Notes on Probability Theory

October 3, 2024

Contents

Part I

Classical Probability

0.1 Theorems on Measures

Given two probability measures, we would like to know under what condition they agree on a σ algebra (field). Two measures agreeing on the generator of the measure is not a sufficient condition for those two measures agreeing on the entire σ algebra. So natural question to ask is that are there some sub-family of the entire σ field such that if two probability measures agree on the family, then they agree on the σ field? There are, here are two of them:

Definition 0.1.1 (Monotone Class). *Let $\mathcal{F}$ be a σ algebra, let $\mathcal{G} \subset \mathcal{F}$, we say $\mathcal{G}$ is a monotone class if it is closed under countably infinite intersections of decreasing sets and closed under countably many unions of increasing sets. That is,*

1. $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{G}$ *with* $\mathcal{A}_i \subset A_{i+1}$, *then* $\bigcup_{i\geq 0} A_i \in \mathcal{G}$.
2. $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{G}$ *with* $\mathcal{A}_i \supset A_{i+1}$, *then* $\bigcap_{i\geq 0} A_i \in \mathcal{G}$.

We list this useful theorem w/o proof, which can be found in ([LG16],[Bil08],[Dur19])

Theorem 0.1.1. *Let $\mathcal{F}$ be a collection of sets that is closed under finite intersection, then the smallest monotone class contains $\mathcal{F}$ is equal to the σ algebra generated by $\mathcal{F}$.*

Here is why the theorem is useful:

Theorem 0.1.2. *Let $(E, \mathcal{F})$ be a measurable space, and let $\mathcal{A} \subset \mathcal{F}$ where $\sigma(\mathcal{A}) = \mathcal{F}$ and $\mathcal{A}$ closed under finite intersections and unions. Then if μ and ν, two σ finite measures, agrees on $\mathcal{A}$, then they agree on $\mathcal{F}$.*

The proof of this theorem is called a monotone class argument:

Proof. By continuity of measures (from below and above), we see

$$\mathcal{G} \triangleq \{A \in \mathcal{F} : \mu(A) = \nu(A)\} \tag{1}$$

is a monotone class that contains $\mathcal{A}$, so $\sigma(\mathcal{A}) \subset \mathcal{G}$ by monotone class theorem. □

The second sub-families of sets that gives us similar results are

Definition 0.1.2 (λ-system). *Let X be a set, and $L \subset \mathcal{P}(X)$ where $\mathcal{P}$ denotes the power sets (collection of all sets), then we say L is a λ-sysmte if*

1. $\emptyset \in L$;
2. *L is closed under complement;*
3. *l is closed under countable disjoint unions.*

Definition 0.1.3 (π-system). *Let X be a set, and $P \subset \mathcal{P}(X)$ where $\mathcal{P}$ denotes the power sets (collection of all sets), then we say P is a π-sysmte if it is closed under finite intersections.*

The following theorem is called the $\pi - \lambda$ theorem, also stated w/o proof:

Theorem 0.1.3 ($\pi - \lambda$ Theorem). *Let P be a π system and L be a λ system. Suppose $P \subset L$, then $\sigma(P) \subset L$.*

Here is why it is useful:

Theorem 0.1.4. *Let μ, ν be two probability measures on the measurable space $(\Omega, \mathcal{F})$, and say μ and ν agree on a π system, call it P, that generates $\mathcal{F}$, then ν, μ agree on the whole σ field.*

Proof. Let ν, μ be probability measure and denote the following set

$$\mathcal{A} \triangleq \{E \in \mathcal{F} : \mu(E) = \nu(E)\}$$

then we note $\mathcal{A}$ is a λ system since $\emptyset \in \mathcal{A}$, and if $\mu(A) = \nu(A)$, then $\mu(A^c) = 1 - \mu(A) = 1 - \nu(A) = \nu(A^c)$, and it is also closed under countable disjoint unions, so $\mathcal{A} \supset \sigma(P) = \mathcal{F}$. □

0.2 Inequalities

Proposition 0.2.1 (Fatou's Lemma)**.** *Let $\{X_n\}$ be random variables with $X_n \geq 0$ a.s. for all n, then*

$$\int_\Omega \underline{\lim\inf}_{n\to\infty} X_n d\mathbb{P} \leq \underline{\lim\inf}_{n\to\infty} \int_\Omega X_n d\mathbb{P}$$

Definition 0.2.1 (Convex function)**.** *Let $\varphi : \mathcal{B} \to \mathbb{R}$, where $\mathcal{B}$ is a topological vectors space on $\mathbb{R}$. we say φ is convex if*

$$\varphi(rx + (1-r)y) \leq r\varphi(x) + (1-r)\varphi(y), \quad \forall x, y \in \mathcal{B}, r \in [0,1].$$

Proposition 0.2.2 (Jessen's Inequality)**.** *Let φ be a positive convex function and let $|X|$ be an integrable random variable, then we have*

$$\varphi(\mathbb{E}[|X|]) \leq \mathbb{E}(\varphi(|X|))$$

Remark 0.2.1. *We can often use Jessen's Inequality to determine integrability of random variables. Jessen's Inequality is true in more general space, such as locally convex topological vector space, or any vector space where we can use Hanh-Banach.*

Proof. We prove this in locally compact topological vectors space that has a probability measure $\mathbb{P}$.

Let $C = \{(x,y) \in \mathcal{B} \times \mathbb{R} : y > \varphi(x)\}$, so C consists of all the points that is "strictly above" the graph of φ. Note that C is a convex set: $(x_i, y_i) \in C$ where $i = 1,2$, then $\gamma y_1 + (1-\gamma)y_2 > \gamma\varphi(x_1) + (1-\gamma)\varphi(x_2)$. Now let $A = \{(s, \varphi(s))\}$ which is obviously convex. By Hanh-Banach separation theorem, there is a hyperplane in $\mathcal{B} \times \mathbb{R}$ that separates C, A, that is, there are $a \in \mathcal{B}'$ and $c \in \mathbb{R}$ such that

$$a(s) + c\varphi(s) = b; \quad a(x) + by \geq b, \forall (x,y) \in \mathcal{B}.$$

That is, there is an affine function $f_s(x) = a(x) + b$ such that $f_s(s) = \varphi(s)$ and $\varphi(x) \geq f_s(x)$ (where we used continuity of $a(\cdot)$ to extend the Inequality given by Hanh-Banach to the boudnary of C).

Now let $s = \mathbb{E}[|X|]$, then one has the following:

$$\mathbb{E}[\varphi(|X|)] \geq \mathbb{E}[f_s(|X|)] = f_s(\mathbb{E}[|X|]) = \varphi\left(\mathbb{E}[X|X|]\right).$$

□

Proposition 0.2.3 (Chebyshev-Markov)**.** *Let $\varphi : \mathbb{R} \to \mathbb{R}^+$ be a strictly increasing on $(0,\infty)$, even and convex function and let X be a random variable and suppose $\mathbb{E}[\varphi(X)]$ finite then*

$$\mathbb{P}[|X| \geq \lambda] \leq \frac{\mathbb{E}[\varphi(X)]}{\varphi(\lambda)}$$

Proof.

$$\mathbb{E}[\varphi(X)] = \int_\Omega \varphi(X) d\mathbb{P} \geq \int_{|X|\geq\lambda} \varphi(X) d\mathbb{P} \geq \varphi(\lambda)\mathbb{P}[|X| \geq \lambda].$$

□

There is a special case of Chebyshev-Markov that will be useful later:

Proposition 0.2.4. *Let X be a random variable such that $var(X) < \infty$ and has mean μ, then*

$$\mathbb{P}[|X - \mu| \geq \epsilon] \leq \frac{var(X)}{\epsilon}.$$

Proof. □

We note that Chebyshev Markov can be made more general in L^p for $0 < p < \infty$ in the sense that φ does not have to be convex

Proposition 0.2.5. *Let X be a random variable such that $\mathbb{E}[|X|^r] < \infty$ for some positive r, then*

$$\mathbb{P}[|X| \geq \lambda] \leq \frac{\mathbb{E}[|X|^r]}{\lambda^r}.$$

Proof.

$$\mathbb{P}[|X| \geq \lambda] = \int_{|X|\geq\lambda} 1 d\mathbb{P} = \int_{|X|\geq\lambda} \frac{\lambda^r}{\lambda^r} d\mathbb{P} \leq \int_{|X|\geq\lambda} \frac{|X|^r}{\lambda^r} d\mathbb{P} \leq \frac{\mathbb{E}[|X|^r]}{\lambda^r}$$

□

Remark 0.2.2. *In application, usually use $\varphi(\cdot) = |\cdot|^m$ for $m \in \mathbb{N}$ so we can bound the tail probability by the moments.*

Remark 0.2.3. *On independence: There are two notion of independence of a sequence of random variables $\{X_n\}$*

- *(complete) independence: $\mathbb{P}\left(\bigcap_{n\in\mathcal{I}}\{X_n \in A_n\}\right) = \prod_{n\in\mathcal{I}} \mathbb{P}\left(X_n \in A_n\right)$ for any index set $\mathcal{I}$.*
- *Pairwise independence: $\mathbb{P}\left(\{X_n \in A_n\} \cap \{X_m \in A_m\}\right) = \mathbb{P}\left(\{X_n \in A_n\}\right)\mathbb{P}\left(\{X_m \in A_m\}\right)$.*

We note that independence implies Pairwise independence, but the converse is not true by the following example.

Example 0.2.1. *Let X_i $i = 1, 2, 3$ independent random variable taking values in $\{0, 1\}$ with probability half and half, and let $A_1 = \{X_1 = X_2\}, A_2 = \{X_1 = X_3\}, A_3 = \{X_2 = X_3\}$. Then clearly A_i's are pairwise independent but not complete independent, which can be seen simply by σ-algebra they generates.*

0.3 Modes of Convergence

Definition 0.3.1. *Convergence a.e or almost surely; Convergence in probability: $\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0, \forall \epsilon > 0$; convergence in $L^p, p \geq 1$: $\lim_{n\to\infty} \mathbb{E}|X_n - X|^p = 0$, weak convergence/convergence in distributions: $\int_\Omega f(x) d\mu_n(x) \to \int_\Omega f(x) d\mu(x)$ for all bounded continous (C_c, C_c^∞ all works) where $\mu_n = \mathbb{P} \circ X_n^{-1}$ the pushforward measure, same for μ, the pushforward measure for X.*

Here is a useful characterization of a.s. convergence

Proposition 0.3.1. *Let $(E, |\cdot|)$ be a complete metric space the random variables are taking values on E. $\{X_n\}$ be sequence of random variables converges to X a.s. if and only if*

$$\lim_{m\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon, \forall n \geq m\right) = 1, \quad \forall \epsilon > 0, \tag{2}$$

or equivalently,

$$\lim_{m\to\infty} \mathbb{P}\left(|X_n - X| \leq \epsilon \text{ for some } n \geq m\ \right) = 0, \quad \forall \epsilon > 0.$$

Proof. I will use the first limit. ($\Longrightarrow$) Suppose $X_n \to X$ a.e., then $\exists \Omega_0 \subset \Omega$ with $\mathbb{P}[\Omega_0] = 1$ and for all $\omega \in \Omega_0$, $\lim_{n\to\infty} |X_n(\omega) - x(\omega)| = 0$. Translating this into set theory language is

$$\mathbb{P}\left(\bigcup_m \bigcap_{n\geq m} \{|X_n - X| > \epsilon\}\right) = 1 \tag{3}$$

where $\bigcup_m \bigcap_{n\geq m} A_n$ is called $\liminf A_n$, and this reads " there exists m such that for all $n \geq m$, A_n happens". We observe that $\cap_{n\geq m}\{|X_n - X| > \epsilon\}$ is a family of increasing set, so by continuity of measure, this direction is proven.

($\Longleftarrow$) Now suppose (2) holds, then by continuity of the measure, (3) also hold. □

Remark 0.3.1. *Obviously, convergence in probability (measure) is weaker than convergence a.e.*

Proposition 0.3.2. *Convergence in L^p for $p \geq 1$ implies convergence in probability.*

Proof. Suppose $\lim_{n\to\infty} \mathbb{E}\left(|X_n - X|^p\right) = 0$, then

$$\mathbb{P}\left(|X_n - X| \geq \epsilon\right) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p} \to 0, \quad \forall \epsilon > 0.$$

□

Remark 0.3.2. *Convergence a.e. can also be formulated in terms of Cauchy sequence: $X_n \to X$ a.s. if and only if*

$$\lim_{k\to\infty} \mathbb{P}[|X_n - X_m| \geq \epsilon; \textit{ for all } n, m \geq k] = 0$$

We note that the converse of (7) can be true when X_n's are "nice" in sense of sequence of functions:

Theorem 0.3.1. *Let $\{X_n\}$ be a sequence that is dominated by some integrable random variable Y and suppose $X_n \to X$ in probability (measure), the $X_n \to X$ in L^p when $X \in L^p$.*

Proof.

$$\begin{aligned}
\mathbb{E}[|X_n - X|^p] &= \int_\Omega |X_n - X|^p d\mathbb{P} \\
&= \int_{|X_n - X|\geq\epsilon} |X_n - X|^p d\mathbb{P} + \int_{|X_n - X|<\epsilon} |X_n - X|^p d\mathbb{P} \\
&\leq \left(\mathbb{P}[|X_n - X| \geq \epsilon]\right)^{\frac{1}{p}} \|X + Y\|_p + \epsilon^p \to \epsilon^p
\end{aligned}$$

where in the last Inequality we used Holder, and this holds for arbitrary small ϵ. □

0.4 Borel-Cantelli

Definition 0.4.1. *Here we define the limit sup/inf in set theoritical setting:*

$$\limsup A_n = \bigcap_{m\geq 0} \bigcup_{n\geq m} A_n; \quad \liminf A_n = \bigcup_{m\geq 0} \bigcap_{n\geq m} A_n.$$

It is convenient to understand $\bigcup$ as "exists" and $\bigcap$ as "for all" in elementary analysis settings, and under stand $\{A_n\}$ as a sequence of "events". Also, $\limsup$ can be understood as A_n happens infinitely often, denote as I.O.

Proposition 0.4.1 (Borel-Cantelli Lemma 1)**.** *For any sequence of measurable set $\{E_n\}$ in a probability space, we have*

$$\sum_{n\in\mathbb{N}} \mathbb{P}[E_n] < \infty \Rightarrow \mathbb{P}[E_n, I.O] = \mathbb{P}[\limsup E_n] = \mathbb{P}\left(\bigcap_{m\geq 0}\bigcup_{n\geq m} E_n\right) = \lim_{m\to\infty]} \mathbb{P}\left(\bigcup_{n\geq m} E_m\right) = 0$$

Proof. Note

$$\sum_{n\geq m} \mathbb{P}[E_n] \geq \mathbb{P}\left(\bigcup_{n\geq m} E_n\right) \to 0$$

due to convergence of the sum. □

There is a converse to the first B-C lemma, with some restrictions on $\{E_n\}$ of course, otherwise they would be made into one lemma:

Proposition 0.4.2 (Borel-Cantelli 2)**.** *Suppose $\{E_m\}$ are independence events, then the converse of (8) is true, namely,*

$$\sum_{n=0}^{\infty} \mathbb{P}[E_n] = \infty \Rightarrow \mathbb{P}[E_n, I.O] = \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m} E_n\right) = 1.$$

Proof. Here we use the identity $(\limsup E_n)^c = \liminf E_n^c$, and we show that $\liminf E_n^c$ has measure zero.

$$\begin{aligned}
\mathbb{P}\left(\left(\bigcap_{m\geq 0}\bigcup_{n\geq m} E_n\right)^c\right) &= \mathbb{P}\left(\bigcup_{m\geq 0}\bigcap_{n\geq m} E_n^c\right) \\
&= \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{n\geq m} E_n^c\right) \\
&\leq \lim_{m\to\infty}\lim_{k\to\infty} \mathbb{P}\left(\bigcap_{m\leq n\leq k} E_n^c\right) \\
&= \lim_{m\to\infty}\lim_{k\to\infty} \prod_{m\leq n\leq k} \mathbb{P}\left(E_n^c\right) \\
&= \lim_{m\to\infty}\lim_{k\to\infty} \prod_{m\leq n\leq k} \left(1 - \mathbb{P}\left(E_n\right)\right)
\end{aligned}$$

By the divergence of the sum given, we see the inner limit goes to zero (heuristically, since the convergence of infinite product is defined by the sum of log). □

If we combine the two B-C lemmas, then we obtain a zero one law:

Proposition 0.4.3. *Suppose $\{E_n\}$ is a sequence of independence events, the*

$$\mathbb{P}\left(\limsup E_n\right) \in \{0, 1\}$$

This is because a infinite sum with positive summants either converges or diverges.

Remark 0.4.1. *There are severl applications of Borel-Cantelli Lemmas, one of which is use it to determine convergence of a random variable; we can also use Borel-Cantelli to construct a almost surely convergence subsequence from a convergent sequence of random variables in probability, which I will describe in the following two lemmas.*

Lemma 0.4.1. *Let $\{X_n\}$ be a sequence of random variables, then X_n converges to X if*

$$\sum_{n\in\mathbb{N}} \mathbb{P}[|X_n - X| \geq \epsilon] < \infty \quad \forall \epsilon > 0.$$

And if $\{X_n\}$ is a collection of independent random variables, then it converges to zero if and only if

$$\sum_{n\in\mathbb{N}} \mathbb{P}[|X_n| \geq \epsilon] < \infty \quad \forall \epsilon > 0.$$

The above lemma will be made more general in the zero-one law that will appear later.

Lemma 0.4.2. *For all X_n that converges in probability, there is a subsequence converges a.e. to the same random variable.*

Proof. Here we will construct such subsequence from $\{X_n\}$. By assumption, there is X such that

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0, \quad \forall \epsilon > 0.$$

Then for all $k \in \mathbb{N}$ there is X_{n_k} such that $\mathbb{P}[|X_{n_k} - X| > \epsilon] \leq \frac{1}{k^2}$. Note that this subsequence $\{X_{n_k}\}$ satisfies the condition for (1), so it converges a.e.

Another such construction: assume WLOG that $X_n \to 0$ (since we can take the sequence to be $X_n - X$), then by assumption,

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \frac{1}{2^k}] = 0; \quad \forall k \geq 0$$

So for all k there is n_k with $\mathbb{P}[|X_{n_k} - X| > \frac{1}{2^k}] \leq \frac{1}{2^k}$. Then by Borel-Cantelli we have

$$\mathbb{P}\left(|X_{n_k}| \geq \frac{1}{2^k}, I.O.\right) = 0.$$

□

Here is a characterization of random variables converges in probability:

Proposition 0.4.4. *$\{X_n\}$ be a sequence of random variables, they converges in probability if and only if every subsequence has a further subsequence that converges a.e. to X.*

Proof. ($\Leftarrow$) is obvious by above lemma.

($\Rightarrow$): Suppose X_n does not converges X in probability, then there is a $\delta > 0$ such that $\mathbb{P}\left(|X_n - X| \geq \delta\right)$ does not converges to zero. then there is a $\epsilon > 0$ and a subsequence $\{X_{n_k}\}$ such that $\mathbb{P}\left(|X_n - X| \geq \delta\right) > \epsilon$ for some ϵ, and clearly there is no subsequence converges to 0, hence no convergence a.e. □

A direct consequence of (11) is we can pass convergence to probability to bounded continuous functions:

Proposition 0.4.5. *Let f be bounded continuous function and let $X_n \to X$ in prop, then $f(X_n) \to f(X)$ in L^p.*

Proof. By (11) for all subsequence of X_n, there is a further subsequence converges to X a.e. Which tells us for all subsequence of $f(X_n)$, there is a further subsequence converges to $f(X)$. Since f bounded, we can apply Dominated Convergence theorem to see that every subsequence of the sequence of real numbers, $\alpha_n = \mathbb{E}\left(|f(X_n) - f(X)|^p\right)$, there is a convergence further subsequence.

Now suppose α_n does not converge to zero, then there is a subsequence call β_n that is uniformly bounded away by some small number $\delta > 0$ and clearly no subsequence of this number converges to 0, hence a contradiction. □

This theorem implies $f(X_n) \to f(X)$ in probability.

Given $X_n \to X$ in prob, we can create many random variables that converges in probability from this by (5).

Proposition 0.4.6. *Let $X_n \to X$ in probability, then for any bounded continuous function f, $f(X_n) \to f(X)$ in probability as well.*

Proof. Only thing we need to prove is that for all subsequence of $f(X_n)$ there is a further subsequence that converges a.e. but this is obvious since f is continuous. □

Definition 0.4.2. *Converges in distributions: Let $\{X_n\}$ be a sequence of random variables taking values in a complete seperable metric space, then we define converges in distribution by $\int f(x)\mu_n(dx)$ for all $f \in C_b$. Here we denote convergence in distribution by $X_n \Rightarrow X$ or $\mu_n \Rightarrow \mu$.*

Convergence in distribution is weaker than convergence in probability:

Proposition 0.4.7. *E, ρ be complete seperable metric space and let $\{X_n\}$ be a sequence of random variables taking values in E such that $X_n \to X$ in probability, then $X_n \Rightarrow X$.*

Proof. By (12) we see that $f(X_n) \to f(X)$ in L^p for $p \geq 1$, so we have

$$\int f(X_n) d\mathbb{P} \to \int f(X) d\mathbb{P}$$

for all bounded continuous function f. □

Finally, we know in measure theory, $f_n \to f, g_n \to f$ a.e. implies $f_n g_n \to fg, f_n + g_n \to f + g$ a.e., here are similar properties of convergence in prob and distributions

Proposition 0.4.8. *If $X_n \to X$ and $Y_n \to Y$ in probability, then the following holds*

$$X_n Y_n \to XY \quad \textit{in probability}$$
$$X_n + Y_n \to X + Y \quad \textit{in probability}$$

Proof. Note $X_n Y_n - XY = X_n Y_n - XY_n + XY_n - XY = Y_n (X_n - X) + X(Y_n - Y)$. It reduces to show the first assertion for $X_n \to X$ and $Y_n \to 0$, and the second assertion.

For second assertion

$$\mathbb{P}\left(|X_n + Y_n - X - Y| > \epsilon\right) \leq \mathbb{P}(|X_n - X| \geq \frac{\epsilon}{2}) + \mathbb{P}(|Y_n - Y| \geq \frac{\epsilon}{2}) \to 0.$$

Now for the frist case we note that for any subsequence n_k of the natural number, there is a further subsequence call $\alpha_k = \psi(n_k)$ such that $Y_{\alpha_k} \to Y$ a.e. Since $\psi(n_k)$ itself is a subsequence of the natrual number, there is a further subsequence call $\beta_k = \phi(\psi(n_k))$ such that $X_{\beta_k} \to 0$ a.e. So $X_{\beta_k} Y_{\beta_k} \to 0$ a.e. which is a further subsequence of n_k. □

The following similar property for convergence in distributions is a direct consequence of the previous proposition

Proposition 0.4.9. *If $X_n \to X$ and $Y_n \Rightarrow Y$, then the following holds*

$$X_n Y_n \Rightarrow XY$$
$$X_n + Y_n \Rightarrow X + Y$$

0.5 Uniformly Integrability

Uniformly integrability provides conditions for reversing the implications:

$$L^p\text{convergence} \Rightarrow \text{convergence in probability} \Rightarrow \text{convergence in distributions} \tag{4}$$

Definition 0.5.1 (Uniformly integrability). *Suppose $\{X_t\}_{t\in T}$ is a family of random variables where T is an index set, we say the family is uniformly integrable if*

$$\lim_{M\to\infty} \sup_{t\in T} \mathbb{E}\left(|X_t| 1_{|X_t|\geq M}\right) = 0.$$

The idea is from uniformly bounded in L^p, but "a little" stronger, which is shown in the following characterization of uniformly integrability:

Proposition 0.5.1. *The family of random variable $\{X_t\}_{t\in T}$ is uniformly integrable if and only if the following two conditions holds:*

- $\sup_{t\in T} \mathbb{E}|X_t|^r < \infty$ *for some $r \geq 1$.*
- *For all $\epsilon > 0$ there is $\delta(\epsilon)$ such that*

$$\mathbb{P}[E] \leq \delta(\epsilon) \Rightarrow \sup_{t\in T} \int_E |X|^r d\mathbb{P} \leq \epsilon.$$

Remark 0.5.1. *The second condition can be understood as a continuity condition at zero on the set function $E \to \sup_t \mathbb{E}[|X_t|^r 1_E]$*

Proof. We show this for $r = 1$, the cases for $r \geq 1$ is trivial due to Jessen's Inequality.

Suppose uniformly integrability then

$$\int |X_t| d\mathbb{P} \leq \int |X_t| 1_{|X_t|\geq M} d\mathbb{P} + M$$

and by definition, there is an $M \geq 1$ such that the first term on the right hand side is bounded. The second asssertion: let $E_M = \{|X_t| \geq M\}$

$$\sup_t \int_E |X_t| d\mathbb{P} = \sup_t \left(\int_{E\setminus E_M} + \int_{E\cap E_M}\right) |X| d\mathbb{P} \leq \sup_t \int |X| 1_{|X|\geq M} d\mathbb{P} + M\mathbb{P}[E]$$

Let M be so large that the first term on the right hand side less than $\frac{\epsilon}{2}$ and let $\mathbb{P}[E] \leq \frac{\epsilon}{2M}$.

Now for the converse, let the L^1 norm of the family be bounded by A.

$$\mathbb{P}\left(|X_t| \geq M\right) \leq \frac{\mathbb{E}|X_t|}{M} \leq \frac{A}{M}$$

By the second condition, for all $\epsilon > 0$, there is $M \geq 1$ such that $\mathbb{P}[|X_t| \geq M] \leq \delta(\epsilon)$ implies

$$\sup_t \int |X_t| 1_{|X_t|\geq M} d\mathbb{P} \leq \epsilon.$$

□

Proposition 0.5.2. *Let $X_n \to X$ in probability, then the followings are equivalent:*

1. $\{|X_t|^r\}$ *is uniformly integrable.*
2. $X_t \to X$ *in* L^r.
3. $\mathbb{E}[|X_t|^r] \to \mathbb{E}|X|^r < \infty$.

Proof. $(1) \Rightarrow (2)$ We note that

$$\underline{\lim}\mathbb{E}[|X_{n_k}|^r] \leq \mathbb{E}[|X|^r]$$

by Fatou's lemma where X_{n_k} is a subsequence that converges to X a.e. So $|X|^r$ is integrable. Note we have the Inequality:

$$|X_n - X|^r \leq C_r\left(|X_n|^r + |X|^r\right)$$

so $Y_n = |X_n - X|^r$ is also uniformly integrable.

$$\int |X_n - X|^r d\mathbb{P} \leq \int_{|X_n - X| \geq \lambda} |X_n - X|^r d\mathbb{P} + \lambda \mathbb{P}[|X_n - X| \geq \lambda]$$

Let n be so large such that $\mathbb{P}[|X_n - X| \geq \lambda] \leq \delta(\epsilon)$ where $\delta(\epsilon)$ is given by (17), then above integrals are bounded by $\epsilon + \lambda\delta(\epsilon)$, let $\epsilon \to 0$ and $\delta(\epsilon) = 0$ we have desired result.

$(2) \Rightarrow (3)$ is trivial.

$(3) \Rightarrow (1)$. Define

$$\psi_M(x) = \begin{cases} x^r & |x|^r \leq M \\ 0 & |x|^r \geq M \end{cases}$$

We observe that $\mathbb{E}[|X_n|^r 1_{|X_n|^r \geq M}] = \mathbb{E}[|X_n|^r - \psi_M(|X_n|)]$ so for large n

$$\begin{aligned} \mathbb{E}[|X_n|^r 1_{|X_n|^r \geq M}] &= \mathbb{E}[|X_n|^r - \psi_M(|X_n|)] \\ &\leq \mathbb{E}[|X|^r] - \mathbb{E}[\psi_M(X)] + \epsilon \end{aligned}$$

This is because $\psi_M(X_n)$ converges to $\psi_M(X)$ in L^r and the condition (3) given in the proposition. Now let M be so large such that the difference between the first two terms be small, so this can be made arbitrary small for large n, then use M to even larger if neccessary to ensure for small n's, the left hand side is also small. □

We shall need the above result for Martingale Theories.

