similarly, there exists ξ' between 0 and x such that

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

so

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

holds. Thus, we get

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

and put tX into x then we get

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

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

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

□

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,

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

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 $X_1, \cdots, X_n$ be i.i.d. random variables such that $EX_1 = 0$ and $EX_1^2 = \sigma^2 > 0$. Then, for $S_n = X_1 + X_2 + \cdots + X_n$,*

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

Proof. Let $\varphi(t) = Ee^{itX_1}$ be a characteristic function of X_1. Then characteristic function of $\dfrac{S_n}{\sigma\sqrt{n}}$ 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 $\dfrac{t}{\sigma\sqrt{n}} \approx 0$. Also note that, for a sequence c_n such that $nc_n \xrightarrow[n\to\infty]{} c$,

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

holds. Therefore,

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

and by Lévy's continuity theorem, we get the conclusion. □

1.2 Double arrays

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

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

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

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

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

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,*

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

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:

<u>**Claim.**</u> $\varphi_n(t) \to e^{-t^2/2}$

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,

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

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

$$\begin{aligned}
A_n &\le \sum_{k=1}^{r_n} E\min\left(|tX_{nk}|^3, |tX_{nk}|^2\right)\\
&= \sum_{k=1}^{r_n}\int \min\left(|tX_{nk}|^3, |tX_{nk}|^2\right) d\mathbb{P}\\
&\underset{(*)}{\le} \sum_{k=1}^{r_n}\int_{|X_{nk}|<\epsilon} |tX_{nk}|^3 d\mathbb{P} + \sum_{k=1}^{r_n}\int_{|X_{nk}|\ge\epsilon} |tX_{nk}|^2 d\mathbb{P}\\
&\le \sum_{k=1}^{r_n}\int |t|^3\epsilon|X_{nk}|^2 d\mathbb{P} + \sum_{k=1}^{r_n}\int_{|X_{nk}|\ge\epsilon s_n} |tX_{nk}|^2 d\mathbb{P}\\
&= \underbrace{\sum_{k=1}^{r_n} |t|^3\epsilon\sigma_{nk}^2}_{=|t|^3\epsilon} + \underbrace{\sum_{k=1}^{r_n}\int_{|X_{nk}|\ge\epsilon s_n} |tX_{nk}|^2 d\mathbb{P}}_{\xrightarrow[n\to\infty]{} 0 \text{ (Lindeberg)}}
\end{aligned}$$

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

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

so

$$\max_{1\le k\le r_n} \sigma_{nk}^2 \le \epsilon^2 + \underbrace{\sum_{k=1}^{r_n}\int_{|X_{nk}|\ge\epsilon} X_{nk}^2 d\mathbb{P}}_{\xrightarrow[n\to\infty]{} 0}$$

holds. It implies that,

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

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

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

$$= K \cdot \frac{t^4}{4} \max_{1 \le k' \le r_n} \sigma_{nk'}^2 \xrightarrow[n\to\infty]{} 0$$

holds, and it implies the conclusion. □

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

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

For this, see figure 1.1.

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

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

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

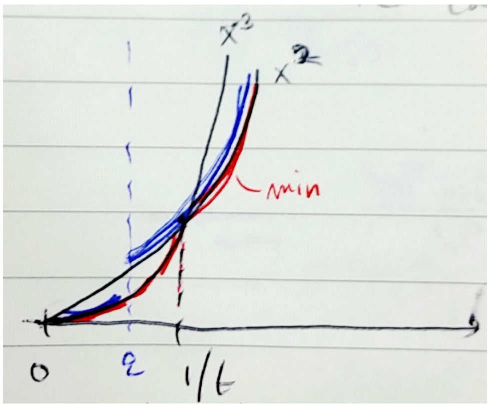

Figure 1.1: The graph of $\min(|tx|^3, |tx|^2)$.

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

(a) $EX_{nk} = 0$

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

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

Proposition 1.2.6. *Lyapunov condition implies Lindeberg condition.*

Proof.

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

□

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 $n=2,3,\cdots$. 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 $Z\sim N(0,1)$,

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

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* + $\left[\frac{\max_{1\leq k\leq r_n}\sigma_{nk}^2}{s_n^2}\xrightarrow[n\to\infty]{}0\right]$.