Finally, there is a useful formula to calculate the L^p norm of a random variable using the tail probability:

Proposition 0.5.3. *Let X be a random variable such that $\mathbb{E}[|X|^p] < \infty$ for some p, then we have the following formula for $\mathbb{E}|X|^p$:*

$$\mathbb{E}[|X|^p] = \int_{t=0}^{\infty} pt^{p-1}\mathbb{P}[|X| > t]dt$$

Proof.

$$\begin{aligned} \int_0^\infty pt^{p-1}\mathbb{P}[|X| > t]dt &= \int_0^\infty pt^{p-1} \int_\Omega 1_{|X|(\omega) \geq t} d\mathbb{P}(\omega) dt \\ &= \int_\Omega \int_0^\infty pt^{p-1} 1_{|X|(\omega) \geq t} dt d\mathbb{P}(\omega) \\ &= \int_\Omega \int_0^{|X|(\omega)} pt^{p-1} dt d\mathbb{P}(\omega) \\ &= \int_\Omega |X|^p(\omega) d\mathbb{P}(\omega) = \mathbb{E}[|X|^p] \end{aligned}$$

where we use Fubini for change of order of integration. □

0.6 Levy's Continuity Theorem/Some Fourier analysis

It is useful to talk about the Fourier transform of a measure. In harmonic analysis, Bochner's theorem (not proven here) tells us that a measure is uniquely determined by its fourier transform, which gives a positive definite function. In this section I choose to use tools from harmonic analysis (from ([MS13])) because it gives an eaiser (not so technical) proof for Levy's continuity theorem. Here are some definitions and results we need to know:

Definition 0.6.1. *Let $f \in C_c(\mathbb{R}^d)$, we define the Fourier transform of f as*

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} \exp\left(-2\pi i \xi \cdot x\right) f(x) dx.$$

Remark 0.6.1. *We would note that this definition naturally extends to all functions such that the above integral makes sense, namely, $L^1(\mathbb{R}^d)$ functions. It also extends naturally to the set of finite borel measures.*

Definition 0.6.2. *Let $\mu \in \mathcal{M}(\mathbb{R}^d)$ be a finite borel measure, then we define the Fourier transform of μ as*

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^d} \exp\left(-2\pi i \xi x\right) \mu(dx).$$

We will make use of the Schwartz Class:

Definition 0.6.3. *Define $\mathcal{S}$, the Schwartz Class, to be a family of smooth (infinitely differentiable functions) such that*

$$\|x^\alpha \partial^\beta f(x)\|_\infty < \infty, \quad \forall \alpha, \beta \in \mathbb{Z}^d$$

and the convergence in the Schwartz class is defined by

$$f_n \to_{\mathcal{S}} f \iff \lim_{n\to\infty} \sum_{\alpha,\beta \in \mathbb{Z}^d} \left(\|x^\alpha \partial^\beta (f_n - f)\|_\infty \right) \wedge \frac{1}{2^{|\alpha+\beta|}} \to 0$$

We define $\mathcal{S}'$ to be the dual of $\mathcal{S}$(continuous linear functionals), which is also called tempered distributtions. Convergence in the space of tempered distributionns is $\omega_t, \omega \in \mathcal{S}'$ then $\omega_t \to \omega$ in tempered distribution if $\langle \omega_t, f \rangle \to \langle \omega, f \rangle$ for all $f \in \mathcal{S}$. Note that all probability measures are in the space of tempered distributions, and we shall see later in the proof of Levy's theorem that converges in tempered distribution is the same as converges in distribution for probability measures.

Here are some properties of Fourier transforms that we state without proof, and we will use them in the future without reference:

Lemma 0.6.1.

1. *let $f, g \in \mathcal{S}$ then $\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi)$.*
2. *$\widehat{\partial_x^n f}(\xi) = (2\pi i \xi)^n \hat{f}(\xi)$.*
3. *Inverse Fourier tranform: $f(x) = C \int \exp(-2\pi i \xi x) \hat{f}(\xi) d\xi$ for some universal constant C. This formula holds for a much wider class of functions, e.g. L^1.*

Now we identify an important property of Fourier transform

Proposition 0.6.1. *Let $f, g \in \mathcal{S}$, then we have the following identity:*

$$\int \hat{f} g = \int f \hat{g}.$$

Proof.

$$
\begin{aligned}
\int \hat{f} g &= \int \int \exp(-2\pi i \xi x) f(\xi) d\xi g(x) dx \\
&= \int f(\xi) \int \exp(-2\pi i \xi x) g(x) dx d\xi \\
&= \int f \hat{g}.
\end{aligned}
$$

□

Note that the identity in (20) holds whenever the change of order of integration makes sense. So, actually, we can use this to define Fourier transform for tempered distributions:

Definition 0.6.4. *Let $\mu \in \mathcal{S}'$, we define the Fourier transform of μ, call $\hat{\mu}$ to be the element in $\mathcal{S}'$ such that*

$$\langle \hat{\mu}, \phi \rangle = \langle \mu, \hat{\phi} \rangle, \quad \forall \phi \in \mathcal{S}.$$

Remark 0.6.2. *Since Fourier transform is an isometry from $\mathcal{S}$ to itself, so above definition for the Fourier transform of distribution is meaningful and determines a element in $\mathcal{S}'$ uniquely and it is also an isometry in $\mathcal{S}'$ ("bijection" that preserves "distance"). Therefore, inverse of Fourier transform exists.*

Remark 0.6.3. *Note that we can embed the collection of Borel signed measures to the space of tempered distribution, that is, $\mathcal{M}(\mathbb{R}^d) \hookrightarrow \mathcal{S}'$. So the convergence of (tempered) distribution is $\langle \mu_n, f \rangle \to \langle \mu, f \rangle$ for all $f \in \mathcal{S}$ which is dense in $C_0(\mathbb{R}^d)$ under $\|\cdot\|_\infty$ by Stone Weierstrass Theorem, which requries the underlying space to be locally compact Hausdorff. We will see below that convergence of tempered distribution is the same as weak (distirbutional) convergence for probability measures. So from this point of view, the most natural definition of weak convergence is convergence in the space of tempered distributions, at least for $\mathbb{R}^d$. Fourier transforms does not work well (to my knowledge) in infinite dimensional space, so if one needs similar result in that case, this method would not work.*

Definition 0.6.5. *We say $f : E \to \mathbb{C}$ is positive definite if for all finite set of $\{z_n\} \subset \mathbb{C}$ then for all $\{e_n\} \subset E$ we have*

$$\sum_{n,k} f(e_n - e_k) z_n \overline{z_k} \geq 0.$$

Finally, there is a representation theorem tells us there is one to one correspondence between the family of probability measures and the family of positive definite functions. Here we state it without proof, which can be found in ([CZ01]) for the real case, and ([MS13]) for the locally compact abelian group case, and ([DPZ14]) for the infinite dimensional case, here we state the theorem in ([CZ01])

Proposition 0.6.2. *f is a characteristic (fourier transform) of a probability measure if and only if f is positive definite, f continuous at 0 and $f(0) = 1$.*

Before we talk about the main result here, there is a good application of Fourier transform: the independence relation of random variables can be determined through their characteristic function (fourier transform):

Theorem 0.6.1 (Kac'). *Let X, Y be an $\mathbb{R}^n$ valued random variable, then $X \perp Y$ if and only only if*

$$\mathbb{E}[e^{i\xi \cdot X}]\mathbb{E}[e^{i\eta \cdot Y}] = \mathbb{E}\left[e^{i(\eta,\xi)\cdot(X,Y)}\right]$$

for all $\xi, \eta \in \mathbb{R}^n$.

Remark 0.6.4. *Note that independence relation is really the relationship between measures instead of random variables.*

Proof. ($\Rightarrow$) direction is by definition.

($\Leftarrow$): Let $\widetilde{X}, \widetilde{Y}$ be independent random variables whose distributions are idenitcal to X, Y respectively. Denote $\mu_\cdot$ as the measure induced by $\cdot = X, Y$, then we have

$$\widehat{\mu_{(X,Y)}} = \widehat{\mu_X}\widehat{\mu_Y} = \widehat{\mu_{\widetilde{X}}}\widehat{\mu_{\widetilde{Y}}} = \widehat{\mu_{\widetilde{X},\widetilde{Y}}}$$

so by uniqueness of characteristic functions, we see that (X, Y) and $(\widetilde{X}, \widetilde{Y})$ has the same distributions, so X, Y are independent. □

Since we are talking about convergence, Levy's continuity theorem characterizes weak convergence of measure completely.

Proposition 0.6.3 (Levy). *Let $\{\mu_n\} \subset \mathcal{M}(\mathbb{R}^n)$ be a sequence of probability measures, then μ_n converges weakly if and only if its fourier transform $\hat{\mu}_n$ converges pointwise.*

Proof. Note that the definition of Fourier transform of measures and Fourier transform of tempered distributions are identical. Note also that $C_c^\infty(\mathbb{R}^d) \subset \mathcal{S}$ and they are dense in $C_0(\mathbb{R}^d)$ by Stone-Weierstrass, so $\mathcal{S}$ is dense in $C_0(\mathbb{R}^d)$.

First suppose weak convergence of measure, then the pointwise convergence is immediate by definition of weak convergence.

Now suppose $\hat{\mu}_n \to \mu$ pointwise, then we have

$$\int_{\mathbb{R}^d} \phi(\xi) d\mu_n(\xi) = \langle \mu_n, \hat{\phi} \rangle = \langle \hat{\mu_n}, \phi \rangle = \int_{\mathbb{R}^d} \hat{\mu}_n(\xi)\phi(\xi) d\xi \to \int_{\mathbb{R}^d} \hat{\mu}(\xi)\phi(\xi) d\xi = \langle \mu, \hat{\phi} \rangle = \int_{\mathbb{R}^d} \hat{\phi}(\xi) d\mu(\xi)$$

where the convergence is given by dominated convergence theorem. By the fact that fourier transform is a surjective map from $\mathcal{S}$ to itself, then we have $\langle \mu_n, \phi \rangle \to \langle \mu, \phi \rangle$ for all $\phi \in \mathcal{S}$. However, μ_n, μ are also continuous linear functional of $C_0(\mathbb{R})$, so by density the convergence holds for all $C_0(\mathbb{R}^d)$ functions.

Here our last step is to show the convergence also holds for all bounded continuous function. Let $f : \mathbb{R}^d \to \mathbb{R}$ be any bounded continuous function and let $\alpha > 0$ such that $\mu(B_0(\alpha)) \geq 1 - \epsilon$ for some small ϵ where $B_0(r)$ is the closed ball with radius r (regularity of borel measures). Finally let $h \in C_c^\infty(\mathbb{R}^d)$ with $h = 1$ in $B_0(\alpha)$ and $h = 0$ on $B_0(\alpha + 1)$, this is possible by Lusin's theorem. For simplicity in typing, let $K_\epsilon = B_0(r)$ here. Now we consider the following:

$$\begin{aligned}\left| \int f d\mu_n - \int f d\mu \right| &\leq \left| \int f d\mu_n - \int f d\mu - \int hf d\mu + \int hf d\mu + \int fh d\mu_n - \int hf d\mu_n \right| \\ &\leq \left| \int f d\mu_n - \int fh d\mu \right| + \left| \int fh d\mu_n - \int fh d\mu \right| + \left| \int fh d\mu_n - \int f d\mu \right|\end{aligned}$$

Now, when $n \to \infty$, the middle term disapear by what we proved before. The first and third terms are similar, so we only need to show that one of them can ger arbitrarily small. Here we consider the first term, and note that $\int f d\mu_n = \int_{K_\epsilon}$

$$\begin{aligned}\left| \int f d\mu_n - \int fh d\mu \right| &\leq \left| \int f d\mu_n - \int fh d\mu_n \right| + \left| \int fh d\mu_n - \int fh d\mu \right| \\ &\leq \|f\|_\infty \left| \int 1 - h d\mu_n \right| + \left| \int fh d\mu_n - \int fh d\mu \right| \\ &= \|f\|_\infty \left| 1 - \int h d\mu_n \right| + \left| \int fh d\mu_n - \int fh d\mu \right|\end{aligned}$$

and we note that limit sup of the first term is bounded by $\|f\|_\infty \mu(K_\epsilon^c) \leq \|f\|_\infty \epsilon$ which can be made arbitrarily small, and the limit of the second term is zero. So we conclude that $|\int f d\mu_n - \int f d\mu|$ can be made arbitrarily small, hence it is zero. $\square$

From the last step of the above proof, we also see that we can define convergence of probability distribution under more relaxing condition

Lemma 0.6.2. *The following are the "same"*

- $\int f d\mu_n \to \int f d\mu$ *for* $f \in C_c(\mathbb{R})$.
- $\int f d\mu_n \to \int f d\mu$ *for* $f \in C_b(\mathbb{R})$.

This is pretty cool since $C_c \subset C_b$ is not dense in sup norm.

Since we are talking about convergence of measures, here is another equivalent way of defining it

Definition 0.6.6. $\mu_n \Rightarrow \mu$ *if* $\lim_n \mu_n(E) = \mu(E)$ *for all* E *borel set with* $\mu(\partial E) = 0$.

The equivalence of this and the above definiton are easily seen in $\mathbb{R}^n$ with step function approximation and Urysohn's lemma.

Definition 0.6.7 (Weakly Compact)**.** *Let* $\{\mu_n\}$ *be a family of probability measures on a complete seperable metric space* E*, then we say* $\{\mu_n\}$ *is weakly compact if for every subsequence, there is a further subsequence that converges weakly.*

Remark 0.6.5. *Here, the concept of weakly compactness arises naturally from functional analysis when we treat borel probability measures as a subset of the dual of* $C_b(E)$*, the space of bounded continuous functions on* E *with* $\|\cdot\|_\infty$ *norm. For this reason, the above definition of weakly convergence, or convergence in distribution is the most natural way to define this concept.*

There is an easy characterization of weakly compactness of probability measures, which is the tightness. We now define and prove the equivalence between those two concepts.

Definition 0.6.8. *Let* Λ *be a family of probability measures on the complete seperable metric space* E*, we say the family* Λ *is said to be tight if for all* $\epsilon > 0$*, there is a compact* K_ϵ *such that*

$$\mu_\lambda(K_\epsilon) \geq 1 - \epsilon \quad \forall \mu_\lambda \in \Lambda.$$

Remark 0.6.6. *Here, inner regularity of borel measures are assumed.*

Proposition 0.6.4. *Let* Λ *be a family of probability measures, then* Λ *is weakly compact if and only if* Λ *is tight.*

Proof. We first prove the case for which E itself is already compact, then extend it to the general case with diagonalizatio argument.

Suppose first that E is compact, then any family on E is automatically tight. We recall that $C(E)$ is also seperable (separability $\iff$ metrizable), so take $\{\mu_n\}$ be any sequence contained in Λ and let $\{f\} \subset C(E)$ be a dense subset. Here we use diagonalization to create a weakly convergent subsequence. We note that $\{\int_E f_n \mu_k\}_{k\in\mathbb{N}}$ is a sequence of bounded real numbers for each $n \in \mathbb{N}$, therefore, it contains a convergent subsequence. For $n = 1$, call the corresponding subsequence of measures $\left\{\mu_k^{(1)}\right\}_{k\in\mathbb{N}}$. Now we construct subsequence inductively: suppose $\left\{\mu_k^{(n)}\right\}_{k\in\mathbb{N}}$ is a subsequence of $\{\mu_k\}$ such that

$$\int_E f_j d\mu_k^{(n)} \to \alpha_j \quad \text{for some } \alpha_j \in \mathbb{R} \text{ for all } j \leq n.$$

For the sequence $\left\{ \int_E f_{n+1} d\mu_k^{(n)} \right\} \subset \mathbb{R}$, there is a subsequence that converges to some number α_{n+1}, then call the corresponding subsequence of probability measures $\left\{ \mu_k^{(n+1)} \right\}_{k\in\mathbb{N}} \subset \left\{ \mu_k^{(n)} \right\}_{k\in\mathbb{N}}$. Let $\nu_n \triangleq \mu_n^{(n)}$, then

$$\lim_{n\to\infty} f_k d\nu_n = \lim_{n\to\infty} f_k d\mu_n^{(k)} = \alpha_k.$$

Now observe that $f_k \to \lim_{n\to\infty} \int_E f_k d\nu_n$ defines a continuous linear functional on $\{f_n\}$ which is dense in $C(E)$. Then we can extend it to $C(E)$ and by Rietz representation theorem on compact Hausdorff space, this map is the integration with respect to some probability measure.

Now for the general case: First assume it is tight, let $K_m \subset E$ be compact set such that $\mu_n(E_m) \geq 1 - \frac{1}{m}$ for all n and for all m. Now we restrict μ_n to K_m and call the restriction $\left\{ \mu_n^{(m)} \right\}$. Again, choose $\{f_k\} \subset C(E)$ to be a sequence that is dense. Now, let $E_m = \bigcup_{k=1}^m K_m$, which is again compact. So by above argument, there is a subsequence, call $\nu_n^{(m)}$ such that

$$\lim_{n\to\infty} \int_{E_m} f_k d\nu_n^{(m)} = \int_{E_m} f_k d\nu^{(m)}$$

for some probability measure on E, note that $\nu^{(n)}$ and $\nu^{(k)}$ agrees on E_k assuming $n \geq k$. We define the measure

$$\mu(A) = \lim_{n\to\infty} \nu^{(n)} (A \cap E_n)$$

The converse: assume by contradiction that $\{\mu_n\}$ is not tight but there is a subsequence that converges weakly. That is, there is a number $\alpha \in (0,1)$ such that

$$\mu_n(K) < 1 - \alpha, \quad \forall K \text{ compact}.$$

Now choose A be compact, continuity set such that $\mu(A) > 1 - \frac{a}{2}$ would get a contradiction. □

Those are all the general convergence result for almost surely, in probability and in distribution I have. The following two sections are special results like Law of large number, central limit theorem and Kolmogorov's condition for convergence of series of random variables.

0.7 Law of Large Numbers

The idea of law of large numbers is that when we have a sequence of similar random variables, say $\{X_n\}$ that have the same expectation ($\mathbb{E}[X_n] = \mu$ for all n), then the average of partial sum would converge to their mean in some sense (prop or a.e.), formally, if we let $S_n = \sum_{i=1}^n X_i$, then we expect

$$\frac{S_n - \mu}{n} \to 0 \quad \text{,as } n \to \infty \text{ in some sense.} \tag{5}$$

There are many versions of this law, roughly speaking, the ones that have convergence a.e. are called strong law, and the ones that have convergence in probability is called weak law. Since we (I) spent a lot of time on Fourier transforms above, let's first do one version with a simple proof via Fourier transform. But first we need the following lemma:

Lemma 0.7.1. *Suppose $\{X_n\}$ converges to $a \in \mathbb{R}$ in distribution, then $X_n \to a$ in probability.*

Proof. Let μ_n be the induced measure on $\mathbb{R}$ of X_n, and let $B_r(a)$ be the unit ball (interval) centered at a with radius r. For fixed r let f_r be a continuous function defined as follows:

$$f_r(x) = \begin{cases} 1 & x \in B_{\frac{\epsilon}{2}}(a) \\ 0 & x \notin B_\epsilon(a) \end{cases}$$

Such function exists by Urysohn's lemma.

$$\mathbb{P}\left(|X_n - a| < \epsilon\right) = \int_\Omega 1_{B_\epsilon(a)}(X(\omega))d\mathbb{P}(\omega) = \int_{\mathbb{R}} 1_{B_\epsilon(a)}(x)d\mu_n(x) \geq \int_{\mathbb{R}} f_\epsilon(x)d\mu_n(x) \to 1.$$

□

Before we dive into the weak law of large number, we need certain differentiation result for Fourier transform of probability measure.

Lemma 0.7.2. *Let μ be a probability measure with mean μ, that is, $\int_x x d\mu(x) = \mu$. Then the fourier transform of μ, call $\hat{\mu}$ is differentiable and $\frac{d\hat{\mu}}{d\xi}(0) = 2\pi i\mu$.*

Proof.

$$\begin{aligned} \frac{\hat{\mu}(\xi + h) - \hat{\mu}(\xi)}{h} &= \frac{1}{h}\left(\int_{\mathbb{R}} \exp(-2\pi i(\xi + h)x) - \exp(-2\pi i\xi x)d\mu\right) \\ &= \int_{\mathbb{R}} \exp\left(-2\pi i\xi x\right) \frac{\exp\left(-2\pi ihx\right) - 1}{h} d\mu(x) \end{aligned}$$

Taylor expansion of the exponential function tells us $\exp(-2\pi ihx) = 1 + 2\pi ixh + o(h)$, where $o(h)$ term means $\lim_{h\to 0} \frac{o(h)}{h} = 0$. So by dominated convergence theorem we see that when we take the limit, the above integral is equal to

$$-2\pi i \int_{\mathbb{R}} \exp(-2\pi i\xi x)x d\mu(x)$$

which is integrable, and at $\xi = 0$, the derivative is $-2\pi i\mu$. □

The following proposition is one of the weak law of large number that is easy to prove now.

Proposition 0.7.1. *Let $\{X_i\}$ be a sequence of i.i.d. random variable such that $X_i \in L^1(\mathbb{P})$ and $\mathbb{E}[X_i] = \mu$ for all i. Then if we let $S_n = \sum_{i=1}^n X_i$, then*

$$\frac{S_n - n\mu}{n} \to_p 0. \quad or \quad \frac{S_n}{n} \to_p \mu$$

Proof. We show that the Fourier transform of the left hand side limit converges to the fourier transform of delta function at zero. In that case we have convergence in distribution to a constant in $\mathbb{R}$, hence convergence in probability. Here we assume WLOG that $\mu = 0$.

Fourier transform of a point mass at zero: δ_0.

$$\int_{\mathbb{R}} \exp(-2\pi i\xi x)d\delta_0(x) = 1$$

Now consider the Fourier transform of the average of the partial sums:

$$\begin{aligned}
\mathbb{E}\left(\exp\left(-2\pi i\frac{\xi}{n}\left(\sum_{j=1}^{n}X_n\right)\right)\right) &= \prod_{j=1}^{n}\mathbb{E}\left[\exp\left(-2\pi i\frac{\xi}{n}X_j\right)\right] \\
&= \left\{\mathbb{E}\left[\exp\left(-2\pi i\frac{\xi}{n}X_j\right)\right]\right\}^n \\
&= \left(1-\frac{2\pi i\mu\xi}{n}+o(\frac{\xi}{n})\right)^n \\
&\to \exp(-2\pi i\mu\xi)
\end{aligned}$$

Since $\mu = 0$, we have pointwise convergence, hence convergence in distribution by Levy's theorem. Then by (5) we have convergence in probability. □

Note that if the assumption is that $\mathbb{E}[|X_n|^p] < \infty$, then by Jessen we see that $\mathbb{E}[|X_n|] < \infty$.

There is another weak law of large numbers that does not require total independence of the random variables, but rather pairwise independence. To prove such thing, we need the notion of equivalence of two sequences of random variables:

Definition 0.7.1. *Let $\{X_n\}$ and $\{Y_n\}$ be two sequence of random variables, then we say they are equivalent if*

$$\sum_{n\in\mathbb{N}}\mathbb{P}[X_n \neq Y_n] < \infty.$$

In other words, $\mathbb{P}\left(X_n \neq Y_n, I.O\right) = 0$, or $\mathbb{P}\left(\bigcup_{m\geq 1}\bigcap_{n\geq m} X_n = Y_n\right) = 1$ by Borel Cantelli.

The word equivalence should give us a hint that the behavior of the partial sums of the two sequence are basically the same:

Theorem 0.7.1. *let $\{X_n\}$ and $\{Y_n\}$ be two equivalent sequence, then*

$$\sum_{n=1}^{\infty}(X_n - Y_n) < \infty \quad a.e.$$

and if $\alpha_n \uparrow \infty$, then

$$\frac{1}{\alpha_n}\sum_{n=1}^{\infty}(X_n - Y_n) = 0 \quad a.e.$$

Proof. Note that the second convergence result is a trivial consequence of the first one.

By Borel-Cantelli, we see that there is $\widetilde{\Omega} \subset \Omega$ such that $\mathbb{P}[\widetilde{\Omega}] = 1$ and for all $\omega \in \widetilde{\Omega}$, there is $n_0 \geq 1$ such that

$$X_n(\omega) = Y_n(\omega) \quad \forall n \geq n_0.$$

More explicitely, for all $\omega \in \widetilde{\Omega}$, the sum $\sum_{n=1}^{\infty}(X_n(\omega) - Y_n(\omega))$ only has finitely many nonzero terms, hence convergence a.e. □

Now we are ready to prove the weak law of large number under a more relaxing condition:

Proposition 0.7.2. *Let $\{X_n\}$ be a sequence of random variable that is identically distributed and pairwise independent, and assume that $\mathbb{E}[X_n] = \alpha < \infty$. Let S_n denote its n^{th} partial sums, then*

$$\frac{S_n}{n} \to_p \alpha$$

Proof. Since we talked about equivalence sequence, it gives a hint that the convergence is proven by constructing an equivalent sequence that is easier to work with. We shall use truncation method to construct such a sequence: define $\{Y_n\}$, a sequence of random variables as follows:

$$Y_n = \begin{cases} X_n & \text{on the set } |X_n| \le n \\ 0 & \text{elsewhere.} \end{cases}$$

Now it is easy to see that $\{X_n\}$ and $\{Y_n\}$ are equivalent because

$$\mathbb{E}[|X_1|] < \infty \Leftrightarrow \sum_{n=1} \mathbb{P}[|X_1| > n] < \infty \quad \text{(Prop (19))}$$

$$\text{(identically distributed)} \Leftrightarrow \sum_{n=1}^{\infty} \mathbb{P}[|X_n| > n] = \sum_{n=1}^{\infty} \mathbb{P}[X_n \neq Y_n] < \infty.$$

So now we only need to show convergence for Y_n, we use Chebyshev Inequality for that. Denote the partial sums of Y_n by S'_n and assume WLOG that $\alpha = 0$. Note that since each of Y_n is a bounded, then it is square integrable. To do that, we first consider the variance of S'_n. Denote μ the measure on $\mathbb{R}^n$ generated by X_1 (X_n's are identically distributed!), then

$$\text{var}(S'_n) = \sum_{j=1}^{n} \text{var}(Y_j) = \sum_{j=1}^{n} \int_{|x| \le j} x^2 d\mu \le \sum_{j=1}^{n} j \int_{|x| \le n} |x| d\mu \le \frac{n(n-1)}{2} \mathbb{E}[|X_1|].$$

Clearly this is not the bound we want, but it does give us an hint on what to do. To improve the bound, we use $\log(n)$ to divide the sum, and note that $\lim_{n \to \infty} \frac{\log(n)}{n} = 0$.

$$\begin{aligned} \sum_{j=1}^{n} \int_{|x| \le j} x^2 d\mu &= \sum_{1 \le j \le \log(n)} \int_{|x| \le j} x^2 d\mu + \sum_{\log(n) < j \le n} \int_{|x| \le j} x^2 d\mu \\ &\le \sum_{1 \le j \le \log n} \log(n) \mathbb{E}|X_1| + \sum_{\log n < j \le n} \left(\log n \int_{|x| \le \log n} |x| d\mu + j \int_{\log n < |x| \le j} |x| d\mu \right) \\ &\le n \log n \mathbb{E}|X_1| + n^2 \int_{|x| > \log n} |x| d\mu \end{aligned}$$

With this bound, we see that

$$\frac{\text{var}(S'_n)}{n^2 \epsilon^2} = \frac{1}{\epsilon^2} \left(\frac{\log n}{n} \mathbb{E}[|X_1|] + \int_{|x| > \log n} |x| d\mu \right) \to 0.$$

So we have

$$\mathbb{P}\left(|S'_n - \mathbb{E}[S'_n]| / n > \epsilon \right) \le \frac{\text{var}(S'_n)}{(n\epsilon)^2} \to 0 \quad \forall \epsilon > 0.$$

So

$$\frac{S'_n}{n} - \frac{\mathbb{E}[S'_n]}{n} \to 0, \quad \text{in probability.}$$

However, since $\mathbb{E}[Y_n] \to \mathbb{E}[X_1] = 0$ by DCT, the second term on the left hand side tends to zero, since it is an average of things that tends to zero. The convergence result for S_n is given by the second part of Theorem 7. □

0.8 Convergence of Random Series

The next goal is to prove the Strong Law of Large Numbers. As far as I know, there are at least two ways of doing this, both of which are contained in ([Dur19]). Here I will take the longer route of proving it using Kolmogorov's three series theorem due to its imporance.

Let's first talk about the Kolmogorov's Maximal Inequality (to distinguish this from Hardy-Littlewood maximal Inequality).

Proposition 0.8.1. *Let $\{X_n\}$ be a sequence of independent random variable with mean zero and variance $\sigma_n = \mathbb{E}[|X_n|^2] < \infty$ for all $n \in \mathbb{N}$. Then we have the following Inequality for maximum of its partial sum $S_n = \sum_{k=1}^n X_k$:*

$$\mathbb{P}\left[\max_{1\le k\le n} |S_k| \ge \epsilon\right] \le \frac{\mathbb{E}|S_n|^2}{\epsilon^2} = \frac{var(S_n)}{\epsilon^2}. \tag{6}$$

Proof. Remove max, then it is just Chebyshev's Inequality. Treating $S_\cdot : \mathbb{N} \to \mathbb{R}$ as a process, then we can divide the case into at which step $|S_k|$ first greater than ϵ. Here is what I mean:

Let $A_1 = \{|X_1| \ge \epsilon\}$, and

$$A_{k+1} = \left(\bigcup_{1\le j\le n} A_j\right)^c \cap \{|S_{n+1}| \ge \epsilon\}.$$

Note that $A_i \cap A_j = \emptyset$ and A_j is the event that the first time the process $|S_\cdot|$ corsses ϵ is at j^{th} step. Now consider the right hand side of the Inequality in the statement: Let μ be the underlying probability measure,

$$\begin{aligned}
\int |S_n|^2 d\mu &= \sum_{k=1}^n \int_{A_k} |S_n|^2 d\mu \\
&= \sum_{k=1}^n \int_{A_k} ((S_n - S_k) + S_k)^2 \, d\mu \\
&= \sum_{k=1}^n \int_{A_k} (S_n - S_k)^2 + 2S_k(S_n - S_k) + S_k^2 d\mu \\
&= \sum_{k=1}^n \left(\int_{A_k} (S_n - S_k)^2 d\mu + \int_{A_k} 2S_k(S_n - S_k) d\mu + \int_{A_k} S_k^2 d\mu \right)
\end{aligned}$$

We note $1_{A_k} S_k \in \sigma(X_1, ..., X_k)$ and $S_n - S_k \in \sigma(X_{k+1}, ..., S_n)$, so they are independent, so the middle term in the above sum disapears. Now if we ignore the first term, then we have the following Inequality

$$\int |S_n| d\mu \ge \sum_{k=1}^n S_k^2 d\mu \ge \sum_{1\le k\le n} \epsilon \mathbb{P}[A_k] = \epsilon^2 \mathbb{P}\left[\bigcup_{1\le k\le n} A_k\right] = \epsilon^2 \mathbb{P}[\max_{1\le k\le n} |S_k|^2 \ge \epsilon].$$

□

We note that the Maximal inequality in seperable Banach space holds if in addition X_n's are symmetric, means $X_n =_d -X_n$ (equal in distirbution). The proof is a bit different since we don't have "compete the square" in general Banach space, it can be found in ([DPZ14]).

In ([CZ01]), according to Chung, the next is step is to have a different bound for (6), and Durret ([Dur19]) uses the Central Limit Theorem to prove the Kolmogorov three series theorem. Time is limited for me, I'll have to use the easier method of CLT from the "future" to avoid technical details, I might come back and edit it later though.

Proposition 0.8.2. *Let $\{X_n\}$ be a sequence of independent random variables such that $\mathbb{E}[X_n] = \mu_n$ and $var(X_n) = \sigma_n^2$. Suppose $\sum_{n\in\mathbb{N}} \mu_n$ and $\sum_{n\in\mathbb{N}} \sigma_n^2$ both converges, then the series converges almost surely, that is,*

$$\lim_{N\to\infty} S_N = \sum_{n=1}^{N} X_n \quad \textit{converges a.s.} \tag{7}$$

Proof. We will show that the sequence of sum S_n are Cauchy using the maximal Inequality given in proposition (6). Let's say $n \geq m$, and assume WLOG that $\mu_n = 0$ for all n,

$$\begin{aligned} \mathbb{P}\left(\sup_{n\leq k\leq m} |S_n - S_k| \geq \epsilon\right) &= \mathbb{P}\left(\sup_{n\leq k\leq m} \left|\sum_{i=n}^{k} X_i\right| \geq \epsilon\right) \\ &\leq \frac{1}{\epsilon^2}\sum_{k=n}^{m} \sigma_n^2. \end{aligned}$$

Note that this goes to zero as $n, m \to \infty$, note further that this shows that the partial sums forms a cauchy sequence in sup norm, hence converges a.e.. □

Now we are ready to prove the three series theorem:

Proposition 0.8.3. *Let $\{X_n\}$ be a sequence independent random variables, its partial sum, $S_n = \sum_{k=1}^{n} X_k$ converges a.e. if some $A > 0$ such that the truncated sequence $\{Y_n\}$ where $Y_n = X_n 1_{|X_n|<A}$ we have that*

$$(1) \sum_{n=1}^{\infty} \mathbb{E}[Y_n] \quad \textit{converges;} \quad (2) \sum_{n\in\mathbb{N}} \mathbb{P}\left(|X_n| \geq A\right) \quad \textit{converges} \quad \textit{and } (3) \sum_{n=\in\mathbb{N}} var(Y_n) \quad \textit{converges.}$$

S_n converges only if the three converges is true for all $A > 0$.

Proof. **Sufficient**: It is easy to see that if the three series condition holds for some $A > 0$, then X_n and Y_n are equivalent sequence, so they either both converges or both diverges. By the two series theorem (prop 0.8), the the partial sums of Y_n converges, so S_n converges.

Necessity: First suppose $\sum_{n\in\mathbb{N}} \mathbb{P}[|X_n| > A]$ diverges for arbitrary large A, then by the second Borel-Cantelli, we have that for all $n \geq 1$ there is a $m \geq n$ such that $X_n > A$, hence the series would diverge.

Now if S_n converges, then (2) holds, then Kolmogorov's two series theorem tells us $\sum_{k=1}^{\infty} Y_n - \mathbb{E}[Y_n]$ converges, but since Y_n is truncated version of X_n, so $\sum_{n\in\mathbb{N}} Y_n$ converges, hence $\sum_{n=1}^{\infty} \mathbb{E}[Y_n]$ converges.

Finally, given S_n converges, then Y_n converges a.e. and bounded pointwise (in n), so we can use central limit theorem

$$\frac{1}{\sqrt{\text{var}(\sum_{n=1}^{N} Y_n)}} \sum_{n=1}^{N} (Y_n - \mathbb{E}[Y_n]) \to_d \mathcal{N}(0,1).$$

and without finiteness of the sum of variance of Y_n, this is not possible. □

Here is a useful result for convergence of random series:

Proposition 0.8.4. *Let $\{x_i\}$ be independent sequence of random variables, let $S_N = \sum_{n=1}^{N} X_n$. Then S_N converges almost surely if and only if it converges in distributions.*

Remark 0.8.1. *This proposition shows us that the three concepts of convergence are the same thing when we talk about sum of independence random variables.*

Proof. Now note that sequence of random variables converges if and only if they are cauchy in probability. One direction here is obvious, so we assume convergence in distribution to some probability measure ν, and assume by contradiction S_N does not converge a.e., that is, S_N is not cauchy in probability. So by Levy's continuity theorem and definition of cauchy in probability, we have

$$\hat{\mu}_n(\xi) \to \hat{\nu}(\xi) \quad \text{pointwie;} \tag{8}$$

where μ_n denotes the measure induced by S_n. Also, there exists $(N(n), M(n))$ increasing pair of natural numbers such there are $\delta, \epsilon > 0$ with

$$\mathbb{P}\left(\left|S_{N(n)} - S_{M(n)}\right| > \epsilon\right) > \delta. \tag{9}$$

We note that $\mu_n = \mu_k * \mu_{k,n}$ for $k < m < n$, where $\mu_{i,j}$ is the measure induced by $\sum_{k=i+1}^{j} X_n$. This is because $\sigma(S_n) \perp \sigma(\sum_{k=n+1}^{m} X_k)$. $*$ denotes convolution of measures defined as $\mu * \nu(\Gamma) = \int_\Omega \mu(\Gamma - \omega) d\nu(\omega)$ whenever the space Ω have nice linear structure (eg. seperable Banach space).

Now, by properties of Fourier transforms, $\widehat{\mu_k * \mu_{k,n}} = \hat{\mu}_k \hat{\mu}_{k,n}$, hence the convergence in distribution implies $\hat{\mu}_{N(n),M(n)} \to 1 \iff \mu \to_d \delta_0$ by a simple application of Levy's continuity theorem. However, (8) tells us that $\mu_{N(n),M(n)}(|x| > \epsilon) > \delta$ so a contradiction. □

Remark 0.8.2. *Here we see that we do not use any property that is unique to $\mathbb{R}$, so the same proof works for $\mathbb{R}^n$ or any space where the continuity theorem holds.*

0.9 The Strong Law of Large Numbers

To prove the strong law of large numebrs, we'll need the a very believable theorem from analysis which we state without proof:

Theorem 0.9.1 (Kronecker's Lemma). *Suppose $\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ converges for real numbers x_n, a_n for $a_n \uparrow \infty$, then*

$$\frac{1}{a_n} \sum_{k=1}^{n} x_n \to 0. \tag{10}$$

Theorem 0.9.2 (The Strong Law of Large Numbers). *Let $\{X_n\}$ be i.i.d with mean μ, then*

$$\lim_{n\to\infty} \frac{S_n - n\mu}{n} \to 0 \quad a.e. \tag{11}$$

Where $S_n = \sum_{i=1}^{n} X_i$ is the partial sum.

Proof. Assume WLOG that $\mu = 0$, let $Y_n = X_n 1_{|X_n| \le n}$, the truncated version of X_n, then by equivalence relation, we only need to show

$$\frac{\sum_{n=1}^{N} Y_n}{N} \to 0 \quad \text{as } N \to \infty.$$

Now by Kronecker's lemma (theorem 8), we only need to show

$$\sum_{n=1}^{N} \frac{Y_n}{n} \text{ converges.}$$

However, note that $\{\frac{Y_n}{n}\}$ forms a sequence of independent random variables, so by Proposition (29), we only need to show it converges in measure or probability or in L^2, so consider:

$$\mathbb{E}\left[\left(\sum_{n=1}^{N} \frac{Y_n}{n}\right)^2\right] = \sum_{n=1}^{N} \frac{\mathbb{E}\left[Y_n^2\right]}{n^2}$$

and we only have to bound the second moment of Y_n by some constant.

$$\begin{aligned}\sum_{n\in\mathbb{N}} \frac{\mathbb{E}[Y_n^2]}{n^2} &= \sum_{n\in\mathbb{N}} \frac{1}{n^2}\int_0^\infty 2t\mathbb{P}[|Y_n| \geq t]dt \\ &= \int_0^\infty \sum_{n\in\mathbb{N}} \frac{1}{n^2} 1_{t\leq n} 2t\mathbb{P}[|X_n| \geq t]dt \\ &= \int_0^\infty 2\left(\sum_{t\leq n} \frac{t}{n^2}\right)\mathbb{P}\left[|X_1| \geq t\right]dt\end{aligned}$$

To show this thing is finite, our task becomes show the sum in the integral is finite. However, observe that this sum is comparable to the following integral

$$2t\int_t^\infty \frac{1}{x^2}dx \leq 10.$$

Now by the fact that X_n's are integrable, we have the desired result. □

0.10 Central Limit Theorem

There are many versions of CLT, I only have the energy to write about couple of them, perhaps I'll add more later.

Remark 0.10.1. *To match with other texts in probability (avoid proofs), we redefine fourier transform of a measure to be $\hat{\mu}(\xi) = \int e^{-it\xi}d\mu(t)$ and use the term "fourier transform of" and "characteristic function of" interchanagably.*

First let's review what is a normal distribution:

Definition 0.10.1. *We say $X \sim \mathcal{N}(m, \sigma^2)$ on the real line if X has density function*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{\sigma^2}\right)$$

For $\mathbb{R}^n$, the definition is similar: We say $X \sim \mathcal{N}(m, \Sigma)$, where Σ is a positive definite matrix, if the measure induced by X, call it μ, is absolutely continuous with respect to λ^n, the n-dim lebesgues measure and

$$\frac{d\mu}{d\lambda^n}(x) = \frac{1}{(2\pi\det(\Sigma))}\exp\left(-\frac{1}{2}\langle x-m, \Sigma^{-1}(x-m)\rangle\right)$$

I will focus on the one dimensional case. I will refer the measure induced by an normal random variable by gaussian measure.

Proposition 0.10.1. *Let μ be a Guassian measure with mean m and variance σ^2, then its Fourier transform is*

$$\hat{\mu}(\xi) = \exp\left(-im\xi - \frac{1}{2}(\sigma\xi)^2\right).$$

Here is one that only requries a simple application of Levy's Continuity theorem, and the proof is basically the same as Proposition 24:

Proposition 0.10.2 (CLT 1)**.** *Let $\{X_n\}$ be i.i.d with finite mean m and finite variance σ^2, then we have the following weak convergence in distribution:*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} X_n - nm}{\sqrt{n}\sigma} \to_d \mathcal{N}(0,1) \tag{12}$$

where $\mathcal{N}(0,1)$ denote the standard normal distribution.

Proof. Here we assume WLOG that $m = 0$, let $\varphi_n(\xi) = \hat{\mu}_n(\xi)$, then we only need to show $\varphi_n \to \exp\left(-\frac{1}{2}\xi^2\right)$ pointwise.

$$\begin{aligned}
\varphi_n(\xi) &= \mathbb{E}\left[\exp\left(-i\frac{1}{\sqrt{n}\sigma}\sum_{k=1}^{n} X_n\right)\right] \\
&= \prod_{k=1}^{n} \mathbb{E}\left(\exp\left(-i\frac{X_n}{\sqrt{n}\sigma}\right)\right) \\
&= \left(\varphi_1\left(\frac{\xi}{\sqrt{n}\sigma}\right)\right)^n = \left(1 - \frac{1}{n\sigma^2}\sigma^2\xi^2 + o\left(\left(\frac{\xi}{\sqrt{n}\sigma}\right)^2\right)\right)^n \to \exp\left(-\frac{1}{2}\xi^2\right)
\end{aligned}$$

□

For triangular arrays that looks like

$$\begin{aligned}
&X_{1,1}, X_{1,2}, ..., X_{1,k_1}; \\
&X_{2,1}, X_{2,2}, ..., X_{2,k_2}; \\
&\vdots \\
&X_{n,1}, X_{n,2}, ..., X_{n,k_n}; \\
&\vdots
\end{aligned} \tag{13}$$

where we assume $\{X_{n,j}\}_{j=1}^{k_n}$ is a collection of independent random variables, but the random variables in different rows might be dependent. Let $\mathbb{E}[X_{n,j}] = \alpha_{nj} < \infty$ and $\operatorname{var}\left(X_{nj}\right) = \sigma_{nj}^2 < \infty$. We assume that $\alpha_{nj} = 0$ and $\sum_{j=1}^{k_n} \sigma_{n,j}^2 = 1$ for simplicity (rescaling makes this doable). We call the sum of the random variable in then^{th} row S_n, that is,

$$S_n = \sum_{j=1}^{k_n} X_{n,j}.$$

To state the Lindeberg-Feller's Central Limit Theorem in its full, we will need the following definition:

Definition 0.10.2. *We say a triangular array like (13) is holospoudic if*

$$\lim_{n\to\infty} \max_{1\le k\le k_n} \mathbb{P}\left[|X_k| > \epsilon\right] = 0, \quad \forall \epsilon.$$

Finally, let $\mu_{n,j}$ be the measure induced by the random variable X_{nj}. Here is the theorem:

Theorem 0.10.1. *Under the above set up we have*

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \int_{|X_{nj}|>\eta} |x|^2 d\mu_{nj} \to 0 \forall \eta > 0$$

(this is called the Lindeberg condition) if and only if the following two conditions holds:

1. $S_n \to \mathcal{N}(0,1)$ *in distribution.*

2. *The triangular array is holospoudic.*

Remark 0.10.2. *The full proof can be found in ([CZ01]) which he proved both sufficiency and necessity, but I will only prove sufficiency here. Chung used a method that can be applied to more general CLT. A proof similar to our proof can be found in ([Bil08]).*

We'd need the following two calculus theorems for the proof, which we state without proof for now, but I might come back latter to add the proof:

Theorem 0.10.2. *Let* $\{\theta_{n,k}, 1 \le n, 1 \le j \le k_n\}$ *be a triangular array of complex numbers, and suppose the following holds*

1. $\max_{1\le k\le k_n} |\theta_{n,k}| \to 0$ *as* $n \to \infty$.

2. $\sum_{k=1}^{k_n} |\theta_{nk}| < M$ *for all* n.

3. $\sum_{k=1}^{k_n} \theta_{nk} \to \theta$ *for some* $\theta \in \mathbb{C}$.

Then we have

$$\prod_{k=1}^{k_n} (1 - \theta_{n,k}) \to e^{\theta}.$$

Theorem 0.10.3. *Let* $u : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ *be a function of the positive integers such that*

$$\lim_{n\to\infty} u(m,n) \to 0 \quad \forall m \in \mathbb{N},$$

then there is a subsequence of the positive the integers, m_n *such that*

$$\lim_{n\to\infty} u(m_n, n) \to 0.$$

The essential idea of the proof is treat the sum of each raw as a random variable and show it converges in distribution.

Proof. Here is the structure of the proof: we will first use Theorem (12) to create an equivalent sequence of random variables to S_n such that we can apply Theorem(11) to its Fourier transform to show this sequence of random variables converges in distribution to the standarn normal distribution, hence S_n would too.

Note that for all $\eta \in \mathbb{N}$ we have

$$m^2 \sum_{k=1}^{n_k} \int_{|x|\geq m} |x|^2 d\mu_{n,k} \to 0.$$

So by theorem (12), there is a sequence of numbers that decrease to zero, call η_n such that

$$\lim_{n\to\infty} \frac{1}{\eta_n^2} \sum_{k=1}^{k_n} \int_{|x|\geq \eta_n} |x|^2 d\mu_{n,k} = 0. \tag{14}$$

Now we let $Y_{n,k} = X_{n,k} 1_{|X_{n,k}|\geq \eta_n}$, the truncated version of $X_{n,k}$ at the point η_n for each row, and denote $S'_n = \sum_{k=1}^{k_n} Y_{n,k}$, I will show at the end of the proof that S'_n and S_n are equivalent. We first show $|\mathbb{E}[S'_n]|$ goes to zero and show that $\text{var}(S'_n) = \mathbb{E}[S_n^2] - \mathbb{E}[S'_n] \to 1$. First we observe that

$$\mathbb{E}[Y_{n,k}] = \int_{\mathbb{R}} x d\mu_{n,k} - \int_{|x|\geq \eta_n} x d\mu_{n,k} = -\int_{|x|\geq \eta_n} x d\mu_{n,k}$$

$$\left|\mathbb{E}[S'_n]\right| = \sum_{k=1}^{n_k} \int_{|x|\geq \eta_n} |x| d\mu_{n,k} \leq \sum_{k=1}^{n_k} \frac{1}{\eta_n} \int_{|x|\geq \eta_n} |x|^2 d\mu_{n,k} \to 0. \tag{15}$$

Also for variance:

$$\text{var}(S'_n) = \mathbb{E}\left(\sum_{k=1}^{n_k} (Y_{n,k} - \mathbb{E}[Y_{n,k}])\right)^2 = \sum_{k=1}^{n_k} \mathbb{E}[Y_{n,k}^2] - \mathbb{E}[Y_{n,k}]^2$$

However,

$$\sum_{k=1}^{n_k} (\mathbb{E}[Y_{n,k}])^2 \leq \sum_{k=1}^{n_k} \int_{|x|\geq \eta_n} |x|^2 d\mu_{n,k} \int_{|x|\geq \eta_n} 1 d\mu_{n,k} \to 0.$$

However, we note that the Lindeberg condition given is the same as

$$\lim_{n\to\infty} \sum_{k=1}^{n_k} \int_{|x|<\eta_n} |x|^2 d\mu_{n,k} = \lim_{n\to\infty} \sum_{k=1}^{n_k} \mathbb{E}[Y_{n,k}^2] = 1. \tag{16}$$

Now, let's consider the Fourier transform of S'_n:

$$\begin{aligned}
\mathbb{E}\left[\exp\left(-i\xi S;_n\right)\right] &= \prod_{k=1}^{n_k} \mathbb{E}\left[\exp\left(-i\xi Y_{n,k}\right)\right] \\
&= \prod_{k=1}^{n_k} \left(1 - i\xi\mathbb{E}[Y_{n,k}] - \frac{1}{2}\left(\xi^2\mathbb{E}[Y_{n,k}^2]\right) + o(\xi^2\mathbb{E}[Y_{n,k}^2])\right)
\end{aligned}$$

We would notice that we gathered most of the conditions in theorem (11): (15) combined with (16) would get us condition (2) and (3), and (14) would get us condition (1), and since $\mathbb{E}[|Y_{n,k}^2] \to 0$

uniformly in k as $n \to \infty$, we do not have to worry about the small o term. Then by theorem (11) we have convergence in distribution.

Finally, we show that S'_n and S_n are equivalent sequence of random variables:

$$\mathbb{P}\left[S_n \neq S'_n\right] = \mathbb{P}\left[\bigcup_{1 \le k \le n_k} |X_{n,k}| \ge \eta_n\right] \le \sum_{1 \le k \le n_k} \mathbb{P}\left[|X_{n,k}| \ge \eta_n\right] \le \frac{1}{\eta_n} \int_{|x| \ge \eta_n} |x|^2 d\mu_{n,k}$$

which goes to zero by (14). □

Chapter 1

Discrete Martinagles

This section is focused on Discrete Martingale theory and some Markov Theory. Too many theories in the topics of modern probability relies on those two. The structure is like the combination of ([CZ01]) and ([Dur19]). It will be break down into the following: conditional expectation, Markov properties and Discrete Martingales.

1.1 Conditional Probability

Here is an intuitive route to get to the definition of conditional expectation:

Definition 1.1.1. *Let $\Omega, \mathcal{F}, \mathbb{P}$ be a probability measure space, and let $\Lambda \subset \mathcal{F}$, then we define the conditional probability with respect to Λ, call $\mathbb{P}_\Lambda(\cdot)$ or $\mathbb{P}(\cdot|\Lambda)$ to be*

$$\mathbb{P}_\Lambda(A) = \frac{\mathbb{P}(A \cap \Lambda)}{\mathbb{P}(A)}, \quad \forall A \in \mathcal{F}.$$

We define the conditional expectation with respect to Λ, call $\mathbb{E}_\Lambda[\cdot]$ to be

$$\mathbb{E}_\Lambda[X] = \int_\Lambda X d\mathbb{P}, \quad \textit{for all random variable } X \textit{ that makes sense.}$$

Now we let $\{\Lambda_n\}$ be a countable partition of Ω (e.g. generated by discrete random variables), and call $\mathcal{G} \subset \mathcal{F}$ the σ-field generated by $\{\Lambda_n\}$. Then for any random variable $X \in \mathcal{F}$ we define a new function $\mathbb{E}_\mathcal{G}(X)$ by

$$\mathbb{E}_\mathcal{G}(X)(\cdot) = \sum_{n \in \mathbb{N}} 1_{\Lambda_n} \mathbb{E}_{\Lambda_n}(X).$$

So this $\mathbb{E}_\mathcal{G}(X)(\cdot)$ takes countable values as well, and note that it is only $\mathcal{G}$-measurable. Furthermore, for any $A \in \mathcal{G}$, we see that

$$\int_A \mathbb{E}_\mathcal{G}[X] d\mathbb{P} = \int_A X d\mathbb{P}.$$

For general $\mathcal{G}$ (not neccessarily countably generated), we here show uniqueness and existance of conditional expectation.

Proposition 1.1.1. *Let X be a L^1 r.v. defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for any σ-sub-field $\mathcal{G}$ of $\mathcal{F}$, there is a unique random variable that is $\mathcal{G}$ measurable, call it $\mathbb{E}[X|\mathcal{G}]$, such that*

$$\int_A \mathbb{E}[X|\mathcal{G}] d\mathbb{P} = \int_A X d\mathbb{P} \quad \forall A \in \mathcal{G}.$$

Proof. Uniqueness first: suppose φ_1 and φ_2 both satisfies the above property, then

$$\int_A \varphi_1 d\mathbb{P} = \int_A \varphi_2 d\mathbb{P} \quad \forall A \in \mathcal{G}.$$

Note that $\{\varphi_1 > \varphi_2\}$ and $\{\varphi_1 < \varphi_2\}$ are both in $\mathcal{G}$, so

$$\int_{\varphi_1 - \varphi_2 > 0} \varphi_1 - \varphi_2 d\mathbb{P} = 0$$

so this must have measure zero, same for the other set, so we must have $\varphi_1 = \varphi_2$ a.e., and since we do not distinguish functions that disagree on set of measure zero, so this is unique.

Now for existance, define $\nu(\cdot) : \mathcal{G} \to \mathbb{R}$ to be

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

Then by properties of integral, this is a signed measure. Also, if $\mathbb{P}(A) = 0$, then $\nu(A) = 0$ as well, so $\nu << \mathbb{P}$. So by Radon-Nikodym, there is a function, call $\mathbb{E}[X|\mathcal{G}] = \frac{d\nu}{d\mathbb{P}}$. □

In conclusion, conditional expectation is just the Radon-Nikodym derivative of the signed measure generated by integrating X on some sub σ field with respect to the given probability measure. Here we repeat what we talked about

Definition 1.1.2. *Let X be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$, given a sub σ field, we define $\mathbb{E}[X|\mathcal{G}]$ to be the random variable such that*

1. *is $\mathcal{G}$ measurable.*

2. *have the same integral as X on $\mathcal{G}$.*

Remark 1.1.1. *In the example given before the definition we see that in countable case we can represente the conditional expectation explicitely, but we don't always have such representation (see [CZ01]).*

Here are some basic properties of conditional expectation, they are comparable to usual expectations (every equality below means a.s. equal):

Theorem 1.1.1. *Let X, Y, Z be integrable random variables on the probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ and let $\mathcal{G} \subset \mathcal{F}$ be a σ subfield, then the following is true:*

1. *If $X \in \mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = X$.*

2. *If $Z \in \mathcal{G}$ $(\sigma(Z) \in \mathcal{G})$, then $\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$ for ZX integrable.*

3. *If $X \leq Y$, then $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$.*

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

5. $X_n \uparrow X \Rightarrow \mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$.

6. $X_n \downarrow X \Rightarrow \mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$.

7. *$|X_n| \leq Y$ where Y integrable, and if $X_n \to X$, then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$.*

8. *Let φ be a convex function, then $\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}]$ whenever $\varphi(X)$ is integrable.*

9. *Suppose X, Y are square integrable, then* $\mathbb{E}[|XY||\mathcal{G}] \leq \left(\mathbb{E}[|X|^2|\mathcal{G}]\right)^{\frac{1}{2}} \left(\mathbb{E}[|Y|^2]|\mathcal{G}\right)^{\frac{1}{2}}$.
10. *If $\mathcal{H} \subset \mathcal{G}$, then* $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right]|\mathcal{H}\right] = \mathbb{E}\left[\mathbb{E}\left[X|\mathcal{H}\right]|\mathcal{G}\right] = \mathbb{E}[X|\mathcal{H}]$.
11. *If $\mathcal{F}_1 \subset \mathcal{F}_2$, then* $\mathbb{E}[X|\mathcal{F}_1] = \mathbb{E}[X|\mathcal{F}_2]$ *if and only if* $\mathbb{E}[X|\mathcal{F}_2] \in \mathcal{F}_1$.

Here I prove some of them, most of them should be trivial implications of the ordinary cases.

Proof. (2) Here we need to show

$$\forall \Lambda \in \mathcal{G}, \quad \int_\Lambda Z\mathbb{E}[X|\mathcal{G}]d\mathbb{P} = \int_\Lambda ZXd\mathbb{P}.$$

We note that if $Z = 1_A$ for some $A \in \mathcal{G}$, then it is automatically true, hence true for any step functions. We also recall a fact from real analysis that we can approximate any function by a sequence of step function, either bounded by the function itself, or increasing to the function. Then use either monotone or DCT will show the limit equals. (We actually have that for all complete metric space, one can construct a sequence of function that increasing to the targeting function).