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

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

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,

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

holds, so

$$\max_{1\le k\le r_n} |\varphi_{nk}(t) - 1| \le \max_{1\le k\le r_n} t^2\sigma_{nk}^2 \xrightarrow[n\to\infty]{} 0$$

is obtained. Meanwhile, note that

$$|e^z - 1 - z| \le K|z|^2 \ \forall z \ s.t. \ |z| \le 2$$

holds for some K. Hence, we get

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

Now since $|e^z| = e^{\mathcal{R}ez} \le e^{|z|}$,

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

holds, so by lemma,

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

is obtained. Thus by CLT, we get

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

which implies

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

Note that

$$|e^z| = \left|e^{\mathcal{R}e(z)+i\mathcal{I}m(z)}\right| = e^{\mathcal{R}e(z)}$$

holds, so it implies that

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

and hence

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

holds. Thus,

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

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

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

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

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

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

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

□

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

(ii) $\max_{1\le m\le n} p_{nm} \xrightarrow[n\to\infty]{} 0$

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

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

$$\left|e^{p_{nm}(e^{it}-1)}\right| = e^{\mathcal{R}e(p_{nm}(e^{it}-1)} = e^{p_{nm}(\cos t-1)} \le 1$$

and

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

so we get

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

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

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

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

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

and Lévy continuity theorem ends the proof. □

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

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

Assume that

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

(ii) $\max\limits_{1\le m\le n} p_{nm} \xrightarrow[n\to\infty]{} 0$

(iii) $\sum_{m=1}^n \epsilon_{nm} \xrightarrow[n\to\infty]{} 0$

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

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,

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

holds. Now, note that

$$\begin{aligned} P(S_n \ne S'_n) &\le P\left(\bigcup_{m=1}^n (X_{nm} \ne X'_{nm})\right) \\ &\le \sum_{m=1}^n P(X_{nm} \ne X'_{nm}) \\ &= \sum_{m=1}^n P(X'_{nm} \ge 2) \\ &= \sum_{m=1}^n \epsilon_{nm} \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

With this, we get

$$P(\underbrace{|S_n - S'_n|}_{\text{integer}} \ge \epsilon) \le P(S_n \ne S'_n) \xrightarrow[n\to\infty]{} 0$$

so $S_n - S'_n \xrightarrow[n\to\infty]{P} 0$. 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 $\langle\cdot,\cdot\rangle : E \times E \to \mathbb{C}$ is a function satisfies followings.*

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

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

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

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

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

Proposition 2.1.3. *Followings hold.*

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

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

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

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

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

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

$$\|x + y\|^2 = \|x\|^2 + 2\mathcal{R}e\langle x, y\rangle + \|y\|^2,$$

we obtain

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

and hence

$$2\mathcal{R}e\bar{\lambda}\langle y - w, z\rangle \leq |\lambda|^2\|z\|^2$$

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

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

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 $v = v'$. □

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 $\|X_n - X_{n+1}\| \le 2^{-n}$ for any $n = 1, 2, \cdots$, then $\exists X \in \mathcal{L}^2$ such that*

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

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

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

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

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

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

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

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

$\square$ (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

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

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

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

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

$$\|X_n - X\| \le \underbrace{\|X_n - X_{k_n}\|}_{\to 0 \text{ (Cauchy)}} + \|X_{k_n} - X\| \xrightarrow[n\to\infty]{} 0.$$

$\square$

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

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

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

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

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

Note that $\|Y - h_{k_n}(X)\| \xrightarrow[n\to\infty]{} 0$. ("converge" means that $\|Y - h_n(X)\| \xrightarrow[n\to\infty]{} 0$.) Letting

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

and

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

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

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

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

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*

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

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 $h(X) \equiv 1$. □

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 μ, ν be σ-finite measures with $\nu << \mu$. (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*

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

and denote it as $f = \dfrac{d\nu}{d\mu}$. 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-σ-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 $A \in \mathcal{F}$, $\int_A XdP = \int_A YdP$.