(9) The proof is also comparable to the proof in ordinary integration case: Let $\alpha = \left(\mathbb{E}[|X|^2|\mathcal{G}]\right)^{\frac{1}{2}}$ and $\beta = \left(\mathbb{E}[|Y|^2]|\mathcal{G}\right)^{\frac{1}{2}}$. Note

$$\frac{|XY|}{\alpha\beta} \leq \frac{|X|^2}{\alpha} + \frac{|Y|^2}{\beta}.$$

Take the condition expectation with respect to $\mathcal{G}$ on the both sides, we see

$$\begin{aligned}\mathbb{E}\left[\frac{|XY|}{\alpha\beta}\right] &\leq \left[\frac{|X|^2}{2\alpha}\right] + \left[\frac{|Y|^2}{2\beta}\right] \\ \Rightarrow \frac{\mathbb{E}[|XY|]}{\alpha\beta} &\leq \frac{1}{2\alpha}\mathbb{E}[|X|^2|\mathcal{G}] + \frac{1}{2\beta}\mathbb{E}[|Y|^2|\mathcal{G}] \leq 1\end{aligned}$$

then the desired result is obtained by multiplying two sides by $\alpha\beta$ and recall what they are.

(8) Let φ be a convex function and let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of increasing step functions that converges to X. Finally, let φ_n be defined as follows:

$$\varphi_n(x) = \begin{cases} \varphi(x) & |x| \leq n \\ \psi(x) & |x| > n \end{cases}$$

where $|\psi(x)| \leq a_n|x| + b_n$ that is tangential to φ. We see that (8) is true for φ_n and X_m, that is,

$$\varphi_n\left(\mathbb{E}[X_m|\mathcal{F}]\right) \leq \mathbb{E}[\varphi_n(X_m)|\mathcal{F}]$$

taking $m \to \infty$ has no convergence problem since φ_n continuous and at most linear growth out side of $[-n, n]$ and the integrability of X and $\varphi(X)$ are assumed. Note also $\varphi_n \uparrow \varphi$, so monotone convergence allows us to take the limit with no problemm as well. □

Remark 1.1.2. *It is worth noting that if f is a measurable function and $X \in \mathcal{G}$, then $f(X) \in \mathcal{G}$ as well. So* $\frac{1}{\mathbb{E}[X|\mathcal{G}]}, (\mathbb{E}[X|\mathcal{G}])^2 \in \mathcal{G}$.

Here we define Markov property and Markov process, but it will not be used until much later. For details on this, see ([RKS+96]), ([Law18]) and sections in ([Dur19]) and ([CZ01]).

Definition 1.1.3. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables, we say that it has Marov property or it is a Markov process if*

$$\mathbb{P}[X_{n+1} \in A|X_1, ..., X_n] = \mathbb{P}[X_{n+1}|X_n] \tag{1.1}$$

where $\mathbb{P}[X|Y]$ is interpreted as conditional expectation of the indiactor function 1_A with respect to $\sigma(Y)$.

Remark 1.1.3. *$\{X_n\}$ in the above definition an be treated as a discrete random process (i.e. nth step denotes as X_n). So for the rest of this notes, I will refer to such thing as processes.*

Remark 1.1.4. *Also note in above definition, $X_n \in \sigma(X_n) \subset \sigma(X_1, ..., X_n)$. So we can make a sequence of σ-field, say $\mathcal{F}_n$, that is increasing ($\mathcal{F}_n \subset \mathcal{F}_{n+1}$) such that $X_n \in \mathcal{F}_n$ (but not neccessarily in $\mathcal{F}_{n-1}$).*

Definition 1.1.4. *If $\{X_n\}$ is a process defined on some probability measure space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{F}_n$ be a sequence of increasing σ-field such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ and $\mathcal{F}_n \subset \mathcal{F}$ for all $n \in \mathbb{N}$. Then we say X_n is adapted to $\mathcal{F}_n$. We also refer to such $\mathcal{F}_n$ as filtrations.*

1.2 Martingales

The most important part (to me anyway) of this section is Martingale processes:

Definition 1.2.1 (Martingale)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a process defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and adapted to $\mathcal{F}_n$. Then we say $\{X_n\}_{n\in\mathbb{N}}$ is a Martinagle process (or just Martingale) if*

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_n] \quad a.s. \tag{1.2}$$

We call it sub-Martingale if $=$ is changed to $\geq$, and super-martingale if $=$ changed to $\leq$.

Remark 1.2.1. *The definition of super and sub Martingale might seem to be the opposite case, for reason of the naming, see section in ([CZ01]) about application of maringle on super and sub harmonic functions.*

One of the consequences of the above definition is the following:

Lemma 1.2.1. *Suppose $\{X_n\}$ is a martinagle w.r.t. the filtration $\mathcal{F}_n$, then*

$$\mathbb{E}[X_n|\mathcal{F}_m] = X_m \quad a.e. \text{ for all } m \leq n. \tag{1.3}$$

It can be easily seen by induction.

Given a Martingale and an increasing convex function, we can create a submartingale out those two by Jessen's inequality:

Proposition 1.2.1. *Let φ be an increasing convex function, and let $\{X_n\}$ be a submartingale w.r.t. $\mathcal{F}_n$, then the following is a sub-martingale with respect to the same filtration:*

$$\varphi(X_n). \tag{1.4}$$

This is a direct consequence of the Jessen's inequality for conditional expectation: $\mathbb{E}[\varphi(X_n)|\mathcal{F}_{n-1}] \geq \varphi(\mathbb{E}[X_n|\mathcal{F}_{n-1}]) = \varphi(X_{n-1})$.

An immediate corollary:

Corollary 1.2.1. *Let X_n be super-martingale, then $\{X_n \wedge A\}$ is a super martinagle.*

Proof. $\{-X_n\}$ defines a sub martingale, then $-X_n \wedge A$ is a submartingale, so $X_n \wedge A$ is a super martiangle. □

One notice that a submartingale is very close to be a (true) martingale. Say X_n is a sub martingale, we might want to ask if we can subtract a little something from each step X_n to make it a martingale. The answer is yes, the following object is what we need to subtract it from the sub martinagle.

Definition 1.2.2. *Let Z_n be a sequence of nonnegative random variables such that*

1. $Z_0 = 0$ *a.s. and* $Z_n \leq Z_{n+1}$ *for all* $n \in \mathbb{N}$.
2. $\mathbb{E}[Z_n] < \infty$ *for all* $n \in \mathbb{N}$.

Then we say Z_n is a sequence of increasing random variables, or just call it an increasing process.

Here is a theorem that tells us we can decompose a submartingale into sum of a martinagle and an increasing process.

Proposition 1.2.2 (Doob's Decomposition)**.** *Let X_n be a sub martingale adapted to $\mathcal{F}_n$, then there is an increasing predictable process Z_n and a martinagle Y_n, such that*

$$X_n = Y_n + Z_n \quad a.e.$$

Proof. Assume n starts at 1 and $X_0 = 0$.

We can create a martingale out of the difference of $X_n - X_{n-1}$. The following sum is obviously a martiangle:

$$Y_n = \sum_{k=1}^{n} X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]$$

since when conditioning it with $\mathcal{F}_{n-1}$, the last term of the sum disapears and the second to the last term is $\mathcal{F}_{n-1}$ measurable. Now we want to see if there is an increasing Z_n satisfies the definition. We can write

$$X_n = \sum_{k=1}^{n} X_k - X_{k-1}$$

then

$$\begin{aligned} Z_n \triangleq X_n - Y_n &= \sum_{k=1}^{n} [(X_k - X_{k-1}) - (X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}])] \\ &= \sum_{k=1}^{n} (\mathbb{E}[X_k|\mathcal{F}_{k-1}] - X_{k-1}) \geq 0 \end{aligned}$$

where the last equality is due to X_n forms a submartingale. In Z_n, each term in the sum is nonnegative, hence Z_n is increasing. □

Now we take a look at Z_n and note $\lim_{n\to\infty} \mathbb{E}[|Z_n|] = \lim_{n\to\infty} \mathbb{E}[Z_n]$ exists, possibly infinite, by monotone. Let's call the pointwise limit of this process $Z_\infty \geq Z_n$ for all n. So if Z_n is L^1 bounded in the since the limit of integral is finite, then $\mathbb{E}[Z_\infty]$ is finite, so it is uniformly integrable. An consequence of this is the following:

Proposition 1.2.3. *Suppose $\{X_n\}$ is L^1 bounded, then $\{Y_n\}, \{Z_n\}$ are also L^1 bounded. Also, if $\{Y_n\}$ is uniformly integrable, then $\{Y_m\}, \{Z_n\}$ are also uniformly integrable.*

Proof.

$$Z_n = X_n - Y_n \leq |X_n| - Y_n \Rightarrow \mathbb{E}[Z_n] \leq \mathbb{E}[|X_n|] - \mathbb{E}[Y_1]$$

This shows if X_n is L^1 bounded, then $\{Z_n\}$ is L^1 bounded as well, and it is uniformly integrable by above sentences. Then by

$$\mathbb{E}[|Y_n|] \leq \mathbb{E}[|X_n|] + \mathbb{E}[Z_n]$$

we see that if X_n is L^1 bounded (or uniformly integrable), then so is Y_n. □

1.3 Optional Time

Here I adapt Chung's definition in ([CZ01]):

Definition 1.3.1. *Let$(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ be a filtration, let τ be a random time, that is, $\tau(\omega) \in \mathbb{Z}^+$ and τ measurable. We say τ is an optional time if*

$$\{\tau = n\} \in \mathcal{F}_n, \quad \forall n \in \mathbb{N}.$$

Which is equivalent to

$$\{\tau \leq n\} \in \mathcal{F}_n, \quad \forall n \in \mathbb{N}.$$

This same thing is called *stopping time* in ([Dur19]), and the sampling theorem is called stopping theorem in him book.

Definition 1.3.2. *We denote $\mathcal{F}_\tau$ be the σ field that contains all $\Lambda \in \mathcal{F}_\infty \triangleq \bigcup_{n\in N} \mathcal{F}_n$, such that*

$$\Lambda \cap \{\tau \leq n\} \in \mathcal{F}_n.$$

Definition 1.3.3. *Let X_n be a random process and τ be a optiona time, then we define*

$$X_\tau(\omega) \triangleq X_{\tau(\omega)}(\omega)$$

where $\tau(\omega)$ is a positive integer.

Here are some properties of $\mathcal{F}_\tau$, we omit the proof since that is just busy work:

Proposition 1.3.1. *Let τ be a optional time, then we have*

1. *$\mathcal{F}_\tau$ is a σ field.*
2. *$\tau \in \mathcal{F}_\tau$.*
3. *A fixed positive number is a optional time.*
4. *If σ is an optional time, then $\sigma \wedge \tau$ is also an optional time.*
5. *If σ is another optional time, then $\mathcal{F}_{\tau\wedge\sigma} \in \mathcal{F}_\tau \bigcap \mathcal{F}_\sigma$.*
6. *$X_\tau \in \mathcal{F}_\tau$.*

I'll prove the last one since Chung says this is an excercise we should not miss

Proof. Suppose X_n are real valued and suppose B is a Borel set of the real line, then consider

$$\Xi = \{\omega \in \Omega : X_{\tau(\omega)}(\omega) \in B\} = \bigcup_n \left(\{\tau(\omega) = n\} \cap \{X_n(\omega) \in B\}\right)$$

Now we note that $\{\tau(\omega) = n\} \in \mathcal{F}_n$ by definition and $\{X_n \in B\} \in \mathcal{F}_n$ also by definition of $\mathcal{F}_n$ measurable. So

$$\Xi \cap \{\tau = n\} = \{\tau = n\} \cap \{X_n \in B\} \in \mathcal{F}_n$$

so $\sigma(X_\tau) \in \mathcal{F}_\tau$. □

Remark 1.3.1. *Suppose the index set is the positive real line, then similar proof would still work if the filtration is "continuous" in certian sense, and we just need to find a dense set of the real line.*

The next theorem tells use stopping times to create new (sub/super) martinagles:

Theorem 1.3.1 (Optional Sampling Theorem)**.** *Let X_n be a super martingale, and let $\tau, \sigma \leq m$ be two bounded stopping times such that $\sigma \leq \tau$, then $\{X_\tau; X_\sigma\}$ forms a super-martingale with respect to the filtration $\{\mathcal{F}_\tau, \mathcal{F}_\sigma\}$.*

Proof. The idea of the proof is to decompose $\Lambda \in \mathcal{F}_\tau$ into $\Lambda_n = \{\tau = n\} \cap \Lambda \in \mathcal{F}_\tau$ by definition, a collection of disjoint sets. Now look at $\Lambda_n \cap \{\sigma \geq n\} = \Lambda_n \cap \{\sigma < n\}^c \in \mathcal{F}_n$.

The theorem will be true if the following inequality holds

$$\int_{\Lambda_n} X_\tau \geq \int_{\Lambda_n} X_\sigma, \quad \forall n. \tag{1.5}$$

We want to work with non random times, namely, integers as subscripts, so we do the following: rewrite this as

$$\int_{\Lambda_n \cap \{\sigma \geq n\}} X_n - \int_{\Lambda_n \cap \{\sigma \geq m+1\}} X_{m+1} \geq \int_{\Lambda_n \cap \{n \leq \sigma \leq m\}} X_\sigma$$

This is because $\tau \leq \sigma \leq m$ and note that the second term on the left is actually zero. We decompose the right hand side into

$$\sum_{j=n}^{m} \int_{\Lambda_n \cap \{\sigma = j\}} X_\sigma = \sum_{j=n}^{m} \int_{\Lambda_n \cap \{\sigma = j\}} X_j \tag{1.6}$$

and we write the left hand side as a telescoping sum:

$$\sum_{j=n}^{m} \left(\int_{\Lambda_n \cap \{\sigma \geq j\}} X_j - \int_{\Lambda_n \cap \{\sigma \geq j+1\}} X_{j+1} \right) \tag{1.7}$$

Now we compare those two sums term by term, if each terms of (1.7) is greater than each term of (1.6), then (1.5) will be true, so let's check:

$$\begin{aligned} \int_{\Lambda_n \cap \{\sigma \geq j\}} X_j - \int_{\lambda_n \cap \{\sigma \geq j+1\}} X_{j+1} - \int_{\Lambda_n \cap \{\sigma = j\}} X_j &= \int_{\Lambda_n \cap \{\sigma \geq j\}} X_j - \int_{\Lambda_n \cap \{\sigma = j\}} X_j - \int_{\lambda_n \cap \{\sigma \geq j+1\}} X_{j+1} \\ &= \int_{\Lambda_n \cap \{\sigma \geq j+1\}} X_j - \int_{\Lambda_n \cap \{\sigma \geq j+1\}} X_{j+1} \end{aligned}$$

which is nonnegative by definiton of super martingale. So (1.5) is true. □

Remark 1.3.2. *One immediate generalization of this theorem is that if we have a sequence of bounded optional time, $\alpha_1 \leq \alpha_2 \leq ....$ then $\{X_{\alpha_n}\}_n$ will form a super martingale w.r.t $\mathcal{F}_{\alpha_n}$ when $\{X_n\}$ is a super martingale.*

Remark 1.3.3. *Note that if we change super to sub or simply just martinagle, then theorem is still true.*

Now we want to see if we still have similar result when τ and σ are unbounded optional times. In this case, we will need X_∞ to be defined, otherwise $X_\tau 1_{\tau=\infty}$ does not make sense.

Definition 1.3.4. *Let $\{X_n\}_{n\in\mathbb{N}}$ be martinagle, then we say it has an last element when there is some random variable, call X_∞ such that $\mathbb{E}[X_\infty|\mathcal{F}_n] = X_n$ and $X_\infty \in \bigcup_{n\in\mathbb{N}} \mathcal{F}_n$. We denote such process as $\{X_n, n \in \mathbb{N}_\infty\}$, where $\mathbb{N}_\infty$ denote $\mathbb{N} \cup \{\infty\}$.*

Similar definitions for super and sub martinagles with last element: $\mathbb{E}[X_\infty|\mathcal{F}_n] \geq X_n$ and $\mathbb{E}[X_\infty|\mathcal{F}_n] \leq X_n$.

In the case of a martingale, it can be seen as the whole process is generated by a single integrable random variable with a set of filtration. We first state the optional sampling theorem for this simple case for unbounded optional times:

Theorem 1.3.2. *Let let $Y \in L^1$ and let $\{X_n\}$ be such that $X_n = \mathbb{E}[Y|\mathcal{F}_n]$ where $\{\mathcal{F}_n\}_n$ is a filtration. Then for any two optional time $\tau \leq \sigma$, we have first*

$$X_\alpha = \mathbb{E}[Y|\mathcal{F}_\alpha]$$
$$\mathbb{E}[X_\sigma|\mathcal{F}_\tau] = X_\tau.$$

Proof. Let $\Lambda \in \mathcal{F}_\tau$, as before, we can decompose it into a collection of disjoint sets: $\Lambda_n = \{\tau = n\} \cap \Lambda \in \mathcal{F}_n$. First we show integrability:

$$|X_n| \leq \mathbb{E}[|Y||\mathcal{F}_n],$$

so

$$\mathbb{E}[|X_\tau|] = \sum_{n=1}^{\infty} \int_{\tau=n} |X_\tau| = \sum_{n=1}^{\infty} \int_{\tau=n} |X_n| \leq \sum_{n=1}^{\infty} |Y| = \mathbb{E}[|Y|].$$

Now we show the first identity:

$$\int_\Lambda X_\tau = \sum_{n=1}^{\infty} \int_{\Lambda_n} X_\tau = \sum_{n=1}^{\infty} \int_{\Lambda_n} X_n = \sum_{n=1}^{\infty} \int_{\Lambda_n} Y = \int_\Lambda Y.$$

For the second identity, recall the tower property:

$$X_\tau = \mathbb{E}[Y|\mathcal{F}_\tau] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_\sigma]|\mathcal{F}_\tau] = \mathbb{E}[X_\sigma|\mathcal{F}_\tau]$$

□

Recall our task it to generalize the optional sampling theorem for general super(sub) martingales with last element. We observe that we can decompose such super martinagle in the following self explainatory way

Lemma 1.3.1. *Let $\{X_n, \mathcal{F}_n, n \in \mathbb{N}_\infty\}$ be a super martingale with last element, then it can be decomposed into sum of super martinagle with last element being zero and a martingale. Namely,*

$$X_n = Y_n + X'_n$$

where

$$X'_n = \mathbb{E}[X_\infty|\mathcal{F}_n] \quad Y_n = X_n - X'_n \geq 0.$$

Now if we want to show sampling theorem for a super martiangle with an last element, then only thing we have to do is to prove this for super martingale with last element being zero:

Proposition 1.3.2. *Let $\{X_n, \mathcal{F}_n, n \in \mathbb{N}_\infty\}$ be a super martingale with $X_\infty = 0$. Then for any $\tau \leq \sigma$ optional times, we have*

$$\mathbb{E}[X_\sigma | \mathcal{F}_\tau] = X_\tau.$$

Proof. Here, the first part of this proof follows the proof for Theorem 14: Let $\Lambda \in \mathcal{F}_\tau$ and decompose it as $\Lambda_n = \Lambda \cap \{\tau = n\}$ and $\Lambda_\infty = \Lambda \cap \{\tau = \infty\}$. Let $j \geq n \geq 1$ and by definition of super-martinagle, we see that

$$\begin{aligned} &\int_{\Lambda_n \cap \{\sigma \geq j+1\}} X_j \geq \int_{\Lambda_n \cap \{\sigma \geq j+1\}} X_{j+1} \\ \Rightarrow &\int_{\Lambda_n \cap \{\sigma \geq j\}} X_j - \int_{\Lambda_n \cap \{\sigma \geq j+1\}} X_{j+1} \geq \int_{\Lambda_n \cap \{\sigma = j\}} X_j \end{aligned}$$

since $\Lambda_n \in \mathcal{F}_n \subset \mathcal{F}_j$, $\{\sigma \geq j+1\} = \{\sigma \leq j\}^c \in \mathcal{F}_j$. SUmming from $j = n$ to m on both sides gives us

$$\int_{\Lambda_n \cap \{\sigma \geq n\}} X_n - \int_{\Lambda_n \cap \{\sigma \geq m+1\}} X_{m+1} \geq \int_{\Lambda_n \cap \{n \leq \sigma \leq m\}} X_\sigma$$

and since we assumed non-negative, we can drop the minus term on the left hand side, and since $\sigma \geq \tau$, we may also drop $\sigma \geq n$ part on both side of the integral, change X_n to X_τ to get

$$\int_{\Lambda_n} X_\tau \geq \int_{\Lambda_n \cap \{\sigma \leq m\}} X_\sigma$$

Note that the right hand side is positive and non-decreasing with respect to m, so we may take $m \to \infty$ and take the sum over $j \in \mathbb{N}$ to obtain

$$\begin{aligned} &\sum_{j \in \mathbb{N}} \int_{\Lambda_n} X_\tau \geq \sum_{j \in \mathbb{N}} \int_{\Lambda_n \cap \{\triangleleft < \infty\}} X_\sigma \\ \Rightarrow &\int_{\Lambda \cap \{\tau < \infty\}} X_\tau \geq \int_{\{\tau < \infty\} \cap \Lambda_n} X_\sigma \\ \Rightarrow &\int_{\Lambda \cap \{\tau < \infty\}} X_\tau + \int_{\Lambda \cap \{\tau = \infty\}} X_\tau \geq \int_{\{\tau < \infty\} \cap \Lambda} X_\sigma + \int_{\Lambda \cap \{\tau = \infty\}} X_\sigma \end{aligned}$$

the last inequality is true becasue on the set $\sigma = \infty$ we have $X_\sigma = 0$ by assumption, and $\tau \leq \sigma$. Combine the sum we have the desired result. □

Now combining the discussion and the theorems above, we have the following theorem which we call Optional Sampling Theorem:

Theorem 1.3.3 (The Optional Sampling Theorem). *If $\{X_n, \mathcal{F}_n, n \in \mathbb{N}_\infty\}$ is a super martiangle with last element and $\tau \leq \sigma$ are two arbitrary optional times, then we have*

$$X_\tau = \mathbb{E}[X_\sigma | \mathcal{F}_\tau].$$

1.4 Some Convergence Results and Inequalities

For a given, say, submartingale, it would be good to know when it converges, either in probability, a.e, distribution or L^p. Let's first do a.e.

Now here is a wierd way to see when it converges: we want $\overline{\lim} X_n = \underline{\lim}_{n\to\infty} X_n$ a.e. We recall the definition of lim sup and lim inf: $\overline{\lim}_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{m\geq n} x_m$ and $\underline{\lim}_{n\to\infty} x_n = \lim_{n\to\infty} \inf_{m\geq n} x_m$, call them α and β respectively. Note further that the sup is decreasing when taking limit and the inf is increasing when taking limit (might not be strictly increasing or decreasinng). So if it happens that x_n does not converge, then $\alpha < \beta$, but this means that $\{x_n\}$ as a process would oscillate between α and β for infinitely many times. More specifically, it would oscillate from at or above β, say some point $\alpha' < \beta$ to some point at or below β, say $\beta' > \alpha'$ infinitely many times, and this is an if and only if "statement". If this happens, then the process will go from α' up to β' infinitely many times. So if there is some condition such that the process X_n oscillate finitely many times with probability one for any α, β, then we'd show lim sup equals to lim inf and we'd show convergence. Even better, we only need to show it for a dense set, say $\alpha, \beta \in \mathbb{Q}$. So we will proceed according to this strange idea.

To formalize this idea, we need a few things first:

Definition 1.4.1. *We say a process $\{H_n\}_{n\in\mathbb{N}}$ is predictable with respect to $\mathcal{F}_n$ if*

$$H_n \in \mathcal{F}_{n-1} \quad \forall n \in \mathbb{N}.$$

And here is a definition we will never use again (I think) after this:

Definition 1.4.2. *Let $\{X_n\}$ be an adapted process and H_n be a predictable process, then we define the Martinagle transform $(H \cdot X)_n$ as*

$$(H \cdot X)_n = \sum_{k=1}^{n} H_n(X_n - X_{n-1}).$$

Here is why it has this name:

Proposition 1.4.1. *Let $\{X_n\}$ be a super martingale, then $\{(H \cdot X)_n\}$ is still a super martingale assuming $H_n \geq 0$ are predictable and bounded.*

Proof. This is straight forward:

$$\begin{aligned}
\mathbb{E}\left[(H \cdot X)_n \,|\mathcal{F}_{n-1}\right] &= \sum_{1\leq k\leq n} \mathbb{E}[H_k(X_k - X_{k-1})|\mathcal{F}_{n-1}] \\
&= \sum_{1\leq k\leq n-1} H_n(X_k - X_{k-1}) + H_n\left(\mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}\right) \\
&= (H \cdot X)_{n-1} + H_n\left(\mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}\right) \geq (H \cdot X)_{n-1}.
\end{aligned}$$

Where we used X_n is super martinagle, and $H_n \in \mathcal{F}_n$. □

Remark 1.4.1. *It works for super-martingale, then it works for sub-martinagle.*

After this, we want to formalize the concept of the process go from α to β for $\alpha < \beta$. So we define the following thing inductively:

Let $K_0 = -1$ and define K_n for $n \geq 1$ to be

$$\begin{aligned}
K_{2n-1} &\triangleq \min_k\{N_{2n-2} < k \text{ and } X_k \leq \alpha\} \\
K_{2n} &\triangleq \min_k\{N_{2n-1} < k \text{ and } X_k \geq \beta\}.
\end{aligned}$$

for some any $\alpha < \beta$. Now it would be easy to verify that N_k's are optional times. Now we want to have something to represent how many time the process X_n goes from α to β, here is one: let H_m be defined as follows:

$$H_n = \begin{cases} 1 & N_{2k-1} < n \leq N_{2k} \\ 0 & \text{else.} \end{cases}$$

By the name, we note that H_n is predictable: $\{N_{2k-1} < n \leq N_{2k}\} = \{N_{2k-1} \leq n-1\} \cap \{N_{2k} \leq n-1\}^c \in \mathcal{F}_{n-1}$. So we can use martinagle transform to make a submartingale out of a submartingale. Finally, we want to count how many times the "upcross" happened from step one to step n, so we define $U_{\alpha,\beta,n} = \max_k\{N_{2k} \leq n\}$, and we see that $U_{\alpha,\beta,n}$ does this job.

Then, our next task to to get some sort of bound for $U_{\alpha,\beta,n}$ such that it gets us $U_{\alpha,\beta,n} < \infty$ a.e. Here is a theorem that does that

Theorem 1.4.1 (Up-Crossing Inequality)**.** *Let $X_n, U_{\alpha,\beta,n}$ be as above, then we have*

$$(\beta - \alpha)\mathbb{E}[U_{\alpha,\beta,n}] \leq \mathbb{E}[(X_n - \alpha)^+] - \mathbb{E}[(X_0 - \alpha)^+].$$

Remark 1.4.2. *From this theorem, it is clear that one of the posible conditions to make the number of crossing finite a.e. is* $\sup_{n\in\mathbb{N}} \mathbb{E}[X_n^+] < \infty$.

Proof. Let $Y_n = (X_n - \alpha)^+ + \alpha$, then the path of Y_n would cross the interval $[\alpha, \beta]$ the same number of times as X_n (Y_n would stay at α if X_n is less than α, and the rest of the behaviors of the two processes are the same). We note that $\sum_{k=1}^n K_n(Y_n - Y_{n-1}$ record the part of the trajectories of Y_n that crosses $[\alpha, \beta]$ from below to above, but with a bit longer since it also record the distance from β to a little above β. Let $K_n = 1 - H_n$ so K_n is predictable and nonnegative, so it turns a submartinagle into a submartinagle via Maringale transform as well. Then we note that

$$\begin{aligned} Y_n - Y_0 &= \sum_{k=1}^n Y_n - Y_{n-1} \\ &= \sum_{k=1}^n (K_n + H_n)(Y_n - Y_{n-1}) \\ &= \sum_{k=1}^n K_n(Y_n - Y_{n-1}) + \sum_{k=1}^n H_n(Y_n - Y_{n-1}) \\ &= (K \cdot Y)_n + (H \cdot Y)_n. \end{aligned}$$

Then as indicated before, both $H \cdot Y$ and $K \cdot Y$ forms a martinagle, and they both start as nonnegative random variables, so we when take the expectation, we get the following Inequality

$$\mathbb{E}[Y_n] - \mathbb{E}[Y_0] \geq \mathbb{E}\left[(K \cdot Y)_n\right] \geq \mathbb{E}[(\beta - \alpha)U_{\alpha,\beta,n}].$$

□

Now we are ready to prove a pointwise convergence theorem:

Theorem 1.4.2 (Submartinagle Convergence Theorem)**.** *Let X_n be a submartinagle w.r.t $\mathcal{F}_n$. Suppose* $\sup_{n\in\mathbb{N}} \mathbb{E}[X_n^+] < \infty$, *then X_n converges pointwise to a random varibale X that is integrable.*

Remark 1.4.3. *Note that the theorem says two things: $X_n \to X$ pointwise a.e., and $\mathbb{E}|X| < \infty$. It does not mean $X_n \to X$ in L^1.*

The proof of this theorem is just a formalization of the intuition we had in the beginning.

Proof. By Up-Crossing inequality, we see that

$$\lim_{n\to\infty} \mathbb{E}[U_{\alpha,\beta,n}] < \sup_{n\in\mathbb{N}} \mathbb{E}[X^+] + |a| < \infty$$

which implies $\mathbb{P}\left(U_{\alpha,\beta,\infty}\right) = 0$, which is true for all $\alpha, \beta \in \mathbb{Q}$. Take union of those countable sets to get

$$\mathbb{P}\left[\bigcup_{\alpha,\beta\in\mathbb{Q}} \{U_{\alpha,\beta,\infty} < \infty\}\right] = 1$$

and observe (or recall the description before)

$$\bigcup_{\alpha,\beta\in\mathbb{Q}} \{U_{\alpha,\beta,\infty} < \infty\} \supset \left\{\overline{\lim}_{n\to\infty} X_n > \underline{\lim}_{n\to\infty} X_n\right\}$$

so we have desired convergence. Now the task is to show $\mathbb{E}[|X|] < \infty$. That is a result of Fatou's lemma:

$$\liminf_{n\to\infty} \mathbb{E}[X_n^+] \geq \mathbb{E}[X^+]$$

Also, by the nature of submartingale, we see

$$\mathbb{E}[X_n^-] = \mathbb{E}[X_n^+] - \mathbb{E}[X_n] \leq \mathbb{E}[X_n^+] - \mathbb{E}[X_0]$$

So apply Fatou's lemma again we see $\mathbb{E}[X^-] \leq \liminf_{n\to\infty} \mathbb{E}[X_n^+] - \mathbb{E}[X_0] \leq \sup_n \mathbb{E}[X^+] - \mathbb{E}[X_0] < \infty$. Hence the theorem is proved. □

Here is a trivial consequence of submartingale convergence theorem:

Corollary 1.4.1. *Let X_n be an negative sub martinagle, or positive super martinagle, then it converges a.e. to an integrable random variable X.*

Before we go any furhter, let's revisit optional time for a short while. Let τ be a optiona time, that is, $\{\tau \leq n\} \in \mathcal{N}$. Note $\{\tau \geq n\} = \{\tau < n\}^c = \{\tau \leq n-1\}^c \in \mathcal{F}_{n-1}$. So recall Proposition 38 and let $H_n = 1_{\tau \geq n}$, then if $\{X_n\}$ is a martinagle with respect to $\{\mathcal{F}_n\}_n$, then

$$(H \cdot X)_n = \sum_{k=1}^{n} 1_{\tau\geq k}\left(X_k - X_{k-1}\right) = X_{n\wedge\tau} - X_0$$

is also a martinagle, so we conclude:

Theorem 1.4.3. *If $\{X_n\}$ is a martinagle, and τ is a stopping time, then $\{X_{n\wedge\tau}\}$ is also a martingale.*

We proved (basically) the same theorem in previous section with a more straight forward method in the sense of we did not use additional tools.

We also proved the following theorem in the previous section, but there is also an way to prove it via Maringale transform:

Theorem 1.4.4. *X_n be a submartingale and let $\tau \leq \sigma$ be two stopping times bounded by $m \geq 1$, then*

$$\mathbb{E}[X_\tau] \leq \mathbb{E}[X_\sigma].$$

Proof. The proof to previous theorem gives us a hint that we need to use maringale transform for some predictable $H_n = 1.$ for which $(H \cdot X)$ gives a submartingale. Since this transform gives us some telescoping sum, we want the first nonzero term to contain $-X_\tau$ and the last nonzero term to contain X_σ. This indicates that we should look lat

$$H_k = 1_{\tau < k < \sigma}.$$

Note $\{\tau < k\} = \{\tau \le k-1\} \in \mathcal{F}_{k-1}$ and $\{k \le \sigma\} = \{\sigma < k\}^c = \{\sigma \le k-1\}^c \in \mathcal{F}_{k-1}$, so H_k is predictable. Now consider the maringale transform:

$$(H \cdot X)_n = \sum_{k=1}^{n} 1_{\tau < k \le \sigma} (X_k - X_{k-1}) = X_{\sigma \wedge n} - X_{\tau \wedge n}.$$

W Which is a submartingale by previous theorem. Now we take $n = m+1$ to get

$$\mathbb{E}[X_\sigma] - \mathbb{E}[X_\tau] = \mathbb{E}[X_{\sigma \wedge (m+1)} - X_{\tau \wedge (m+1)} \ge \mathbb{E}[X_{\sigma \wedge 0} - X_{\tau \wedge 0}] = 0$$

which proves our statement. □

The following two inequalities are useful (important):

Proposition 1.4.2 (Doob's Maximal Inequality)**.** *Let $\{X_n\}$ be a submartingale, and let $\lambda > 0$, then*

$$\lambda \mathbb{P}[\max_{1 \le m \le n} X_n^+ \ge \lambda] \le \mathbb{E} X_n 1_A \le \mathbb{E}[X_n^+].$$

Proof. Let $\tau = \min_m \{X_n^+ \ge \lambda\}$ which is a bounded stopping time, and denote $A = \{\max_{1 \le m \le n} X_n^+ \ge \lambda\}$

$$\lambda \mathbb{P}[A] = \int_A \lambda d\mathbb{P} \le \int_A X_{\tau \wedge n} d\mathbb{P} \le \mathbb{E}[X_n] \le \mathbb{E}[X_n^+]$$

where the last two inequalities are due to Theorem 20 and $X_n = X_n^+ - X_n^-$. □

Remark 1.4.4. *Note that if we change* $\max X_n^+$ *to* $\max X_n$*, we get the same result since $\lambda > 0$.*

Remark 1.4.5. *Compare this to the Kolmogorov's Maximal Inequality, we see K-Max inequality is a special case of Doob's Maximal Inequality.*

Remark 1.4.6. *We can easily extend the the above inequality to the case of countably infinite index*

$$\lambda \mathbb{P}[\sup_{n \ge 1} X_n^+ \ge \lambda] \le \sup_{n \ge 1} \mathbb{E}[X_n^+]$$

Other forms of Doob's Maximal Inequality:

Proposition 1.4.3. *Let $\{X_n\}$ be a super-martingale and let $\lambda > 0$, then*

$$\lambda \mathbb{P}\left[\max_{1 \le m \le n} |X_n|\right] \le \mathbb{E}[X_0] + 2\mathbb{E}[|X_n|].$$

Proof. The moral of the proof is the same. Let $\tau = n \wedge \max_{1 \le m \le n} \{|X_n| \ge \lambda\}$, so τ is a stopping time. Denote $A = \{\max_{1 \le m \le n} |X_n| \ge \lambda\}$.

$$\mathbb{P}[A] = \int_A 1 \le \int_A \frac{|X_\tau|}{\lambda} \le \frac{1}{\lambda} \mathbb{E}[|X_\tau|].$$

Decompose $\mathbb{E}[X_\tau]$ into positive and negative parts

$$\mathbb{E}[|x_\tau|] = \mathbb{E}[X_\tau^+] - \mathbb{E}[X_\tau^-] = \mathbb{E}[X_\tau] - 2\mathbb{E}[X_\tau^-] \le \mathbb{E}[|X_0|] + 2\mathbb{E}[X_\tau^-].$$

Now we note $-X_n$ is sub-martinagle, and $X_n^- = (-X_n)^+$ which is also a submartinagle. So we can apply optional sampling theorem for bounded stopping time for X^-:

$$\mathbb{E}[|X_0|] + 2\mathbb{E}[X_\tau^-] \le \mathbb{E}[|X_0|] + 2\mathbb{E}[X_n^-] \le \mathbb{E}[X_0] + 2\mathbb{E}[|X_n|].$$

Plug those all back into the first inequality we get desired result. □

Uniformly Integrable (s)Martingales

Uniformly integrability provides nice convergence results for submartingales. Recall the definition and a convergence theorem from the first chapter:

Definition 1.4.3 (Uniformly integrability)**.** *Suppose $\{X_t\}_{t\in T}$ is a family of random variables where T is an index set, we say the family is uniformly integrable if*

$$\lim_{M\to\infty}\sup_{t\in T}\mathbb{E}\left(|X_t|1_{|X_t|\geq M}\right)=0.$$

Proposition 1.4.4. *Let $X_n \to X$ in probability, then the followings are equivalent:*

1. *$\{|X_t|^r\}$ is uniformly integrable.*
2. *$X_t \to X$ in L^r.*
3. *$\mathbb{E}[|X_t|^r] \to \mathbb{E}|X|^r < \infty$.*

Here is the main theorem of this uniformly integrable martingales:

Theorem 1.4.5. *Let $\{X_n\}$ be submartingales, then the following are the same:*

1. *$\{X_n\}$ is uniformly integrable.*
2. *$X_n \to X_\infty$ in L^1.*
3. *$X_n \to X_\infty$ in L^1 and almost surely.*

Proof. (1) $\Rightarrow \sup_{n\in\mathbb{N}}\mathbb{E}[|X_n|] \Rightarrow X_n \to X_\infty$ a.e. by maringle convergence theorem $\Rightarrow X_n \to_p X_\infty \Rightarrow X_n \to X_\infty$ in L^1.

Now suppose (2) is true, then $\{X_n\}$ must be bounded in L^1, then by mtg convergence theorem again it converges to X_∞ a.e. so in probability, by Theorem 41 again it is uniformly integrable. So (1) $\Leftarrow$ (2) $\Rightarrow$ (3). and (3) $\Rightarrow$ (2) is obvious. □

The above theorem is for submartingales, but if we have a true Martinagle, we can improve the above theorem.

Lemma 1.4.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a martingale with filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, and suppose $X_n \to_{L^1} X$ for some X, then we have*

$$X_n = \mathbb{E}[X|\mathcal{F}_n], \quad \textit{a.s. for al } n \in \mathbb{N}.$$

Proof. Note here besides L^1 convergence we also have a.e. convergence. We need to show that for all $A \in \mathcal{F}_n$ we have

$$\int_A X_n = \int_A \lim X_n$$

From the definition of Martingales, we also have the following expression for the left hand sdie:

$$\int_A X_n = \int_A \mathbb{E}[X_m|\mathcal{F}_n] = \int_A X_m \quad \forall n \in \mathbb{N}. \tag{1.8}$$

Take the limit on the both sides of (1.8) on m to see $\int_A X_n = \lim_{m\to\infty}\int_A X_m$, so we need to show

$$\lim_{n\to\infty}\int_A X_n = \int_A \lim_{n\to\infty} X_n$$

but this comes directly from the definition. □

Now, realizing the fact that a martingale is also a submartingale, combined with Lemma 9, we have the following improved theorem for martingales:

Theorem 1.4.6. *Let $\{X_n\}$ be a martingale, then the following are equivalent:*

1. *$\{X_n\}$ is uniformly integrable.*
2. *$X_n \to X_\infty$ in L^1.*
3. *$X_n \to X_\infty$ in L^1 and almost surely.*
4. *There is X_∞ integrable such that $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ a.e.*

Proof. Only thing we need to prove here is (4) implies (1) and it is straight forward

$$\mathbb{E}\left(|X_n|1_{|X_n|\geq M}\right) \leq \int_{|X_n|\geq M} |X|_\infty d\mathbb{P}.$$

However, by Chebyshev Inequality we see that

$$\mathbb{P}[|X_n| \geq M] \leq \frac{1}{M}\int |\mathbb{E}[X_\infty|\mathcal{F}_n]| \leq \frac{1}{M}\int \mathbb{E}|X_\infty||\mathcal{F}_n = \frac{1}{M}\mathbb{E}[|X_\infty|]$$

which goes to zero as $M \to \infty$, so DCT will give us uniformly integrability result. □

Now we state the converse of Lemma 9:

Lemma 1.4.2. *Let X be an integrable random variable and let $\{\mathcal{F}_n\}_n$ be a filtration such that $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, where $\mathcal{F}_\infty = \sigma(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n)$. Then we have*

$$X_n \triangleq \mathbb{E}[X|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{F}_\infty] \quad \text{a.e. and in } L^1.$$

Proof. This is a simple observation that $\{X_n\}_{n\in\mathbb{N}}$ defines a martinagle w.r.t the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ with last element being $\mathbb{E}[X|\mathcal{F}_\infty]$ (the tower property of conditional expectation). Then apply previous theorem. □

Here is an immediate consequenceL

Theorem 1.4.7 (Levy's 0-1 Law)**.** *Let $\mathcal{F}_n \uparrow \mathcal{F}_\infty$. Suppose $A \in \mathcal{F}_\infty$, then*

$$\mathbb{E}[1_A \mathcal{F}_n] \to 1_A \quad \text{a.e.}$$

Then we can derive Kolmogorov's zero one law fairly easily:

Theorem 1.4.8 (Kolmogorov's 0-1 Law)**.** *Let $\{X_n\}$ be a seqeunce of independent random variables. Let $\mathcal{T} = \cap_{n\in\mathbb{N}}\sigma(X_n, X_{n+1}, ...)$. Then if $A \in \mathcal{T}$, then $\mathbb{P}[A] \in \{0,1\}$.*

Proof. Let $\mathcal{F}_n = \sigma(X_1, ..., X_n)$, and let $\mathcal{F}_\infty = \sigma(X_1, X_2, ...)$, then $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, and $A \in \mathcal{F}_\infty$. So by Levy's zero one law, $\mathbb{E}[1_A|\mathcal{F}_n] \to 1_A$. However, $A \perp \mathcal{F}_n$ for all n, so $\mathbb{E}[1_A|\mathcal{F}_n] = \mathbb{P}[A]$. This means $\mathbb{P}[A] = 1_A$, so it is either one or zero. □

The last thing I want to write about is the Doob's L^p inequality in its general form for discrete sub martinagle that will be useful in continuous martingale case.

Theorem 1.4.9. *[L^p Inequality] Let $1 < p < \infty$ and q be its Holder conjugate, that is, $\frac{1}{q} + \frac{1}{p} = 1$. Suppose that $\{X_n\}$ be a positive submartingale such that*

$$\sup_{n \geq 1} \mathbb{E}[|X_n|^p] < \infty \tag{1.9}$$

Then $\sup_{n \in \mathbb{N}} X_n \in L^p$ and

$$\|\sup_{n \in \mathbb{N}} X_n\|_p \leq q \sup_{n \in \mathbb{N}} \|X_n\|_p$$

,

Proof. By Jessen's inequality, condition 1.9 implies $\sup_n \mathbb{E}[|X|] < \infty$, then by Mtg convergence theorem it converges a.e. to some X and by Proposition 41, it is uniformly integrable and $X_n \to X$ in L^p and $X \in L^p$.

Now we consider the extension of Doob's maximal inequality (by continuity of measure and monotone) we have

$$\lambda \mathbb{P}[\sup_n X_n \geq \lambda] \leq \int_{\sup_n X_n \geq \lambda} X_\infty.$$

Now let $Y = \sup_n X_n$ and let's pretend for a moment that $Y \in L^p$, then we can calculate the moments of Y via its tail probability:

$$\begin{aligned}
\mathbb{E}[Y^p] &= \int_0^\infty pt^{p-1} \mathbb{P}[Y \geq t] dt \\
&\leq \int_0^\infty pt^{p-1} \int_{Y \geq t} X_\infty d\mathbb{P} dt \\
&= \int_0^\infty pt^{p-1} \int 1_{Y \geq t} X_\infty d\mathbb{P} dt \\
&= \int_\Omega X_\infty \int_0^Y \frac{1}{\lambda} pt^{p-1} dt d\mathbb{P} \\
&= q\mathbb{E}[X_\infty Y^{p-1}]
\end{aligned}$$

Now we use Holder's inequality with exponent p on X_∞ and $q = \frac{p}{p-1}$ on Y^{p-1} to get $\mathbb{E}[Y^p] \leq q\|X_\infty\|_p \|Y\|_p^{p-1}$ which implies $\|Y\|_p \leq q\|X_\infty\|_p$. So we need to check the integrability of Y.

In this case, replace Y by $Y \wedge M$ for positive M in above calculate and get the same bound for $Y \wedge M$ in the end, and we see that the bound for $\|Y \wedge M\|_p$ does not depend on M, so take the limit inf and use Fatou's lemma we should see the result. □

Remark 1.4.7.

Chapter 2

Stochastic Integration

The structure here is a combination of ([KS12]) and ([LG16]). We will first generalize the the concept of maringle from discrete to continuous time, and talk about the analogies of Doob's inequalities, optional sampling and uniform integrability in continuous setting (by the way, they are exact same result from density argument, at least in the case of continous martingales).

Then, we will talk about gaussian process, and then dive into existance of Brownian motion, which will be shown by existance of white noise, a more general wiener process along with Kolmogorov's continuity theorem. Then we investigate pathwise properties of Brownian motion including reflection principle, zero-one law, distributional properties etc.

Then we will talk about construction of stochastic integral and their basic properties, as well as the change of variable formula which is called Ito's lemma due to Ito.

2.1 Continous time Martingales

The Setup

Again, we need to establish the concept of filtration in continuous setting:

Definition 2.1.1 (Filtration)**.** *Let $\{\mathcal{F}_t\}_{\mathbb{R}^+}$ be an increasing family of σ-fields indexed by the positive real line. increasing means*

$$\mathcal{F}_s \subset \mathcal{F}_t \forall s \leq t,\ s, t \in \mathbb{R}^+.$$

We call this family of σ-field a filtration and $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_n, \mathbb{P})$ a filtered probability space.

In principle, we want to make our fintration as nice as possible for a given random process index by the positive real line, so we define

$$\mathcal{F}_{t+} = \bigcap_{\sigma > t} \mathcal{F}_\sigma.$$

If we have a filtration like this, it is called right continuous. Futher more, let $\mathcal{F}_0$ contians all subsets of the null sets of the underlying probability measure. This can be done as follows: let $\mathcal{N}$ be the collection of sets that contains all the null sets of $\mathbb{P}$ and let $\mathcal{F}_0' = \sigma(\mathcal{F}_0 \cup \mathcal{N})$ and define $\mathcal{F}_t' = \bigcap_{s>t} \sigma(\mathcal{F}_s \cup \mathcal{N})$, then we'd have a filtration that is right continous and contains all the null sets of the probability measure, and we call these two conditions the usual condition for filtrations. Note that contains all subsets of null sets is stronger than completion of a σ algebra.

Now we want measurabilities of a continous stochastic process:

Definition 2.1.2. *Let $\{X_t\}_{t\in\mathbb{R}^+}$ be a continous time stochastic process, we say it is measurable if*

$$X_\cdot(\cdot) : (\omega, t) \to X_t(\omega)$$

is measurable with respect to $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+)$.

But here is the measurability we would want to work with:

Definition 2.1.3. *Let $\{X_t\}_{t\in\mathbb{R}^+}$ be a continous time stochastic process, we say it is progressively measurable if*

$$X_\cdot(\cdot) : (\omega, t) \to X_t(\omega)$$

is measurable with respect to $\mathcal{F}_t \otimes \mathcal{B}([0,t])$.

Here is a theorem that takes care of measurabilities for us (without proof):

Proposition 2.1.1. *If $(X_t)_t$ is an adapted right/left continous process, then above two measurabilities will both be satisfied.*

Now we need the concept of stopping time, which is the "same" as optional time for discrete process, but optional time becomes something else in continuous time setting:

Definition 2.1.4. *Let $\tau : \Omega \to [0, \infty]$ be a random time, we say it is a stopping time to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ if*

$$\{\tau \leq t\} \in \mathcal{F}_t \quad \forall t \geq 0,$$

and the σ-field of the past before τ is defined to be

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty, A \cap \{\tau \leq t\}\} \in \mathcal{F}_t.$$

and τ is saied to be optional time if

$$\{\tau < t\} \in \mathcal{F}_t \quad \forall t \geq 0$$

and

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau < t\} \in \mathcal{F}_t\}$$

Now let's get rid of the concept of optional time:

Proposition 2.1.2. *Let τ be an optional time of $\mathcal{F}_t$, if $\mathcal{F}_t$ is right continous, then τ is also a stopping time.*

Proof.

$$\{\tau \leq t\} = \bigcap_{s>\tau} \{\tau < s\}$$

and we note that $\tau \in \mathcal{F}_\tau$, so $\{\tau < t\} \cap \{\tau < s\} \in \mathcal{F}_{s\wedge t}$. So above set is also in $\mathcal{F}_t$ by right continuous. □

From now on, we will always assume our filtration is right continous and contains all the subsets of null sets. In this case, we don't have to work with optional time.

Here are some facts of stopping times and the σ algebra associated with the stopping time from ([LG16]), the proof is mostly set algebra manipulations, so we omit.

Proposition 2.1.3. *Let T, S be stopping time w.r.t $\mathcal{F}_t$, and suppose $S \le T$, then*

1. *$\mathcal{F}_T \subset \mathcal{F}_{T+}$, and if filtration is right continous (which will always be the case), then they are equal.*
2. *$T = t$ for a constant t is a stopping time and $\mathcal{F}_T = \mathcal{F}_t$.*
3. *T is $\mathcal{F}_T$ measurable.*
4. *Let $A \in \mathcal{F}_\infty$ and let*

$$T^A(\omega) = \begin{cases} T(\omega) & \text{if } \omega \in A \\ \infty & \omega \notin A \end{cases}$$

   Then $A \in \mathcal{F}_T$ iff T^A is a stopping time.

5. *$S \wedge T, S \vee T$ are stopping times, and $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$.*
6. *If $(S_n)_n$ is a monotone increasing/decreasing sequence of stopping times, then $S = \lim \downarrow S_n$ or $S = \lim \downarrow S_n$ are stopping times.*

Here is a way to verify if some random time is a stopping time and how to construct stopping time take only countably many values that decreases to arbitrary stopping time:

Proposition 2.1.4. *Let T be a stopping time and $S \in \mathcal{F}$ be a random variable such that $S \ge T$. Then S is a stopping time.*

Also, the following is a sequence of stopping times that only take countably many values:

$$T_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} 1_{\{k^{-n} < T \le (k+1)2^{-n-1}\}} + \infty 1_{\{T=\infty\}}.$$

Second part is obvious, let's show first part

Proof. By definition we have

$$\{S \le t\} \cap \{T \le t\} \in \mathcal{F}_T$$

but if $S \le t$, then $T \le t$, so

$$\{T \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_T.$$

□

Here are two useful examples of stopping time

Proposition 2.1.5. *Let $\{X_t\}$ be a stochastic process takes value in a complete metric space that is adapted to $\{\mathcal{F}_t\}_t$, let O be an open set and F be an closed set. Then the followings are true*

1. *if X_t is right continous, then the following is a stopping time*

$$T = \inf_t \{X_t \in O\};$$

2. *if X_t is continous, then the following is a stopping time*

$$T = \inf_t \{X_t \in F\}.$$

The idea of the proof is to use right continuity and continuity to make uncoutable unions into countable unions and still gives the same infomation on sets like $\{T \le t\}$.

Martingale

The defintion of martinagle in the continuous setting is exactly like the definition in the discrete setting, just change $\mathbb{N}$ to $\mathbb{R}^+$. So we do not repeat here.

The following fact could be usefulThe following facts might be useful:

Lemma 2.1.1. *Submartinagles with constant expectation is a martinagle.*

Proof. Let $A \in \mathcal{F}_s$ where $s \leq t$, consider

$$\int_\Omega X_t - \int_\Omega X_s = \int_\Omega |\mathbb{E}[X_t|\mathcal{F}_s] - X_s| = 0$$

so $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ a.e. □

Here is an important property:

Definition 2.1.5 (Independent Increments)**.** *Let X_t be a process that is adapted to $\mathcal{F}_t$ (filtration), we say X_t has independent increments if for all $s \leq t$ we have*

$$X_t - X_s \perp \mathcal{F}_s,$$

in particular,

$$\mathbb{E}[X_t - X_s|\mathcal{F}_s] = \mathbb{E}[X_t - X_s]$$

We note martinagles almost have this property:

Proposition 2.1.6. *Let X_t be martingale adapted to $\mathcal{F}_t$, then X_t has orthogonal increments.*

The proof is simple:

Proof.

$$\mathbb{E}[X_t - X_s|\mathcal{F}_s] = \mathbb{E}[X_t|\mathcal{F}_s] - X_s = X_s - X_s = 0.$$

□

Now apperently, if a process has independent increments, and it has constant expectation, then it would be a martingale (assuming integrability). Here is a theorem on that, which will come in handy later:

Theorem 2.1.1. *Let X_t be a process with independent increments, that is, $X_t - X_s \perp \mathcal{F}_t$. Then*

1. *If $X_t \in L^1$ for all t, then $Z_t = X_t - \mathbb{E}[X_t]$ is a martinagle.*
2. *If $X_t \in L^2$ for all t, then $Y_t = Z_t^2 - \mathbb{E}[Z_t^2]$ is a martingale, where Z_t as in (1).*
3. *If for some θ, $E_t = \mathbb{E}[\exp(\theta X_t)] < \infty$ for all $t \geq 0$, then $\frac{\exp(\theta X_t)}{\mathbb{E}[E_t]}$ is a martinagle.*

Proof. (1):

$$\begin{aligned}\mathbb{E}[Z_t - Z_s|\mathcal{F}_s] &= \mathbb{E}[X_t - \mathbb{E}[X_t] - X_s + \mathbb{E}[X_s]|\mathcal{F}_s] \\ &= \mathbb{E}[X_t - X_s|\mathcal{F}_s] - \mathbb{E}[X_t - X_s] = 0\end{aligned}$$

where last equality from independence.

(2):

$$\begin{aligned}\mathbb{E}[Z_t^2|\mathcal{F}_s] &= \mathbb{E}[(Z_t - Z_s + Z_s)^2\,|\mathcal{S}] \\ &= \mathbb{E}[(Z_t - Z_s)^2 + 2Z_s(Z_t - Z_s) + Z_s^2|\mathcal{F}_s] \\ &= \mathbb{E}[(Z_t - Z_s)^2] + \mathbb{Z}_s^2 \\ &= \mathbb{E}[Z_t^2] - 2\mathbb{E}(Z_tZ_s) + Z_s^2\end{aligned}$$

However, $\mathbb{E}[Z_tZ_s] = \mathbb{E}[\mathbb{E}[Z_t|\mathcal{F}_s]Z_s] = \mathbb{E}[Z_s^2]$, plug it in, we have the desired result.

(3):

$$\begin{aligned}\mathbb{E}\left[\frac{\exp(\theta X_t)}{E_t}|\mathcal{F}_s\right] &= \mathbb{E}\left[\frac{\exp(\theta(X_t - X_s))\exp(\theta X_s)}{\mathbb{E}[\exp(\theta(X_t - X_s))\exp(\theta X_s)]}|\mathcal{F}_s\right] \\ &= \mathbb{E}[\exp(\theta(X_t - X_s))]\frac{1}{\mathbb{E}[\exp(\theta(X_t - X_s))]}\frac{\exp(\theta X_s)}{\mathbb{E}[\exp(\theta X_s)]} \\ &= \frac{\exp(\theta X_s)}{\mathbb{E}[\exp(\theta X_s)]}\end{aligned}$$

where we used independence for the condition expectation and the ordinary expectation. □

An immediate result is the following

Corollary 2.1.1. *Let X_t be a stochastic process adapted to $\{\mathcal{F}_t\}_t$, and suppose X_t has independent increments and has mean zero, then $X_t^2 - \mathbb{E}[X_t^2]$ is a martinagle.*

We will recall this fact when doing quadratic variation and Brownian motion.