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

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

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

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

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

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

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.. □

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

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

and

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

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

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 $\varphi : \mathbb{R} \to \mathbb{R}$, *if* $E|\varphi(X)| < \infty$, *then* $E[\varphi(X)|X] = \varphi(X)$.

(d) If $\mathcal{F} = \{\phi, \Omega\}$, *then* $E(X|\mathcal{F}) = EX$. *("trivial σ-field")*

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

$$E\left[E[X|\mathcal{F}_1]|\,\mathcal{F}_2\right] = E\left[E[X|\mathcal{F}_2]|\,\mathcal{F}_1\right] = E[X|\mathcal{F}_1].$$

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

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

(i) *(Markov)* $P(|X| \ge c|\mathcal{F}) \le c^{-1}E[|X||\mathcal{F}]$ *for* $c > 0$.

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

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

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

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

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

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.,

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

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 $X \ge 0$, there is a sequence of simple r.v.'s such that $X_n \nearrow X$, so $|X_nY| \le |XY|$ holds.

Thus by DCT ((k)),

$$E[X_nY|\mathcal{F}] \xrightarrow[n\to\infty]{} E[XY|\mathcal{F}],$$

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

$$\int_A E[XY|\mathcal{F}]dP = \int_A XYdP \overset{DCT}{=} \lim_{n\to\infty}\int_A X_nYdP = \lim_{n\to\infty}\int_A E[X_nY|\mathcal{F}]dP \overset{DCT}{=} \int_A \lim_{n\to\infty} X_nE[Y|\mathcal{F}]dP$$

and hence

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

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

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

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

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

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

(i) Clear.

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

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

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

(k) Let

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

Then Y_n is monotone, $Y_n \xrightarrow[n\to\infty]{a.s} 0$, and $Y_n \le 2Y$. Then $EY_n \xrightarrow[n\to\infty]{} 0$ 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,

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

and hence

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

(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

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

Proposition 2.2.12. *Let X, Y be integrable independent random variables with $E|\varphi(X, Y)|\infty$, where $\varphi : \mathbb{R}^2 \to \mathbb{R}$ 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,

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

□

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

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

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

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

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

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,

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

Thus we can say that

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

2.3 Martingales and Stopping Times

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

Definition 2.3.1. *Let $\{\mathcal{F}_n\}$ be a sequence of sub σ-fileds 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 $E[X_n \mathcal{F}_{n-1}] = X_{n-1}$ $\forall n \ge 1$.*

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

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

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

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

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

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

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

where $h_n : \mathbb{R}^n \to \mathbb{R}$ is Borel. Assume that X_n's are integrable. Then letting $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$, 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 $\varphi : \mathbb{R} \to \mathbb{R}$ 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 submartinagle and $\varphi : \mathbb{R} \to \mathbb{R}$ 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 supermartinagle and $\varphi : \mathbb{R} \to \mathbb{R}$ 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 $\varphi(x) = |x|,\ x^+, |x|^p\ (p \geq 1), |x-a|,\ (x-a)^+, \cdots$.

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

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

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

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

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 (& ∞). 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.

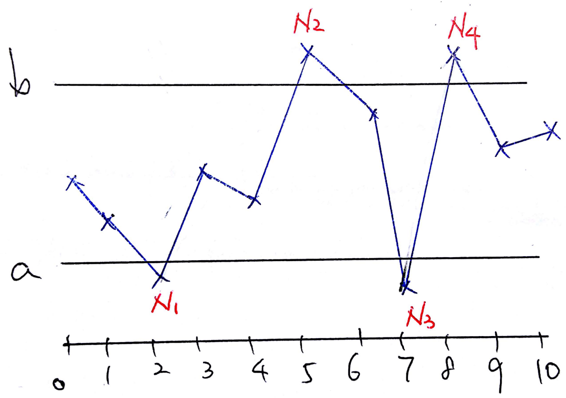

Figure 2.1: X_n and N_n's. For example, $N_4 = 8$.

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

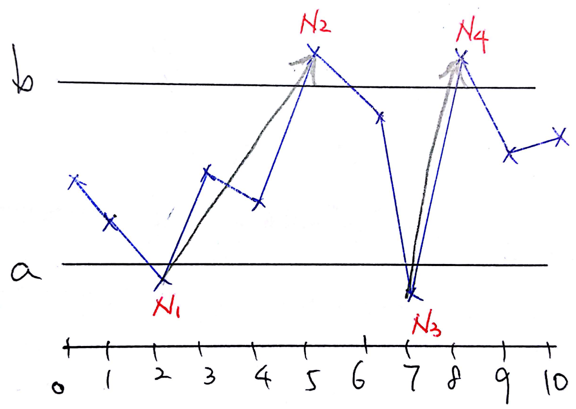

Figure 2.2: Upcrossing process. For example, in this figure, $U_{10} = 2$.

consider Y_n instead of X_n without loss of generality.

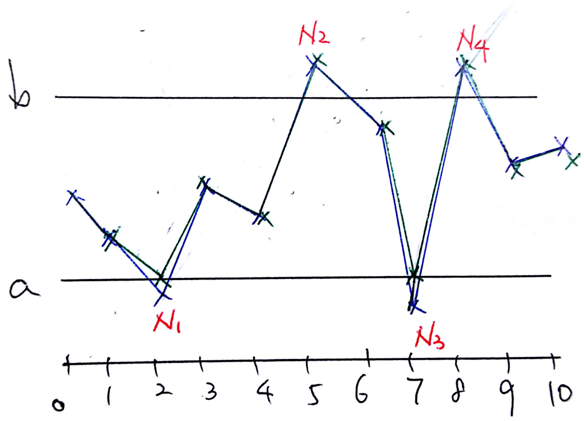

Figure 2.3: Upcrossing process and Y_n.

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 $J_k = \{N_{2k-1} + 1, \cdots, N_{2k}\} = \{m : N_{2k-1} < m \leq N_{2k}\}$ 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

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

(Note that $N_{2U_n} \le n$) 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 $(N_{2k-1} < m \le N_{2k}) = (N_{2k-1} \le m-1) \cap (N_{2k} \le m-1)^c \in \mathcal{F}_{m-1}$.) 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 $E(\tilde{H} \cdot Y)_n \ge E(\tilde{H} \cdot Y)_1 \ge 0$ and hence

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

i.e.,

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

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

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

□

Remark 2.3.15. The key fact is that $E(\tilde{H} \cdot Y)_n \ge 0$, 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 $n \to \infty$, X_n converges a.s. to a limit X with $E|X| < \infty$.*

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

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

Note that U_n is monotone, so $\exists U$ s.t. $U_n \nearrow U$. Then from

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

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

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

(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

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

i.e., $\liminf X_n = \limsup X_n$ P-a.s., which implies $\exists \lim X_n =: X$ P-a.s.. Now by Fatou's lemma,

$$EX^+ \le \liminf_{n\to\infty} EX_n^+ < \infty$$

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

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

holds, and by Fatou again, we get

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

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 $X_n \ge 0$ is a $((\mathcal{F}_n)$-$)$supermartingale, then as $n \to \infty$, $X_n \to X$ a.s. and $EX \le EX_0$.*

Proof. $Y_n = -X_n \le 0$ is a submartingale with $EY_n^+ = 0$. Thus by previous theorem, Y_n has a

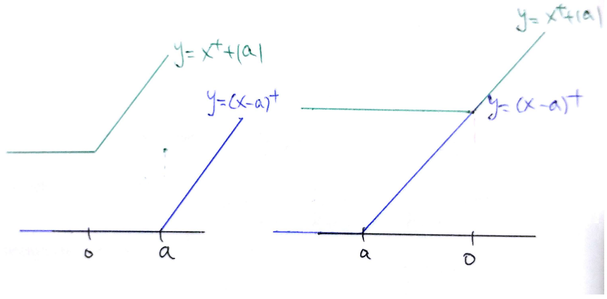

Figure 2.4: $y = (x-a)^+$ and $y = x^+ + |a|$.

limit Y, and $X_n \xrightarrow[n\to\infty]{a.s} -Y =: X$. As X_n is a supermartingale, we get $EX_0 \geq EX_n$, and with Fatou's lemma, we obtain $EX \leq EX_0$.