Here is an important formula that will simplify our future calculation greatly:

Proposition 2.1.7. *Let $\{X_t\}$ be an square integrable martingale, that is, $X_t \in L^2$ for all $t \geq 0$. Let $\{t_i\}_{i=0}^n$ be a partition of the interval $[s,t]$, then*

$$\mathbb{E}\left[\sum_{i=1}^{n}\left(X_{t_i} - X_{t_{i-1}}\right)^2|\mathcal{F}_s\right] = \mathbb{E}[X_t^2 - X_s^2] = \mathbb{E}[(X_t - X_s)^2].$$

In particular, above is true if we replace the conditional expectation with ordinary expectation.

Proof. We first calculate the conditional expectation of each term and recall the tower property for conditional expectation:

$$\mathbb{E}[\left(X_{t_i} - X_{t_{i-1}}\right)^2|\mathcal{F}_s] = \mathbb{E}\left[\mathbb{E}\left[\left(X_{t_i} - X_{t_{i-1}}\right)^2|\mathcal{F}_{t_{i-1}}\right]|\mathcal{F}_s\right]$$

Now the inner conditional expectation becomes

$$\mathbb{E}\left[X_{t_i}^2 - 2X_{t_i}X_{t_{i-1}} + X_{t_{i-1}}^2|\mathcal{F}_{t_{i-1}}\right] = \mathbb{E}[X_{t_i}^2|\mathcal{F}_{t_i}] - X_{t_{i-1}}^2.$$

So the whole thing becomes

$$\mathbb{E}[X_{t_i}^2|\mathcal{F}_s] - \mathbb{E}[X_{t_{i-1}}^2|\mathcal{F}_s]$$

so the original sum is telescoping, hence we have the first identity. The second identity is obtianed in the similar way:

$$\mathbb{E}[(X_t - X_s)^2] = \mathbb{E}[\mathbb{E}[X_t^2 - 2X_tX_s + X_s^2]|\mathcal{F}_s]] = \mathbb{E}[X_t^2 - X_s^2].$$

□

Inequalities for Continuous time Martingales

Here we basically have the same result as in discrete martinagle theories.

Proposition 2.1.8. *Let (X_t) be an adapted process and let $f : \mathbb{R} \to \mathbb{R}^+$ (nonnegative) be a convex function such that $f(X_t)$ is integrable for all $t \geq 0$, then*

1. *If X_t is a martiangle, then $f(X_t)$ is a sub-martinagle.*
2. *IF X_t is a sub-martinagle and f is non-decreasing, then $f(X_t)$ is a sub martingale as well.*

Proof. The first assertion is straight forward.
(2)

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] \geq f\left(\mathbb{E}[X_t|\mathcal{F}_s]\right) \geq f(X_s)$$

where the last inequality requires f to be non-decreasing. □

The Doob's Maximal Inequality and Doob's L^p Inequality extend easily to the continuous case:

Proposition 2.1.9. 1. *Let $(X_t)_t$ be a super martinagle with right continuous sample paths, then for all $t > 0$ and $\lambda > 0$, we have*

$$\lambda \mathbb{P}[\sup_{0 \leq s \leq t} |X_t|] \leq 2\mathbb{E}[|X_t|] + \mathbb{E}[|X_0|].$$

2. *Let (X_t) be a martinagle with continous sample paths, and suppose $\sup_{0 \leq s \leq t} \mathbb{E}[|X_t|^p] < \infty$ for $p > 1$, then*

$$\mathbb{E}\left[\sup_{0 \leq s \leq t} |_s|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_t|^p].$$

Proof. For (2), it is just an extension of the discrete case of the Theorem 25. Let $D = \mathbb{Q} \cap [0,t] \cup \{t\}$, then apply Theorem 25 to the martinagle $(X_t)_{t \in D}$ which is discrete and take the advantage of continuity, the result here is easily seen.

(1) is a consequence of Proposition 40. Let D_m be a partition of $[0,t]$, and it gets finer as m increases and $D_m \in D_{m+1}$, also say $D = \bigcup_{m \geq 1} D_m$ is dense in $[0,t]$, then by the discrete case we have

$$\lambda \mathbb{P}\left[\max_{i \in D_m} |X_i| \geq \lambda\right] \leq 2\mathbb{E}[X_t] + \mathbb{E}[|X_0|].$$

Now take $m \to \infty$ we would obtain this inequality on the dense set D, and take advantage of right continuity will get us the desired result. □

Remark 2.1.1. *If we change the interval $[0,t]$ to $[s,t]$ for $s < t$, above Proposition still holds.*

Next we talk about up-crossing inequality in continous setting, where we require the submartinagle to be right continuous. Here is the set up

For $\alpha \leq \beta$, Let $\tau_1 = \min_{t \geq 0}\{X_t \leq \alpha\}$, and define

$$\sigma_n = \inf_t \{\tau_{n-1} \leq t, X_t > \beta\}$$

$$\tau_{n+1} = \inf_t \{\sigma_n \leq t, X_t < \alpha\}$$

Let F be a finite set, then we define the up-crossing number to be the maximum j such that $\sigma_j \neq \infty$, or in plain words, the number of up-crossings. Here is the upcrossing inequality, we state without proof since it is just a generalization of the up-crossing in discrete case:

Theorem 2.1.2. *Let X_t be a submartinagle with right continuous sample paths, and suppose $-\infty < \alpha < \beta < \infty$ be real numbers, then we have*

$$(\beta - \alpha)\,\mathbb{E}[U_{[s,t],\alpha,\beta}] \leq \mathbb{E}[(X_t - \alpha)^+] + |\alpha|.$$

Convergence of Continuous Time Martinagles

Like above, results are very similar to the discrete case.

Theorem 2.1.3. *Let X_t be a submartingale with right continous sample pathes, and suppose $\sup_{t\geq 0}\mathbb{E}[X_t^+] < \infty$, then there is $X_\infty \in L^1$ such that $X_n \to X_\infty$ a.e.*

Remark 2.1.2. *Note again here it does not say converge in L^1.*

Proof. The proof is of the same spirit as in the discrete case. For all $\alpha < \beta$ that are finite, by Up-crossing inequality and the fact $\sup_{t\geq 0}\mathbb{E}[X_t^+] \leq C < \infty$, we have

$$\mathbb{E}[U_{[0,n],\alpha,\beta}] \leq \frac{\mathbb{E}[(X_n - \alpha)^+] + \alpha}{\beta - \alpha} < C_{\alpha,\beta} < \infty$$

for all α, β where $C_{\alpha,\beta}$ only depens on α and β. So we take $n \to \infty$, we see $U_{[0,n],\alpha,\beta} < \infty$ a.e. so

$$\mathbb{P}\left[\bigcup_{\alpha<\beta,\alpha,\beta\in\mathbb{Q}} U_{[0,\infty],\alpha,\beta} < \infty\right] = 1$$

and so by the same logic as in the discrete case, we have a.e. convergence. To show $X_\infty \in L^1$, we apply Fatou's lemma, but first observe that

$$\mathbb{E}[|X_n|] = \mathbb{E}[X_n^+] + \mathbb{E}[X_n^+] - \left(\mathbb{E}[X_n^+] - \mathbb{E}[X_n^-]\right) = 2\mathbb{E}[X_n^+] - \mathbb{E}[X_n] \leq C' - 2\mathbb{E}[X_0]$$

for some C'. So

$$\infty > \underline{\lim}_n \mathbb{E}[|X_n|] \geq \mathbb{E}[|X_\infty|].$$

□

We give a special name for (sup/sub)martingale with last element:

Definition 2.1.6. *Let (X_t) be a martingale, we say it is closed if there is some $X_\infty \in L^1$ such that*

$$X_t = \mathbb{E}[X_\infty|\mathcal{F}_t], \quad \forall t \in \mathbb{R}^+.$$

Here is the convergence result that is related to uniformly integrability:

Theorem 2.1.4. *Let (X_t) be a martingale with right continuous sample pathes, then the following are equivalent:*

1. *X_t is closed;*
2. *X_t is uniformly integrable.*
3. *X_t converges to some X_∞ both a.e. and in L^1.*

Remark 2.1.3. *If any of the above condition is satisfied, then the process can simply be written as $(\mathbb{E}[X_\infty|\mathcal{F}_t])$, very nice condition.*

Proof. (1) $\Rightarrow$ (2) (exactly same proof as before): Call the last element $X \in L^1$, let $M > 0$ and consider

$$\int_{|X_t|\geq M} |\mathbb{E}[X|\mathcal{F}_t] \leq \int 1_{|X_t|\geq M}|X| \tag{2.1}$$

but

$$\mathbb{P}\left[|X_t| \geq M\right] \leq \frac{1}{M}\mathbb{E}[|X_t|] \leq \frac{\mathbb{E}[|X|]}{M} \to 0, \quad \text{as } M \to \infty.$$

So take M to ∞, the left hand side of (2.1) converges to zero.

(2) $\Rightarrow$ (3): By submartinagle convergence theorem, we have convergence a.e., and uniformly integrability implies convergence in L^1.

(3) $\Rightarrow$ (1): Let $A \in \mathcal{F}_t$, then $\int_A X_\sigma = \int_A X_t$ for all $\sigma \geq t$ by definition of martingale. Take σ to ∞, by L^1 convergence, we see $\int_A X_\infty = \int_A X_t$ for all $A \in \mathcal{F}_t$ and for all t. □

Optional Sampling

We will use the following thing to prove optional sampling theorem.

In the discrete setting, we worked with martingales that have a starting point but not necessarity have an end point. In the proof of Doob's Maximal inequality in the continuous setting, we took a dense set from the interval $[0, t]$, which contains both 0 and t, and this is a case where a discrete martingale that has a first and a last element. There is one case we have not talked about yet, which is a (sub/super) martingale that has an end point but not a starting point. One of such thing can be created by, for example, taking a dense set in $[0, t]$ that includes t but not zero, call this set A. And if (X_t) is a continuous martingale, then $\{X_i\}_{i\in A}$ would be a martinagle that has an end point, but not a starting point.

In discrete setting, we call such (sub/super) martingales the backward (sub/super) martingales, it is nothing but a mtg with an end element but without a starting element. Since it has an end element, we might as well take advantage of it: Let $-\mathbb{N}$ be the set of nonpositive integers with its natrual orders, and suppose $\mathcal{F}_{-n} \supset \mathcal{F}_{-n-1}$, then a stochastic process $\{X_{-n}\}_{-n\in-\mathbb{N}}$ with $\mathbb{E}[X_{-n}|\mathcal{F}_{n-1}] = X_{n-1}$ is a backward mtg.

Here are some convergence results for backward mtg's

Proposition 2.1.10. *Let $(X_{-n})_{-n\in-\mathbb{N}}$ be a backward sub-mtg with respect to $\mathcal{F}_{-n}$ as described above. Then*

$$\lim_{n\to\infty} X_{-n} = X_{-\infty}, \quad \textit{a.e. for some } X_{-\infty},$$

and $X_{-\infty} \in [-\infty, \infty)$.

Remark 2.1.4. *In the case of a backward martingale, we don't need any condition for it to converge a.e., but we lose integrability on the convergence element. We will see this is due to the fact that the upper bound for up-crossing inequality already exists.*

Proof. For any $n > 0$, $\{X_{-k}\}_{-n\leq-k\leq-1}$ is a sub-martingale, so we have up-crossing inequality which is the following (with our usual notation):

$$\mathbb{E}[U_{\alpha,\beta,\{-n,\dots,-1\}}] \leq \frac{\mathbb{E}[(X_{-1}-\alpha)^+ + \alpha]}{\beta-\alpha}.$$

Taking $n \to \infty$ we see that the up-crossing is bounded by a constant (here it can be viewed as down-crossing if we look at it backward in time), so $U_{\alpha,\beta,\{-N\}} < \infty$ a.e.. This holds for any α, β, so we a countable dense union:

$$\mathbb{P}\left[\bigcup_{\alpha<\beta, \quad \alpha,\beta\in\mathbb{Q}} E_{\alpha,\beta,-\mathbb{N}} < \infty\right] = 1.$$

So limit exists. However, since $\mathbb{E}[X_{-1}|\mathcal{F}_{-n}] \geq X_{-n}$, so it does not diverge to infinity, but we can't prevent it to diverge to $-\infty$ here. □

Remark 2.1.5. *Backward mtg's can appears in another form, namely, σ-field decreases as n increases, and there is a starting element. For example:* $\mathcal{F}_n$ *decreases, and* $\mathbb{E}[X_n|\mathcal{F}_{n+1}] \geq X_{n+1}$*, this is also an backward martingale, we only need to reindex the process: identify* X_1 *to be* Y_{-1} *and let* $Y_{-n} = X_n$*, and* $\mathcal{G}_{-n} = \mathcal{F}_n$*, then we have the backward mtg in the above form.*

Here is another convergence result for backward mtg:

Proposition 2.1.11. *Let* $(X_{-n})_{-n\in-\mathbb{N}}$ *be a backward sub-martingale w.r.t.* $\mathcal{F}_{-n}$*'s, then the following conditions are equivalent*

1. X_{-n}*'s forms a uniformly integrable family.*
2. X_{-n} *converges in* L^1 *to* $X_{-\infty}$*.*
3. $\{X_{-n} : -n \in -\mathbb{N} \cup \{-\infty\}\}$ *is a closed submartingale (with first and last element).*
4. $\lim_{n\to\infty} \mathbb{E}[X_{-n}] > -\infty$.

Proof. $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4)$ follows similar proof for regular discrete maringale, we only need to show the case from (4) to (1). So consider the following:

$$\int_{|X_{-n}|>M} |X_{-n}| = \int_{|X_{-n}|>M} X_{-n}^+ + \int_{|X_{-n}|>M} X_{-n}^-$$

where

$$\mathbb{P}\left[|X_{-n}| \geq M\right] \leq \mathbb{P}\left[|\mathbb{E}[X_{-1}|\mathcal{F}_{-n}]| \geq M\right] \leq \mathbb{P}\left[|\mathbb{E}[|X_{-1}||\mathcal{F}_{-n}] \geq M\right] \leq \frac{\mathbb{E}[|X_{-1}|]}{M}$$

which converges to zero. Now we use the usual trick to get rid of X_n^- because we don't know much about negative part of sub martingale:

$$\begin{aligned}
\int_{|X_{-n}|>M} X_{-n}^+ + \int_{|X_{-n}|>M} X_{-n}^- &= \int_{|X_{-n}|>M} X_{-n}^+ + \int_{|X_{-n}|>M} X_{-n}^+ - \left(\int_{|X_{-n}|>M} X_{-n}^+ - \int_{|X_{-n}|>M} X_{-n}^-\right) \\
&= 2\int_{|X_{-n}|>M} X_{-n}^+ - \int_{|X_{-n}|>M} X_{-n} \\
&\leq 2\int_{|X_{-n}|>M} X_{-n}^+ + \left|\int_{|X_{-n}|>M} X_{-1}\right|
\end{aligned}$$

Now we want to bound the integrands, note X_{-n}^+ forms an sub-martingale by Jessen's, so

$$\int_{|X_{-n}|>M} X_{-n}^+ \leq \int_{|X_{-n}|>M} X_{-1}^+$$

and the second term obviously goes to zero when $M \to \infty$, so the whole thing is uniformly integrable. □

Now we are ready to state and prove the optional sampling theorem for right continous mtg's, even though it is likely we will only stick with mtg's with continuous sample pathes.

Theorem 2.1.5. *[Optional Sampling Theorem] Let $(X_t)_{t\in\mathbb{R}^+}$ be an uniformly integrable sub-martingale with right continuous sample pathes, and let $\tau \leq \sigma$ be stopping times, then*

$$\mathbb{E}[X_\sigma|\mathcal{F}_\tau] \geq X_\tau.$$

Proof. The strategy is to break it down to familiar things, namely, discrete stopping times that we worked with in previous chapter. So define τ_n and σ_n by the following

$$\tau_n(\omega) = \begin{cases} \infty & \tau(\omega) = \infty \\ \frac{k}{2^n} & \frac{k-1}{2^n} \leq \tau(\omega) < \frac{k}{2^n} \end{cases}; \quad \sigma_n(\omega) = \begin{cases} \infty & \tau(\omega) = \infty \\ \frac{k}{2^n} & \frac{k-1}{2^n} \leq \sigma(\omega) < \frac{k}{2^n} \end{cases}.$$

Then as discussed previously, τ_n and σ_n are stopping times that takes countably infinitely many values, and we still have $\tau_n \leq \sigma_n$. To have a discrete martingale that resembles the properties we want, let $Y_k^n = X_{\frac{k}{2^n}}$, then $X_{\tau_n} = Y_{\tau_n}^{(n)}$, same for X_{σ_n}. Furthermroe $Y^{(n)}$ is also uniformly integrable sub-martingale. So by discrete optional sampling theorem, we have

$$\int_A X_{\tau_n} \leq \int_A x_{\sigma_n} \quad \forall A \in \mathcal{F}_\tau. \tag{2.2}$$

This is because $\tau \leq \tau_n$ for all n, so $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$ for all n (recall the properties of σ-field associated with stopping times).

Now observe, $\mathcal{F}_{\tau_n}$ is a family of decreasing σ-algebra, and by discrete optional sampling theorem,

$$\mathbb{E}[X_{\tau_n}|\mathcal{F}_{n-1}] \geq X_{\tau_{n-1}},$$

so here we have a backward martingale with starting element being X_{τ_1}. Also by discrete martingale optional sampling theorem, we see that $\mathbb{E}[X_0] \leq \mathbb{E}[X_{\tau_n}]$ for all n, so the expectation is bounded below, so by previous thoerem, we get X_{τ_n} is uniformly integrable and converges to X_τ both in L^1 and a.e. (right continuity). Taking limit $n \to \infty$ of (2.2) we have

$$\int_A X_\tau \leq \int_A X\sigma,$$

since $A \in \mathcal{F}_\tau$ is arbitrary, the proof is done. □

The next lemma is immediate:

Lemma 2.1.2. *Let X_t be a right continous sub-martingale and let $\tau \leq \sigma$ be bounded (by some constant) stopping time, then*

$$\mathbb{E}[X_\sigma|\mathcal{F}_\tau] \geq X_\tau.$$

The next corollary is immediate from the lemma:

Corollary 2.1.2. *Let X_t be a right continous sub-martingale, then $\{X_{\tau\wedge t}; \mathcal{F}_{t\wedge\tau}\}$ is a sub-martinagle when τ is an stopping time.*

2.2 Other Martingales

From now on, everything will be continous (Yeah!)

This section I will talk about local mtgs, semi-martingales and square integrable martingales. The first two "martingales" are not real martinagle, but are two wider classes of stochastic process that will be helpful when doing stochastic integrations.

2.2.1 Local Martingales

We will be working with local martingales often:

Definition 2.2.1 (Local Martingale)**.** *An adapted stochastic process (X_t) w.r.t. $\{\mathcal{F}_t\}$ is a local martingale if there is an increasing sequence of stopping times τ_n with $\tau_n \uparrow \infty$ that makes $(X_{t\wedge\tau_n})$ a (true) martingale for all n.*

Remark 2.2.1. *Here is one of the motivation of local martingales: often a process (X_t) almost satisfies the condition of being a true martingale but false to be in L^1. In that case, we can define certain kinds of stopping times to bound it (localize it) so that it is in L^1. Such stopping times are usually $\tau_n = \inf_t\{|X_t| \geq n\}$, when X_t is continuous.*

Remark 2.2.2. *The above definition is from ([KS12]), the definition for local martingale from ([LG16]) also requres $(X_{t\wedge\tau_n})$ to be uniformly integrable, but I don't see a difference in reality. But gives us a hint that uniformly integrability imposed on a local martingale does not neccessarily make it a true martingale.*

Terminology: if (X_t) is a local martinagle such that $(X_{t\wedge\tau_n})$ are true mtg's for all n and $\tau_n \uparrow \infty$, then we say $\{\tau_n\}$ reduces X_t.

The following lemma is a direct consequence of optional sampling.

Lemma 2.2.1. *If (X_t) is a local martingale that can be redueced by $\{\tau_n\}$, then it can also be reduced by $\{\tau_n \wedge \sigma\}$ for any bounded stopping time σ (if we define local martingale according to ([LG16]), then we can remove "bounded" becuase of uniformly integrability).*

Here are some properties of local mtg's:

Theorem 2.2.1. *1. A nonnegative continuous local martingale with $M_0 \in L^1$ is a super martingale.*

2. Suppose M_t is a continuous local martingale and $Z \in L^1$ such that $|M_t| \leq Z$ for all t, then M_t is a uniformly integrable martingale.

3. For any continuous local martingale (M_t), the following sequence of stopping times reduces (M_t):

$$\tau_n = \inf_{t\geq 0}\{|M_t| \geq n\}.$$

Proof. (1) Let (M_t) be a positive local martingale that starts with an L^1 element, let τ_n reduce M_t and let $0 \leq s \leq t$, then by definition we have

$$\mathbb{E}[M_{t\wedge\tau_n}|\mathcal{F}_s] = X_{s\wedge\tau_n}$$

taking liminf on both side and apply Fatou's lemma for conditional expectation we see

$$\mathbb{E}[M_t|\mathcal{F}_s] \leq X_s.$$

Only problem remains is integrability, but this can be taken care of by setting $s = 0$.

(2)

$$\int |\mathbb{E}[X_{t\wedge\tau_n}|\mathcal{F}_s] - X_{s\wedge\tau_n}| = 0.$$

Taking $n \to \infty$ and apply DCT to get

$$\int |\mathbb{E}[X_t|\mathcal{F}_s] - X_s| = 0.$$

Uniformly integrability is straight forward.

(3) This is a direct consequence of (2) and the previous lemma: $M_{t\wedge\tau_n}$ is a local martingale bounded by n. □

2.2.2 Sqaure Integrable Martingales and Quadratic Variation

Definition 2.2.2. *Let (M_t) be a (true) martingale, we say it is square integrable if $M_t \in L^2$ for all t, and call the collection of square integrable martingales $\mathcal{M}^2$, and call the collection of continous square integrable martingales $\mathcal{M}^{2,c}$ or $\mathcal{M}_2^c$ which we use interchanagably.*

Previously we've seen that a submartinagle can be written as sum of true martingale and an increasing predictable process (Doob's Decomposition) in the discrete setting. There is one in continuous setting as well:

Theorem 2.2.2 (Doob-Meyer's Decomposition)**.** *Let $X_t, \mathcal{F}_t$ be a continous non-negative submartinagle, then there exists unique martingale M_t and an increasing process Z_t starting at 0 such that $X_t = M_t + Z_t$. Uniqueness is up to indistinguishability, meaning almost every path are the same.*

proof is long and tedious, omit.

By Doob-Meyer's decompsoition we can have the following definition:

Definition 2.2.3. *For all $M_t \in \mathcal{M}_2^c$, there is a unique increasing process, we call $\langle M \rangle_t$ such that $M_t^2 - \langle M \rangle_t$ is a martingale. We call $\langle M \rangle_t$ the quadratic variation of M_t.*

Remark 2.2.3. *The existance of such process is given by the fact M_t^2 is a submartinagle and Doob Meyer decompsoition.*

The definition for quadratic variation of local martingale is similar, the uniqueness and existnce is proven in ([LG16]), we just state the corresponding theorem w/o proof:

Theorem 2.2.3. *Let M_t be a local martiangle. Then there is an increasing process $\langle M \rangle_t$ such that $M_t^2 - \langle M \rangle_t$ is a local martingale, such process is unique up to indistinguishability.*

The following theorem gives us insight on what exactly is quadratic variation:

Theorem 2.2.4. *Let (M_t) be a local martingale and let Π_n be partitions of $[0, t]$ that gets finer as n gets larger and $\|\Pi_n\| \to 0$ as $n \to \infty$. Then*

$$\lim_{n\to\infty} \sum_{t_i, t_{i-1} \in \Pi} \left(M_{t_i} - M_{t_{i-1}}\right)^2 \to \langle M \rangle_t \quad \textit{in probability.}$$

Proof can be found on both ([KS12]) and ([LG16]).

Proposition 2.2.1. *suppose (M_t) is a local mtg and τ is a stopping time, then $\langle M^\tau \rangle_t = \langle M \rangle_{t\wedge\tau}$.*

Proposition 2.2.2. *Let (M_t) be a local martingale and suppose $\langle M\rangle_t = 0$ for $t \in [0, T]$, then $M_t = M_0$ on $[0, T]$.*

Proof. Assume WLOG that $M_0 = 0$.

$$\mathbb{E}[M_t^2] = \mathbb{E}[M_t^2 - \langle M\rangle_t] = 0$$

this is enough to show $M_t = 0$. □

The following theorem tells us quadratic variaiton is the right thing to look at for sqaure integrable martingales:

Theorem 2.2.5. *Let $\{X_t, \mathcal{F}_t, 0 \le t < \infty\}$ be a continuous stochastic process and let Π_n's be increasing partitions of $[0, t]$ with $\|\Pi\| \to 0$ as $n \to \infty$. Suppose that*

$$\lim_{n\to\infty} V_t^{(p)} \triangleq \lim_{n\to\infty} \sum_{t_i, t_{i-1} \in \Pi} \left|M_{t_i} - M_{t_{i-1}}\right|^p \to_{in\ prob} L_t$$

for some a.e. finite L_t. Then for all $\epsilon > 0$, $V^{(p+\epsilon)} \to 0$ and $V_t^{(p-\epsilon)} \to \infty$ on the event $\{L_t > 0\}$.

Proof.

$$\begin{aligned}\lim_{n\to\infty} V_t^{(p+\epsilon)} &= \lim_{n\to\infty} \sum_{t_i, t_{i-1} \in \Pi} \left|M_{t_i} - M_{t_{i-1}}\right|^{p+\epsilon} \\ &\le \lim_{n\to\infty} \max_i |M_{t_i} - M_{t_{i-1}}|^\epsilon \sum_{t_i, t_{i-1} \in \Pi} \left|M_{t_i} - M_{t_{i-1}}\right|^p.\end{aligned}$$

Continuity tells us $\max_i |M_{t_i} - M_{t_{i-1}}| \to 0$ as $n \to \infty$ a.e., and the other part of the sum converges to L_t in probability, hence the whole thing converges to 0 in probability.

For the second part, sitll the same method, but notice that on the event $L_t > 0$, for large enough n, we have $\max_i |M_{t_i} - M_{t_{i-1}}|^\epsilon > 0$.

$$\begin{aligned}\lim_{n\to\infty} V_t^{(p-\epsilon)} &= \lim_{n\to\infty} \sum_{t_i, t_{i-1} \in \Pi} \left|M_{t_i} - M_{t_{i-1}}\right|^{p-\epsilon} \\ &\ge \lim_{n\to\infty} \left(\max_i |M_{t_i} - M_{t_{i-1}}|\right)^{-\epsilon} \sum_{t_i, t_{i-1} \in \Pi} \left|M_{t_i} - M_{t_{i-1}}\right|^p \to \infty.\end{aligned}$$

□

The following theorem tells also when a local martingale is a true martingale bounded in L^2:

Proposition 2.2.3. *Let (M_t) be a local martingale with $M_0 \in L^2$, then the following are equivalent:*

1. *(M_t) is a true martingale bounded in L^2.*
2. $\mathbb{E}[\langle M\rangle_\infty] < \infty$.

Futhermore, if these properties holds, then $M_t^2 - \langle M\rangle_t$ is an uniformly integrable mtg.

Proof. (1) $\Rightarrow$ (2): Say $M_0 = 0$. Suppose $\sup_{t \in \mathbb{R}^+} \mathbb{E}[M_t^2] < \infty$ and M_t is a true martingale. Recall the relation $\mathbb{E}[M_t^2] = \mathbb{E}[\langle M \rangle_t]$ gives $\liminf_{n \to \infty} \mathbb{E}[\langle M \rangle_t] \geq \mathbb{E}[\liminf \langle M \rangle_\infty] = \mathbb{E}[\langle M \rangle_\infty]$ by Fatou's lemma.

(2) $\Rightarrow$ (1): Let $\{\tau_n\}_n$ be the increasing sequence of stopping times that reduces M_t, then

$$\mathbb{E}[\sup_{0 \leq s \leq t} M_{t \wedge \tau_n}^2] \leq C_2 \sup_{0 \leq s \leq t} \mathbb{E}[M_{t \wedge \tau_n}^2] = C_2 \sup_{0 \leq s \leq t} \mathbb{E}[\langle M \rangle_{t \wedge \tau_n}] \leq C\mathbb{E}\langle M \rangle_\infty.$$

We may take $n \to \infty$ first and take $t \to \infty$ to see $\sup_{t \geq 0} M_t^2$ is bounded in L^1 which dominates M_t^2. Another use of Jessen shows M_t is dominated by $\sup_{t \geq 0} |M_t| \in L^1$, so by previous theorems, M_t is a true mtg. □

The following lemma is an immediate consequence of this Proposition

Lemma 2.2.2. *Let (M_t) be a local martiangle with $M_0 \in L^2$, then the following are equivalent:*

1. *(M_t) is a true martingale and $M_t \in L^2$ for all $t \in \mathbb{R}^+$.*
2. *$\mathbb{E}[\langle M \rangle_t] < \infty$ for all $t \geq 0$.*

Remark 2.2.4. *The above two theorems tells us that local martingales are real (or $\mathbb{R}^n$) valued processes whose quadratic variations are not square integrable.*

Square Integrable Martiangles and Cross-Variation

Definition 2.2.4 (Cross-Variation)**.** *Let M_t, N_t be (local) Martiangles, cross variation is defined to be the stochastic process, called $\langle M, N \rangle$ such that $M_t N_t - \langle N, M \rangle_t$ is a (local) Maringale.*

Since there is a definition, we need to show such thing exists. To see this, see this, we write $N_t M_t$ as

$$N_t M_t = \frac{1}{2}\left((N_t + M_t)^2 - N_t^2 - M_t^2\right), \tag{2.3}$$

and we notice that

$$N_t M_t - \frac{1}{2}\left(\langle N + M \rangle_t - \langle N \rangle_t - \langle M \rangle_t\right)$$

is a martinagle. So the cross-variation of N_t and M_t is defined to be above. It is also unique since the quadratic variation of square integrable martignale is unique. It is of finite variation because quadratic variation has finite variation. Also $\langle M, N \rangle_t = \langle N.M \rangle_t$, and it is bilinear.

Similarly, one can also define the cross variation to be

$$\langle M, M \rangle_t \triangleq \frac{1}{4}\left(\langle M + N \rangle_t - \langle M - N \rangle_t\right) \tag{2.4}$$

which is the same as the above definition.

Here are some properties of quadratic variation for *local martiangles*, which will also be true for true martingales:

Proposition 2.2.4. *Let M_t and N_t be two local martingales with first element being sqaure integrable, then*

1. *The mapping $(M_t, N_t) \mapsto \langle M_t, N_t \rangle$ is bilear and symmetric.*

2. *Let Π_n be an increasing sequence of partitions of $[0,t]$ and $\|\Pi_m\| \to 0$, then*

$$\lim_{n\to\infty} \sum_{t_i,t_{i-1}\in\Pi_n} \left(M_{t_i} - M_{t_{i-1}}\right)\left(N_{t_i} - N_{t_{i-1}}\right) \to \langle N, M\rangle_t \quad \text{in probability.} \tag{2.5}$$

3. *For all stopping time τ, the cross-variaiton of the stopped process equals to the stopped corss-variation:*

$$\langle M^\tau, N^\tau\rangle_t = \langle M, N\rangle_{t\wedge\tau},$$

   where $M_t^\tau = M_{\tau\wedge t}$ is called the stopped process.

4. *Suppose N_n, M_t are two martingales (continuity is always assumed), then $N_t M_t - \langle n, M\rangle_t$ is also uniformly integrable, and $\langle M, N\rangle_\infty$ exists as an a.e. limit of $\langle M, N\rangle_t$ and it is integrable which satisfy the the following equation:*

$$\mathbb{E}[N_\infty M_\infty] - \mathbb{E}[N_0 M_0] = \mathbb{E}[\langle N, M\rangle_\infty].$$

Proof. The only thing that is not obvious is (2): Let's denote $\Delta_i^N = N_{t_i} - N_{t_{i-1}}; \Delta_i^M = M_{t_i} - M_{t_{i-1}}$, so the summants of (2.5) is

$$\Delta_i^N \Delta_i^M = \frac{1}{2}\left(\left(\Delta_i^N + \Delta_i^M\right)^2 - \left(\Delta_i^N\right)^2 - \left(\Delta_i^M\right)^2\right)$$

then the desired result is given by (2.3). □

Notice that $\langle\cdot,\cdot\rangle$ is somewhat like inner product, or it really wants to be an inner product. There are some elements missing though: Cauchy-Schwartz, orthogonality and the underlying space. We will take care of those in the next step.

Definition 2.2.5 (Orthogonal Processes)**.** *Two local martinagles M, N are said to be orthognoal to each other if*

$$\langle M, N\rangle \equiv 0 \quad \text{or } \langle M, N\rangle_t = 0 \forall t \geq 0,$$

this can only happen when $N_t M_t$ itself is an local martingale.

Cauchy-Schwartz:

Theorem 2.2.6 (Kunita-Watanabe)**.** *Let M, N be two conitnous local martingales, and H, K be two measurable processes, then*

$$\int_0^t |H_s||K_s||d\langle N, M\rangle_s| \leq \left(\int_0^t |H_s|^2 d\langle N\rangle_s\right)^{\frac{1}{2}} \left(\int_0^t |K_s|^2 d\langle M\rangle_s\right)^{\frac{1}{2}}$$

for all $t \geq 0$.

Remark 2.2.5. *This is only interesting when all above are finite.*

Proof. Notice that we only need to show this for H, K are simple functions, so only neet to show it for H, K constant, which means we only needs to show it for $H = K = 1$, that is, to show

$$\int_0^t |d\langle M, N\rangle_s| \leq (\langle M\rangle_t)^{\frac{1}{2}} (\langle N\rangle_t)^{\frac{1}{2}}.$$

But the left hand side is nothing but the total variation of $\langle M, N\rangle_t = \frac{1}{2}\left(\langle N+M\rangle_t - \langle N\rangle_t - \langle M\rangle_t\right) = \frac{1}{4}\left(\langle M+N\rangle_t - \langle M-N\rangle_t\right)$ (see (2.4)), which is the difference of two finite variation process. However, the total variation of this is nothing but the sum of those two finite variation process (see, for example, [Fol99] Theorem 3.27). That is, we need to show

$$\frac{1}{4}\left(\int_0^t d\langle M+N\rangle_s + \int_0^t d\langle M-N\rangle_s\right) = \frac{1}{4}\left(\langle M+N\rangle_t + \langle M-N\rangle_t\right) \le \left(\langle M\rangle_t\right)^{\frac{1}{2}}\left(\langle N\rangle_t\right)^{\frac{1}{2}}$$

But this is obvious by the bilinear property of cross variation:

$$\langle M+N\rangle_t + \langle M-N\rangle_t = 2\langle M\rangle_t + 2\langle N\rangle_t \le 4\left(\langle M\rangle_t\right)^{\frac{1}{2}}\left(\langle N\rangle_t\right)^{\frac{1}{2}}.$$

And we note that if $H_s \equiv \alpha$ and $K_s \equiv \beta$, then it stil holds.

Now the only sketchy part of this proof is that the step function might not start at zero, e.g. $H = 1_{[\tau,\sigma]}$. But this case can be taken care of by starting the process N and M at τ. □

Remark 2.2.6. *From the above proof, we see that $\langle\cdot,\cdot\rangle$ has Cauchy-Schwartz inequality. So if there is an space that is closed in some norm induced by $\langle\cdot,\cdot\rangle$, then it would be a Hilbert space when paired with cross variation as the inner product. Now we take care of that part.*

Definition 2.2.6 (The Hilbert Space)**.** *On a filtered probability space, we define $\mathbb{H}$ to be the space of all L^2 bounded continuous martingales that start at zero ($M_0 = 0$), with norm being $\sqrt{\mathbb{E}[\langle M\rangle_\infty]}$ and inner product being $\mathbb{E}\langle M, N\rangle_\infty$.*

Proposition 2.2.5. *The above definition makes sense.*

Proof. We have seen previously that an L^2 bounded martingale, then $\langle M\rangle_\infty$ exists and is integrable (L^p inequality, monotone). $\|\cdot\|_{\mathbb{H}}$ make sense because if $\|M\|_{\mathbb{H}}^2 = \mathbb{E}[\langle M\rangle_\infty] = 0$, then $\langle M\rangle_t = 0$ for all $t \ge 0$ (increasing and nonnegative), so $M_t = 0$ for all $t \ge 0$, so $\|\cdot\|_{\mathbb{H}}$ is indeed a norm. The inner product is also indeed an inner product thanks to Kunita-Watanabe Inequality. □

Theorem 2.2.7. $\left(\mathbb{H}, \sqrt{\mathbb{E}\langle\cdot,\cdot\rangle}\right)$ *defines an Hilbert space.*

Proof. Here, we need to show every Cauchy sequence in $\mathbb{H}$ converges to a L^2 bounded continuous martinagle (unifromly integrable).

Let $\{M^n\} \subset \mathbb{H}$ be a cauchy sequence, that is,

$$\lim_{n,m\to\infty} \mathbb{E}\left(\langle M^n - M^m\rangle_\infty\right) = 0$$

But we also have for all $t \ge 0$:

$$\mathbb{E}\left(\langle M^n - M^m\rangle_\infty\right) \ge \sup_{t\in\mathbb{R}^+} \mathbb{E}[(M_t^n - M_t^m)^2] \ge \frac{1}{C}\mathbb{E}\left[\sup_{t\ge 0}\left(M_t^n - M_t^m\right)^2\right] \to 0 \quad \text{as } n, m \to \infty. \tag{2.6}$$

So there is at least an $L^2(\Omega, \mathbb{P})$ limit of M_t for each t, call it M_t. Now we would need to show M_t is pathwise a.e. continuous and it is a Martingale (it is already square integrable).

First, it is a square integrable martingale: since it is L^2 convergent, then by Jessen, it is L^1 convergent. Let $A \in \mathcal{F}_s$, then

$$\int_A M_s = \lim \int_A M_s^n = \lim \int_A \mathbb{E}[M_t|\mathcal{F}_s] = \int_A \mathbb{E}[M_t|\mathcal{F}_s].$$

So M_t is a martinagle with no problem. Now, continuity is harder to get since M_t is a $L^2(L^1)$ limit of continous mtg, to get continuity, want it to be a uniform limit of a sequence of continous martingales. However, (2.6) tells us something very close to that: $M_\cdot$ is an uniform limit of $M_\cdot^n$ in probability, so there is a subsequence of $M_\cdot^n$, stil call it $M_\cdot^n$ that converges to $M_\cdot$ for almost all $\omega \in \Omega$, so we know $M_\cdot$ is continuous. This also insures $\langle M \rangle_\infty$ exists as limit of $\langle M^n \rangle_\infty$, which is also in L^1. $\square$

The following Proposition is an immediate consequence of the previous Theorem

Proposition 2.2.6. *With above settings,* $\mathbb{H}_T^2 \triangleq \{M_t, 0 \le t \le T; M_t \text{ is square integrable martingale}\}$ *is a Hilbert space with norm defined by* $\|\cdot\|_{\mathbb{H}_T^2}^2 = \mathbb{E}\langle \cdot \rangle_T$.

Remark 2.2.7. *This fact will be important in the development of Stochastic Integration.*

Continuous Semi-Martingales

This is the last "other" Martingales here:

Definition 2.2.7 (Semi-Martingale)**.** *We say* (X_t) *is a continuous semi-martinagle if it can be decomposed into*

$$X_t = M_t + A_t$$

where (M_t) *is a continuous local martingale and* A_t *is an adapted process with finite variation.*

One might have guessed that it also have a quadratic variation. By the approximation given by (2.5), we see (at least intuitively) the finite variation process does not contribute anything to the quadratic variation approximation (in probability).

Definition 2.2.8. *Let* $X_t = M_t + A_t$ *and* $Y_t = N_t + Z_t$ *be two semi-martingales where* M, N *are continous local martingales and* A, Z *be two finite variation process. We define the quadratic variation of* X *to be*

$$\langle X \rangle_t = \langle M \rangle_t$$

and the cross-variation between X *and* Y *to be*

$$\langle X, Y \rangle_t = \langle M, N \rangle_t .$$

The next theorem tells us why this makes sense:

Theorem 2.2.8. *Let* $\Pi_n = \{t_i\}_{i \in I}$ *be a partition of* $[0, t]$ *where* I *is a finite set, where* $\|\Pi_n\| \to 0$ *as* $n \to \infty$. *Let* $X_t = M_t + A_t$ *be a continous semi-maringale where* M_t *and* A_t *are continous local martingale and continuous finite variation process repsectively, then*

$$\lim_{n\to\infty} \sum_{t_i, t_{i-1} \in \Pi_n} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \to \langle X \rangle_t = \langle M \rangle_t \quad \textit{in probability.}$$

Proof. The inuition is from Theorem 35, let's calculate it directly:

$$\begin{aligned} \sum_{t_i, t_{i-1} \in \Pi_n} \left(X_{t_i} - X_{t_{i-1}}\right)^2 &= \sum_{t_i, t_{i-1} \in \Pi_n} \left(M_{t_i} - M_{t_{i-1}} + A_{t_i} - A_{t_{i-1}}\right)^2 \\ &= \sum_{t_i, t_{i-1} \in \Pi_n} \left(M_{t_i} - M_{t_{i-1}}\right)^2 + \sum_{t_i, t_{i-1} \in \Pi_n} \left(A_{t_i} - A_{t_{i-1}}\right)^2 \\ &\quad + \sum_{t_i, t_{i-1} \in \Pi_n} 2\left(M_{t_i} - M_{t_{i-1}}\right)\left(A_{t_i} - A_{t_{i-1}}\right). \end{aligned}$$

The first sum goes to $\langle M\rangle_t$ with no problem, the second sum goes to zero by Theorem 35. Now consider the last summants

$$\sum_{t_i,t_{i-1}\in\Pi_n} 2\left(M_{t_i}-M_{t_{i-1}}\right)\left(A_{t_i}-A_{t_{i-1}}\right) = \max_{t_i,t_{i-1}\in\Pi_n} 2\left|M_{t_i}-M_{t_{i-1}}\right| \sum_{t_i,t_{i-1}\in\Pi_n} \left|A_{t_i}-A_{t_{i-1}}\right|$$

which goes to zero for almost all ω by, again, Theorem (35). □

2.3 Brownian Motion

Here, we first look at some properties of a more general Gaussian process, and then build Brownian motion out of that. Then look at some path properties of Brownian motions.

2.3.1 Gaussian Processes

A preview of this section

1. Elementary facts about Gaussian random variables.
2. The space that Gaussian random variable generates.
3. Gaussian white noise.

Gaussian Random Variables

Let's recall some facts about Gaussian random variables.

Definition 2.3.1. *We say $X \sim \mathcal{N}(m,\sigma^2)$ on the real line if X has density function*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-m)^2}{\sigma^2})$$

For $\mathbb{R}^n$, the definition is similar: We say $X \sim \mathcal{N}(m,\Sigma)$, where Σ is a positive definite matrix, if the measure induced by X, call it μ, is absolutely continuous with respect to λ^n, the n-dim lebesgues measure and

$$\frac{d\mu}{d\lambda^n}(x) = \frac{1}{(2\pi\det(\Sigma))}\exp\left(-\frac{1}{2}\langle x-m,\Sigma^{-1}(x-m)\rangle\right)$$

we call Σ the covariance matrix of $(X_i)_{1\le i\le n}$ where the entries are $(\Sigma)_{i,j} = Cov(X_i, X_j)$

The moment generating function and the characteristic functions for the centered (mean zero) Gaussian measures (random variables), since that's what we are interested in, are the following:

Proposition 2.3.1. *Let μ be a Guassian measure with mean m and variance σ^2, then its Fourier transform is*

$$\hat{\mu}(\xi) = \exp\left(-\frac{1}{2}(\sigma\xi)^2\right); \quad \mathbb{E}[e^{\lambda x}] = \exp\left(\lambda^2\sigma^2/2\right).$$

Proof. Let $\lambda \in \mathbb{C}$ be a complex number and consider the following

$$\begin{aligned}\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} e^{\lambda x-\frac{1}{2}\frac{x^2}{\sigma^2}}dx &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} e^{\frac{1}{2}\sigma^2\lambda^2}e^{-\frac{1}{2}\sigma^2\lambda^2+\lambda x-\frac{1}{2}\frac{x^2}{\sigma^2}}dx \\ &= \exp(\frac{1}{2}\lambda^2\sigma^2)\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} e^{-\frac{(x-\sigma\lambda)^2}{\sigma^2}} \\ &= \exp\left(\frac{1}{2}\lambda^2\sigma\right)\end{aligned}$$

then let λ be either ξ or $\lambda \in \mathbb{R}$ gives us desired result ☐

Definition 2.3.2 (Jointly Guassian)**.** *Suppose X, Y are Guassian random variables, we say X, Y are jointly Guassian random variables if $\alpha X + \beta Y$ are Guassian random variables for all $\alpha, \beta \in \mathbb{R}$. Similar definition for multiple gaussian random variables.*

Following prop is useful

Proposition 2.3.2. *Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ adn $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then $X + y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Proof is easy algebra, omit.

Here is an interesting fact about jointly Guassian random variables:

Proposition 2.3.3. *Let $(X_1, ..., X_n)$ be a Guassian random variable (or Gaussian vector) in $\mathbb{R}^n$, then $\{X_i\}_{1\le i\le n}$ are independent if and only if the covariance matrix is diagonal.*

Proof. If part is siimple, we only show the only if part: suppose Σ, the covariance matrix, is diagonal, say its diagonal elements are λ_i^2's, and assume WLOG that $\mathbb{E}[X_i] = 0$ for all i, then its density is

$$\begin{aligned}\frac{1}{\sqrt{2\pi\det(\Sigma)}}\exp\left(-\frac{1}{2}x\cdot\Sigma^{-1}x\right) &= \frac{1}{\sqrt{2\pi\det(\Sigma)}}\exp\left(-\frac{1}{2}\sum_{1\le i\le n}\frac{x_i^2}{\lambda_i^2}\right) \\ &= \frac{1}{\sqrt{2\pi\prod_{1\le i\le n}\lambda_i^2}}\prod_{1\le i\le n}\exp\left(-\frac{1}{2}\frac{x_i^2}{\lambda_i^2}\right)\end{aligned}$$

hence independent. ☐

Gaussian vectors forms a closed space in L^2:

Theorem 2.3.1. 1. *Let $\{X_n\}_n$ be a sequence of Gaussian vectors such that $\mathbb{E}[X_n] = \mu_n$ and $Cov(X_n) = \Sigma_n$ where $\mu_n \to \mu$ and $\Sigma_n \to \Sigma$ in matrix norm ($\sum_{1\le i,j\le n}\left|\Sigma_{i,j}\right|$), and X_n converges in L^2 to some X, then $X \sim \mathcal{N}(\mu, \Sigma)$.*

2. *Same setting as (1) and assume L^2 convergence, then we have L^p convergence for all $1 \le p < \infty$.*

Proof. (1) is just simple application of Levy's continuity theorem.

(2) Let's work with 1-dimensional case, higher dimensions are similar. Since μ_n and σ_n are all bounded uniformly by some number, we can take advantage of the fact that all moments exists for normal random varibales, and the fact that higher moments of Guassian random variables depens completely on the first and second moments. So $\mathbb{E}|X_n|^p \to \mathbb{E}|X|^p$, then by Proposition 18, we have convergesn in L^p. ☐

There is a one-to-one relation between Guassian measures in $\mathbb{R}^n$ and nonnegative symmetric matricies in $GL(\mathbb{R}^n)$:

Proposition 2.3.4. *Suppose Σ is a symmetric nonnegative matrix, then there is a Gaussian measure with mean 0 and covariance matrix Σ.*

We omit the proof here, but the idea of the proof is to find an eigen-basis of Σ, call then $\{e_i\}_{1\le i\le n}$, and pair them up with i.i.d standard Guassian random variables.

Remark 2.3.1. *The above proposition hold in more general settings, e.g. in a Hilbert space, one can show there is a one-to-one correspondence between the family of centered Gaussian process and the family of nonnegative symmetric trace oprators. For more detail, see ([DPZ14]).*

Gaussian Process and Space

Definition 2.3.3.
1. *The (centered) Gaussian space is a subspace of L^2 that contains only the Gaussian random variables.*
2. *A (centered) Gaussian process is a process $(X_t)_{t\in T}$ where T is any index set, such that any finite linear combination of elements in the process gives a (centered) gaussian random variable. In other words, the family is (finitely) jointly Guassian.*

The next proposition is immediate from Thoerem 39:

Proposition 2.3.5. *Let $(X_t)_{t\in T}$ be a Gaussian process, then the linear span of elements of $(X_t)_{t\in T}$ in L^2 is a Guassian space, which is called Guassian space generated by $(X_t)_{t\in T}$.*

Remark 2.3.2. *Note that a centered Gaussian space is still a Hilbert space with L^2 inner product.*

Gaussian White Noise

Definition 2.3.4 (Gaussian White Noise)**.** *Let $(E,\mathcal{B}(E))$ be a measurable space, and let μ be a σ finite measure on it. A Gaussian white noise with intensity μ is an isometry from $L^2(E,\mu)$ to a centered Gaussian space.*

Remark 2.3.3. *If $\dot W$ is a Guassian white noise, and $f\in L^2(E,\mu)$ (we drop the σ-field here), then*

$$\mathbb{E}\left|\dot W(f)^2\right| = \int_E |f|^2 d\mu$$

and it also preserves inner products:

$$\mathbb{E}\left[\dot W(f)\dot W(g)\right] = \int_E fg d\mu.$$

In particular, it also works with indicator functions:

$$\mathbb{E}\left[\dot W(1_A)\dot W(1_B)\right] = \mu(A\cap B).$$

Note that if $f_n\to f$ in $L^2(\mu)$ and f_n's are simple functions, then $\dot W(f_n)\to \dot W(f)$ in $L^2(\Omega,\mathbb{P})$.

The following theorem shows existance of White noise:

Theorem 2.3.2. *Let $(E,\mathcal{E})$ be a fixed seperable measurable space, then for any σ finite μ, there is a white noise $\dot W$ with intensity μ.*

From the proof, we will see that the underlying space does not have to be $L^2(E,\mu)$ if we do not restrict the intensity to be a measure, it works for any Hilbert space.

Proof. Let $\{e_n\}_{n\in\mathcal{N}}$ be the orthornormal basis of $L^2(E,\mu)$, and let $\{X_n\}_{n\in\mathcal{N}}$ be i.i.d standard normal random variables with a common probability space $(\Omega,\mathcal{F},\mathbb{P})$, and consider the map

$$f \mapsto \sum_{n\in\mathcal{N}} \langle f, f_n\rangle X_n$$

is a gaussian white noise. □

Finally, Gaussian white noises have "quadratic variation":

Proposition 2.3.6. *Let $(E,\mathcal{E},\mu)$ be a seperable measure space where μ is σ finite and let $\dot{W}$ be an gaussian white noise with intensity μ. For $A\subset E$ and $A=\bigcup_{1\le i\le n} A_i^n$ where $\{A_i^n\}_{1\le i\le n}$ is disjoint for all n, and*

$$\lim_{n\to\infty}\max_{1\le i\le n}\mu(A_i^n)=0,$$

Then

$$\sum_{1\le i\le n}\dot{W}(A_i^n)^2\to\mu(A)\quad \text{in } L^2 \text{ as } n\to\infty.$$

Proof. Let's try direct computation:

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{1\le i\le n}\dot{W}(A_i^n)^2-\mu(A)\right)^2\right] &= \mathbb{E}\left[\left(\sum_{1\le i\le n}\dot{W}(A_i^n)^2-\mu(A_i^n)\right)^2\right]\\ &= \sum_{1\le i,j\le n}\mathbb{E}\left[\left(\dot{W}(A_i^n)^2-\mu(A_i^n)\right)\left(\dot{W}(A_j^n)^2-\mu(A_j^n)\right)\right]\end{aligned}$$

We let $B_i^n=\left(\dot{W}(A_i^n)^2-\mu(A_i^n)\right)$ and observe that for a fixed n, $B_i^n\perp B_i^n$, and $\mathbb{E}[B_i^n]=0$, so $\mathbb{E}[B_i^nB_j^n]=0$ for $i\ne j$. So only (i,i) term remains in the above sum:

$$\mathbb{E}\left[\left(\sum_{1\le i\le n}\dot{W}(A_i^n)^2-\mu(A)\right)^2\right]=\sum_{1\le i\le n}\mathbb{E}\left[\left(\dot{W}(A_i^n)^2-\mu(A_i^n)\right)^2\right]$$

computing each term:

$$\begin{aligned}\mathbb{E}\left[\left(\dot{W}(A_i^n)^2-\mu(A_i^n)\right)^2\right] &= \mathbb{E}\left[\dot{W}(A_i^n)^4\right]-2\mathbb{E}\left[\dot{W}(A_i^n)^2\mu(A_i^n)\right]+\mu(A_i^n)^2\\ &= 3\mu(A_i^n)^2-\mu(A_i^n)^2=2\mu(A_i^n)^2\end{aligned}$$

which we see goes to zero in the spirit of theorem 35. □

At this point, one can go off of stochastic calculus and dive right into Martinagle measures and Walsh's theory on stochastic PDE just as in ([DKM+09]), but we are going to develop Brownian motion from here.

2.3.2 Construction of Brownian Motion

We first define Brownian Motion and then show it exists.

Definition 2.3.5 (Brownian Motion/Wiener Process)**.** *A Brownian motion/Wiener process is a stochastic process, denoted by* (W_t) *such that*

1. $W_0 = 0$ *a.e.*
2. $W_\cdot(\omega)$ *is a continous function for almost every* ω.
3. W_t *has independent increments.*
4. $W_t - W_s \sim \mathcal{N}(0, t-s)$ *for all* $0 \le s \le t < \infty$.

We now build Wiener process step by step (honestly, there are only two steps). Also, I will use wiener process and Brownian motion interchanagably.

Definition 2.3.6 (Pre-Brownian motion)**.** *Let* $\dot{W}$ *be a Gaussian process defined on* $\mathbb{R}^+$ *with intensity being the ususal lebesgues measure, the process* B_t *defiend by*

$$B_t = \dot{W}([0,t])$$

is called a pre-Brownian motion.

We see that $\mathbb{E}[B_t B_s] = t \wedge s$.

We can see easily that a pre-Brownian motion satisfies (1),(3),(4) of the definiton of Brownian motion, the following theorem tells us exactly that:

Theorem 2.3.3. *Let* X_t *be a real valued stochastic process, then the followings are equivalent*

1. (X_t) *is a pre-Brownian Motion.*
2. (X_t) *is a centered Gaussian process with covariance* $K(t,s) = t \wedge s$.
3. $X_0 = 0$ *a.e. and* (X_t) *has independent increment and* $X_t - X_s \sim \mathcal{N}(0, t-s)$ *for all* $0 \le s \le t < \infty$.

Proof. $(1) \Rightarrow (2)$: this is obvious.

$(2) \Rightarrow (3)$: let $0u \le s \le t < \infty$ and let $\gamma > 0$, then

$$\mathbb{E}[X_u(X_t - X_s)] = \mathbb{E}[X_u X_t] = \mathbb{E}[X_u X_s] = u \wedge t - u \wedge s = u - u = 0$$

being jointly gaussian we see $X_u \perp (X_t - X_s)$ for all $u \le s$. More generally, $\sigma(X_u; 0 \le u \le s) \perp \sigma(X_t - X_s; s \le t < \infty)$. Now, only thing we need to do is to calculate the variance of $X_t - X_s$ where $s \le t$, since we know it is Guassian:

$$\mathbb{E}[(X_t - X_s)^2] = \mathbb{E}[X_t^2] - 2\mathbb{E}[X_t X_s] + \mathbb{E}[X_s^2] = t - 2t \wedge s - s = t - s.$$

$(3) \Rightarrow (1)$: Here we would like to construct a Guassian white noise with intensity λ where λ is 1-dim lebesgues measure. So we make an guess and see if it is a Gaussian white noise:

$$\dot{W}(\lambda 1_{[s,t]}) \triangleq \lambda (B_t - B_s)$$

which maps simple functions to centered gussian random variables and we see this is an isometry defined on the set of simple functions to a subset of centered gaussian random variables. Since simple functions are dense in L^2, we can extend it to the whole space of $L^2(\mathbb{R}^+)$ and use the fact that Gaussian random variables are closed in $L^2(\mathbb{P})$, we see $\dot{W}$ is a white noise and $B_t = \dot{W}([0,t])$. □

Remark 2.3.4. *So all of the three above can be used as the definition of pre-Brownian motion now.*

Now, only continuity is missing, but the following Kolmogorov's Continuity Criterion tells us when a process will be very close to continuous for almost every sample path. But before that, there are two more concpets we need to know:

Definition 2.3.7 (indistinguishability). *Let X_t, Y_t be two stochasti processes, we say they are indistinguishable if*

$$\mathbb{P}[X_t \neq Y_t; \forall t \geq 0] = 0.$$

That is, for almost all ω, $X_\cdot(\omega)$ and $Y_\cdot(\omega)$ are the same thing. Or, treating $X : \Omega \to \{measurable\ functions\}$, and same for Y, then $X = Y$ a.e.

Definition 2.3.8 (Modification/Version). *Let X_t, Y_t be two stochastic processes, then we say Y is a version of X, or Y is a modification of X if*

$$\mathbb{P}[X_t = Y_t] = 1 \quad \forall t \geq 0.$$

Remark 2.3.5. *Clearly, indistinguishability is stronger than modification. In particular, X has a continuous versio,*

Theorem 2.3.4 (Kolmogorov's Continuity Theorem). *Let (X_t) be a stochastic process on a complete seperable metric space, and suppose there is $\alpha, \beta > 0$ such that fro all $s, t \in [0, T]$*

$$\mathbb{E}|X_t - X_s|^\delta \leq C|t-s|^{1+\epsilon}$$

then there is a version of X that is Holder continuous on $[0, T]$ for all exponent $\alpha < \frac{\delta}{\epsilon}$.

Proof. Assume WLOG that $T = 1$.

The idea of the proof: we want deduce pathwise property from general infomation (such as expectations here), then only thing that can help us is Borel-Cantelli here. It is also hard to construct a continous process from scratch, but if X_t is continuous in a weaker sense, say continous or uniformly continous on a dense set, then we can do a natural extension to the whole set and create a continous process. The natrual dense set of the choice is the set of rational numbers. However, from the condition that is given, we would like to have some control over the distances, and the set of rational numbers does not give us that. The next choice would be dyadic rationals, that is, $\{\frac{a}{2^k} : a \in \mathbb{N}\}$, or I would like to call them the "rational binary numbers". It is a good choice because we can decompose them into layers:

Let $D_n = \{\frac{k}{2^n}; 0 \leq k \leq 2^n\}$, and let

$$D = \bigcup_{n \geq 1} D_n$$

then D is the collection of "rational binary numbers" between 0 and 1, and in each layer, we have some control over the distances. With the help of the elementary fact that binary representation of a rational number is unique, we are ready to prove the theorem.

Fix $n \geq 1$, and let $0 \leq k \leq 2^n$, then by Chebyshev we have

$$\mathbb{P}\left[\left|X\left(\frac{k}{2^n}\right) - X\left(\frac{k-1}{2^n}\right)\right| > \beta\right] \leq \frac{\mathbb{E}[\left|X\left(\frac{k}{2^n}\right) - X\left(\frac{k-1}{2^n}\right)\right|^\delta]}{\beta^\delta} \leq 2^{-(1+\epsilon)}\beta^{-\delta}.$$

Since we want to use Borel-Cantelli, we want to vary β w.r.t. n so we can sum things up. So the natrual choice for β for a fixed n is $\beta = 2^{-\gamma n}$, plug it in we see

$$\mathbb{P}\left[\left|X\left(\frac{k}{2^n}\right) - X\left(\frac{k-1}{2^n}\right)\right| > 2^{-\gamma n}\right] \leq 2^{-n(1+\epsilon-\gamma\delta)}$$

which implies

$$\mathbb{P}\left[\max_{1\leq k\leq 2^n}\left|X\left(\frac{k}{2^n}\right) - X\left(\frac{k-1}{2^n}\right)\right| > 2^{-\gamma n}\right] \leq 2^{-n(\epsilon-\gamma\delta)}$$

Here we want the exponent on the right hand side to be strictly greater than zero, so choose $\gamma < \frac{\epsilon}{\delta}$, and notice that

$$\sum_{n=1}^{\infty}\mathbb{P}\left[\max_{1\leq k\leq 2^n}\left|X\left(\frac{k}{2^n}\right) - X\left(\frac{k-1}{2^n}\right)\right| > 2^{-\gamma n}\right] \leq \sum_{n\geq 1}\left(2^{\epsilon-\gamma\delta}\right)^n < \infty.$$

Now, Borel-Cantelli tells us that there is $\widetilde{\Omega} \subset \Omega$ with $\mathbb{P}\left[\widetilde{\Omega}\right] = 1$, such that for all $\omega \in \widetilde{\Omega}$, there is $N(\omega) \geq 1$ such that for all $n \geq N(\omega)$, the path associated with ω has the property

$$\max_{1\leq k\leq 2^n}\left|X\left(\frac{k}{2^n}\right) - X\left(\frac{k-1}{2^n}\right)\right| \leq 2^{-\gamma n}$$

(for lack of space to put (ω)). This looks like uniform continuous on D, let's see if it is. Suppose $s, t \in D$ where $s \leq t$ and suppose $2^{-n-1} \leq |t-s| < 2^{-n}$, then $s, t \in D_m$ for some $m > n$, so $|t-s| \leq 2^{m-n}2^{-m}$, let m be the smallest such number, then $|t-s| \geq 2^{-m}$, hence $2^{-m} \leq |t-s| < 2^{-n}$, so

$$|t-s| = \sum_{=n+1}^{m} e_j 2^{-j} \quad e_j \in \{0, 1\}$$

and such representation is unique (binary). So we do the decomposition

$$\begin{aligned}
|X(t) - X(s)| &\leq \sum_{j=n+1}^{m-1}\left|X(s + \sum_{i=n+1}^{j+1} e_i 2^{-i}) - X(s + \sum_{i=n+1}^{j} e_i 2^{-i})\right| \\
&\leq \sum_{n+1}^{\infty} 2^{-\gamma j} \\
&= 2^{-\gamma(n+1)}\sum_{j=0}^{\infty} 2^{-\gamma j} \\
&= \frac{2^{-\gamma(n+1)}}{1-2^{-\gamma}} \leq \frac{|t-s|^{\gamma}}{1-2^{-\gamma}}.
\end{aligned}$$

So X is Holder continuous with exponent γ on D, pathwise. So we can extend it naturally to $[0, 1]$: if $t \in D$, then $\widetilde{X}_t = X_t$, if $t \in [0, 1]\backslash D$, then choose $\{t_n\} \subset D$ such that $t_n \to t$, and let $\widetilde{X}_t = \lim_{n\to\infty} X_{t_n}$ (pointwise limit). Then $\widetilde{X}$ is a modification of X. To see this on $t \notin D$, we have $X_{t_n} \to \widetilde{X}_t$ a.e. but it also converges to X_t in probability, so the two limits actually equal. □

Theorem 2.3.5. *Pre-Brownian motion has a continuous modification, which is Holder continuous with exponent strictly less than one half.*

Proof. Let X be a standard normal random variable, then

$$\mathbb{E}\,|B(t)-B(s)|^p = \mathbb{E}[\left(\sqrt{|t-s|}X\right)^p] = |t-s|^{\frac{p}{2}}\mathbb{E}[X^p].$$

we take $p > 2$. □

Now the following thoerem becomes immediate.

Theorem 2.3.6. *Brownian Motion/Wiener Process exists.*

Remark 2.3.6. *Since Brownian motions are versions of pre-Brownian motions, so they inherit all distributional properties pre-Brownian motions have. In particular, we can view Brownian motion as a continous Gaussian process.*

It is nice that Brownian motion is conitnous, and it is also fortunate (or unfortunate) that it is nowhere differentiable, in fact, it is no where p-Holder continous for $p > \frac{1}{2}$:

Theorem 2.3.7. *Let B be a one-dim standard brownian motion and let $p > \frac{1}{2}$. Then B is nowhere Holder continous of order p, meaning for almost all $\omega \in \Omega$, $B_\cdot(\omega)$ is not Holder continous at any point, and hence nowhere differentiable.*

Proof. Suppose by contradiction that B_t is Holder countinuous on, say (WLOG) the unit interval, with strictly positive probability, then we'd have

$$\mathbb{P}\left[\sup_{x\neq y, x,y\in[0,1]} \frac{|B_x - B_y|}{|x-y|^p} < C\right] \geq \alpha > 0$$

Now, let's derive a contradiction: let $|x-y| = h$, so $B_s - B_y =_d B_h$, so we can only look at the case of Holder continous at the point zero. By the scaling property of BM, we have

$$\mathbb{P}\left[|B_h| \leq Ch^p\right] = \mathbb{P}\left[|B_1| \leq Ch^{p-\frac{1}{2}}\right] \geq \alpha$$

This should hold for all $h > 0$, however, if we take $h \downarrow 0$, we'd see a contradiction. □

Wiener Measures: "second" construction of Brownian motion

Suppose we are wokring with 1-dim process, suppose $\{A_i\}_{1\leq i\leq n} \subset \mathcal{B}(\mathbb{R})$, then we call the following set *Cylindrical*:

$$C = \left\{\omega(\cdot) \in C(\mathbb{R}^+, \mathbb{R}) : \omega(t_i) \in A_i \text{ for all } A_i\right\} \tag{2.7}$$

Given a process and a Cylindrical set as above, then the distribution of the process on the Cylindrical sets is called the finite dimensional distributions. Here we can calculate the finite dimensional distributions for Brownian motions directly:

Lemma 2.3.1. *Let B_t be a one dimensional Brownian motion, and let $0 = t_0 < t_1 < t_2 < \cdots < t_n$, then the finite dimensional distribution of Brownian motion has the following density function:*

$$\mathbb{P}\left[B_{t_1} \in dx_1, \ldots, B_{t_n} \in dx_n\right] = \frac{1}{\sqrt{(2\pi)^n}\sqrt{t_1(t_2-t_1)\ldots(t_n-t_{n-1})}} \exp\left(\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\right) \tag{2.8}$$

Proof. Note that (2.8) is fairly similar to the distribution of $(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n - B_{t_{n-1}}})$, which had the density

$$\mathbb{P}\left[B_{t_1} \in dx_1, \dots, B_{t_n} - B_{t_{n-1}} \in dx_n\right] = \frac{1}{\sqrt{(2\pi)^n}\sqrt{t_1(t_2-t_1)\dots(t_n-t_{n-1})}} \exp\left(\sum_{i=1}^n \frac{(x_i)^2}{2(t_i - t_{i-1})}\right)$$

then let $y_i = \sum_{j=1}^i x_i$ would do the trick, but (2.8) is very intuitive itself. □

We recall the $\lambda - \pi$ theorem (or $\pi - \lambda$) and the monotone class theorem from the very first page of this ntoes: if two measures agrees on a generating algebra (collection of subset that is closed under intersection), then two measures agree on the whole σ field. In other words, a measure is Characterized by its behavior on such an algebra. We note that that the σ field generated by the family of Cylindrical sets is the σ field for which the coordinate mapping:

$$C(\mathbb{R}^+, \mathbb{R}) \ni \omega \to \omega(t) \in \mathbb{R}$$

is continous.

The above means the law of Brownian motion is unique! Say we call it W (Wiener measure) then the underlying probability space becomes $(\Omega, \mathcal{F}, W)$ where $\mathcal{F}$ is the σ-field generated by the Cylindrical sets, then brownian motion is just the coordinate mapping process: $X_t(\omega) = \omega(t)$. We call this space the canonical space for Brownian motion.

Remark 2.3.7. *The above discussion can be treated as a second construction of Brownian motion, where we define a density function like 2.8 on the space of continous functions on the positive part of the real line. Then we do not have to show continuity, and such probability measure (Wiener measure) would give us the distributional proerties of pre-Brownian motion.*

2.3.3 Sample Path Properties of Brownian Motions

For now, let the filtration of Brownian motions be

$$\mathcal{F}_t = \sigma(\{B_s : s \le t\})$$

so the filtration is not neccessarily right continous but left continous, and let

$$\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s.$$

Here we state Blumenthal's zero one law

Theorem 2.3.8 (Blumenthal's). *Let $A \in \mathcal{F}_{0+}$, then $\mathbb{P}[A] \in \{0, 1\}$.*

Proof. The idea is the same as of the proof for Kolmogorov's zero one law (even though we derived it from Levy's zero one law), where we want to shwo $\mathcal{F}_{0+}$ is independent to itself.

The intuition is the following:

$$\mathcal{F}_{0+} = \bigcap_{t>0} \mathcal{F}_t = \bigcap_{t>0} \sigma(B_\tau : \tau \le t) \perp \sigma(B_t - B_\tau : t \ge \tau > 0)\ \forall t > 0$$

but since $X_0 = \lim_{\tau \to 0} X_\tau$, indicates that $\sigma(B_t - B_\tau; t \ge \tau > 0) = \sigma(B_t - B_0; t \ge 0) = \sigma(B_t; t \ge 0) \supset \mathcal{F}_{0+}$ (this is heuristic). Now we formalize this idea:

Let $0 < t_1 < \cdots < t_n$ and and let $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded continous function, consider the following

$$\mathbb{E}[1_A f(B_{t_1}, \ldots, B_{t_n})] = \lim_{\epsilon \downarrow 0} \mathbb{E}[1_A f(B_{t_1} - B_\epsilon, \ldots, B_{t_n} - B_\epsilon)] = \lim_{\epsilon \downarrow 0} \mathbb{P}[A]\mathbb{E}[f(B_{t_1} - B_\epsilon, \ldots, B_{t_n} - B_\epsilon)] = \mathbb{P}[A]\mathbb{E}[1$$

where the first and last equalities are due to continuity (actually, right continuity). So $\mathcal{F}_{0+} \perp \sigma(B_{t_1}, \ldots, B_{t_n})$ for all finitely many t_i's that are strictly positive t_i's, but B_0 is pointwise limit, so $\mathcal{F}_{0+} \perp \sigma(B_t : t \geq 0) \supset \mathcal{F}_{0+}$. □

using exactly the above proof with B_0 replaced by B_s for any $s \geq 0$, we can prove the following theorem:

Theorem 2.3.9. *With above setting, we have $\mathcal{F}_{s+} \perp \sigma(B_t; t \geq s)$.*

More generally, in the above proof, we only used right continuity and independent increments, so we can use the exact same proof again for the following more general theorem:

Theorem 2.3.10. *Let X_t be a stochastic process with independent increments and right continous, then $\mathcal{F}_{s+} \perp \sigma(B_t; t \geq s)$ for $\mathcal{F}_t = \sigma(X_s; s \leq t)$.*

Now let's look at the bahavior of Brownian motion near $t = 0^+$:

Theorem 2.3.11. *Brownian motion changes sign infinitely many times near zero. That is, there is a sequence $t_n \downarrow 0$ such that $\mathbb{P}\left[\#\{B_{t_n} > 0; n \in \mathbb{N}\} = \infty\right] = 1$ and $\mathbb{P}\left[\#\{B_{t_n} < 0; n \in \mathbb{N}\} = \infty\right] = 1$, where # is the counting measure.*

Proof. Let's define

$$A = \{\#\{B_{t_n} > 0; n \in \mathbb{N}\} = \infty\} = \bigcap_{n \geq 1} \bigcup_{m \geq n} \{B_{t_m} > 0\} = \bigcap_{n \geq N} \bigcup_{m \geq n} \{B_{t_m} > 0\} \in \mathcal{F}_{t_N}$$

for arbitrarily large N, so $A \in \bigcap_{t_n} \mathcal{F}_{t_n} = \mathcal{F}_{0+}$, so by the zero-one law, $\mathbb{P}[A] \in \{0, 1\}$. Note we note

$$\bigcap_{n \geq k} \bigcup_{m \geq n} \{B_{t_m} > 0\}$$

is a monotone decreasing sequence of set so

$$\mathbb{P}\left[\bigcap_{n \geq 1} \bigcup_{m \geq n} \{B_{t_m} > 0\}\right] = \lim_{n \to \infty} \mathbb{P}\left[\bigcup_{m \geq n} \{B_{t_m}\} \geq 0\right] \geq \frac{1}{2}.$$

so $\mathbb{P}[A] = 1$. Same for the other one. □

Theorem 2.3.12. *Brownian motion is a Martinagle.*

Proof. This is immediate from one of the remarks before: process with mean zero and independent increments are martingales. But here is a quick computation

$$\mathbb{E}[B_t - B_s | \mathcal{F}_s] = 0.$$

□

Before we talk about other path properties, let's look at some distributional properties of Brownian motion, since they can help with the study of path properties. But first, we need a Strong Law of Large Numbers for Brownian motion

Theorem 2.3.13 (SLLN BM)**.** *Let B_t be a Bronwian motion, then*

$$\lim_{n\to\infty} \frac{B_t}{t} = 0.$$

Proof. Decompose the fraction as the following

$$\begin{aligned}\lim_{t\to\infty} \frac{B_t}{t} &= \lim_{t\to\infty} \frac{B_{\lfloor t\rfloor} + B_t - B_{\lfloor t\rfloor}}{t} \\ &= \lim_{t\to\infty} \frac{1}{t}\left(\sum_{n=1}^{\lfloor t\rfloor} B_n - B_{n-1}\right) + \lim_{t\to\infty} \frac{B_t - B_{\lfloor t\rfloor}}{t}\end{aligned}$$

The first limit converges to zero with no problem (usual SLLN), we have to take care of the second limit. Here again we need to show pointwise convergence with only distributional infomation, so it is natural to use Borel-Cantelli. Note that we are only concered when the integer part of t increases, so let $t_n \in [n, n+1)$ and let $\{\epsilon_n\}$ be a sequence of numbers tends to zero, and the values are to be determined, by Doob's L^p inequality we see

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\max_{t\in[n,n+1)} |B_t - B_n| \geq \epsilon_n n\right] \leq C_p \sum_{n\in\mathbb{N}} \frac{1}{\epsilon_n^2 n^2}$$

so let $\epsilon_n = n^{-\frac{1}{100}}$, then the above sum is finite, so by Borel-Cantelli, we see the last term converges to zero. □

Theorem 2.3.14. *Let $\{B_t\}_t$ be a Wiener process (Brownian motion), then the follows are also Brownian motions:*

1. *For all $\alpha > 0$, $\frac{1}{\alpha}B_{\alpha^2 t}$ is also a Brownian motion.*
2. *Define X_t in the following way, then it is a Bronwian motion*

$$X_t = \begin{cases} tB_{\frac{1}{t}} & t > 0 \\ 0 & t = 0. \end{cases}$$

3. $-B_t$.

Proof. (1) $\frac{1}{\alpha}B_{\alpha^2 t}$ is continous, now we check if it satisfies the distributions:

$$\mathbb{E}[\frac{1}{\alpha}B_{\alpha^2 t}\frac{1}{\alpha}B_{\alpha^2 s}] = \frac{1}{\alpha^2}\alpha^2\,(t\wedge s) = s\wedge t.$$

where the difference is obvious Gaussian.

(2) By the SLLN we see X_t is continous at zero, and

$$\mathbb{E}[(tB_{\frac{1}{t}})^2] = t^2\frac{1}{t} = t.$$

(3) is obvious. □

Remark 2.3.8. *Here we observe that (2) implies SLLN since if X_t is a Brownian motion, then X_t continous at zero, but*

$$\lim_{t\downarrow 0} \frac{B_{\frac{1}{t}}}{\frac{1}{t}} = \lim_{s\to\infty} \frac{B_s}{s} = 0.$$

An immediate and important result from (3) is that Bronwian motion is unbounded a.e.

Theorem 2.3.15. *Let* $T_a = \inf_t\{|B_t| = a\}$*, then* $\mathbb{P}[\tau_a < \infty] = 1$.

Proof. From (3) we see that

$$\begin{aligned}
\mathbb{P}\left[T_a < \infty\right] &= \mathbb{P}[\sup_{0\leq t<\infty} |B_t| > a] \\
&= \mathbb{P}\left[\bigcup_{1\leq n<\infty} \sup_{t\leq n}\{|B_t| \geq a\}\right] \\
&= \lim_{n\to\infty} \mathbb{P}[\sup_{0\leq t\leq n} |B_t| \geq a] \\
&\geq \lim_{n\to\infty} \mathbb{P}[|B_n| \geq a] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}|B_1| \geq a\right] = \lim_{n\to\infty} \mathbb{P}\left[|B_1| \geq \frac{a}{\sqrt{n}}\right] = 1
\end{aligned}$$

where the last equality is achieved by continuity of measrues. □

Here is a similar theorem but says the hitting time of brownian motion at any level is finite:

Theorem 2.3.16. *Let* $\tau_a = \inf_{t\geq 0}\{B_t = a\}$*, then* $\tau_a < \infty$ *a.e.*

Proof. Here it is enough to prove it for $a > 0$, and since Brownian motion changes sign infinite many times near the origin, we have

$$\lim_{\epsilon\downarrow 0} \mathbb{P}\left[\sup_{0\leq t\leq 1} B_t > \epsilon\right] = 1$$

By scaling property we have

$$\lim_{\epsilon\downarrow 0} \mathbb{P}\left[\sup_{0\leq t\leq 1} B_t > \epsilon\right] = \lim_{\epsilon\downarrow 0} \mathbb{P}\left[\sup_{0\leq t\leq 1} \frac{1}{\epsilon} B_t > 1\right] = \lim_{\epsilon\downarrow 0} \mathbb{P}\left[\sup_{0\leq t\leq \frac{1}{\epsilon^2}} \frac{1}{\epsilon} B_{\epsilon^2 t} \geq 1\right] = \lim_{\epsilon\downarrow 0} \mathbb{P}\left[\sup_{0\leq t\leq \frac{1}{\epsilon^2}} B_t \geq 1\right]$$

the above expression still equals to 1. So here we showed $\tau_1 < \infty$ a.e. For other numbers, we can use a different scaling. □

The following properties of Brownian motions are entirely trivial from above theorems:

Corollary 2.3.1.
1. *Brownian motion is nowhere monotone.*
2. *Has quadratic variation* $\langle B\rangle_t = t$.
3. *Has infinite variation on any nontrivial interval.*

Strong Markov Property

We've talked about (simple) Markov property very briefly in the discrete martingale section, here is the definition of Strong Markov Property:

Theorem 2.3.17 (Strong Markov for Brownian Motion). *Let τ be a stopping time with $\mathbb{P}[\tau < \infty] > 0$ (infinity is allowed), then the process*

$$B_t^\tau \triangleq 1_{\{\tau<\infty\}} \left(B_{\tau+t} - B_\tau\right)$$

is a Bronwian motion under the probability measure $\mathbb{P}\left(\cdot|\tau < \infty\right)$ which is independent of $\mathcal{F}_\tau$.

Remark 2.3.9. *Just as the proof of the zero one law, we show independence via finite dimensional distributions. Also, we don't know anything about Brownian motion at stopping time, but we can approximate the stopping times with discrete stopping times. In such case, for example, say τ is discrete, and on the set $\{\tau = n\}$, we can use B_n instead of B_τ.*

Proof. Let approximate it with discrete stopping times that decreases to τ:

$$\tau_n(\omega) \triangleq \sum_{k=0}^{\infty} 1_{\{k/2^n\le\tau<(k+1)/2^n\}}(\omega)\frac{k+1}{2^n} + \infty 1_{\{\tau=\infty\}} \tag{2.9}$$

Let $A \in \mathcal{F}_\tau$, so by definition, $A \cap \{\tau \le t\} \in \mathcal{F}_t$, and let $A_k = A \cap \left\{\tau_n = \frac{k}{2^n}\right\} \in \mathcal{F}_{\frac{k}{2^n}}$, and we note that $A_k \cap \{\tau \ne \infty\} = A_k$, and let f be a bounded continous function, then

$$\begin{aligned}\int_{A\cap\{\tau\ne\infty\}} f\left(B_{\tau_n+t} - B_{\tau_n}\right) &= \sum_{k\in\mathbb{N}} \int_{A_k\cap\{\tau\ne\infty\}} f\left(B_{\tau_n+t} - B_{\tau_n}\right)\\ &= \sum_{k\in\mathbb{N}} \int_{A_k} f\left(B_{\frac{k}{2^n}+t} - B_{\frac{k}{2^n}}\right)\\ &= \sum_{k\in\mathbb{N}} \mathbb{P}[A_k] \int f\left(B_{\frac{k}{2^n}+t} - B_{\frac{k}{2^n}}\right)\\ &= \sum_{k\in\mathbb{N}} \mathbb{P}[A_k\cap\{\tau\ne\infty\}]\mathbb{E}\left[f(B_t)\right]\\ &= \mathbb{P}[A\cap\{\tau\ne\infty\}]\mathbb{E}[f(B_t)]\end{aligned}$$

take $n \to \infty$ and by dominated convergence theorem we have

$$\int_A f\left(B_{\tau+t} - B_\tau\right) = \mathbb{P}[A]\mathbb{E}[f\left(B_t\right)].$$

In particular, when $A = \Omega$ this tells us $B_{\tau+t} - B_\tau =_d B_t$. So both independence and identical distribution are proven. □

Next we have reflection principle, and there are many versions of it that all tells the same thing:

Theorem 2.3.18 (Reflection Principle). *Let B be a Brownain motion and τ be an stopping time. Define*

$$\widetilde{B}_t \triangleq 2B_{t\wedge\tau} - B_t,$$

then $\widetilde{B}_t =_d B_t$.

Proof. By the Strong Markov property, we have $B_t^\tau \triangleq B_{t+\tau} - B_\tau$ is a standard Brownian motion, hence so is $-B_t^\tau$. Now we make the observation (or clever change of notation) that

$$\widetilde{B}_t = B_{t\wedge\tau} + B_{(t-\tau)^++\tau} - B_\tau = B_{t\wedge\tau} - B^\tau_{(t-\tau)^+}$$

since when $t \le \tau$, above is $B_t^\tau = B_{t\wedge\tau} - B_{(\tau-t)^++t} - B_\tau = B_t - B_t - B_\tau = B_\tau$ and when $t \ge \tau$ we get $B_t^\tau = B_\tau + B_t - B_\tau = B_t$, same as $\widetilde{B}_t$. But we can write B_t as

$$B_t = B_{t\wedge\tau} - \left(B_{(t-\tau)^++\tau} - B_\tau\right) = B_{t\wedge\tau} + B^\tau_{(t-\tau)^+} \tag{2.10}$$

So they are equal in distirbution due to the fact that $-B^\tau_{(t-\tau)^+} =_d B^\tau_{(t-\tau)^+}$. □

Perhaps the proof to this version does not give insight to the reflection principle, so here is an other version:

Theorem 2.3.19 (Reflection Principle 2). *Let $\tau_a = \inf\{t \geq 0 : B_t = a\}$ where $\{B_t\}$ is an Brownian motion, then*

$$\mathbb{P}\left[\tau_a \leq t, B_t \leq b\right] = \mathbb{P}\left[B_t \geq 2a - b\right]$$

for $b \leq a < \infty$.

Proof. With above setup, do a direct computation

$$\begin{aligned}\mathbb{P}\left[\tau_a \leq t, B_t \leq b\right] &= \mathbb{P}\left[\tau_a \leq t, B_t - a \leq b - a\right] \\ &= \mathbb{P}\left[\tau_a \leq t, B_t - B_{\tau_a} \leq b - a\right] \\ &= \mathbb{P}\left[\tau_a \leq t, B_{(t-\tau_a)+\tau_a} - B_{\tau_a} \leq b - a\right]\end{aligned}$$

However, $W_t = B_{s+\tau_a} - B_{\tau_a}$ itself is a Brownian motion and it is independent of $\mathcal{F}_{\tau_a}$ by the Strong Markov Property about, hence

$$\begin{aligned}\mathbb{P}\left[\tau_a \leq t, B_{(t-\tau_a)+\tau_a} - B_{\tau_a} \leq b - a\right] &= \mathbb{P}\left[\tau_a \leq t, -B_{(t-\tau_a)+\tau_a} + B_{\tau_a} \leq b - a\right] \\ &= \mathbb{P}\left[\tau_a \leq t, B_t \geq 2a - b\right] \\ &= \mathbb{P}\left[B_t \geq 2a - b\right]\end{aligned}$$

where the last equality is due to the fact $\{B_t \geq 2a - b\} \supset \{\tau_a \leq t\}$ since $2a - b \geq a$ because $b \leq a$. □

Remark 2.3.10. *If one is unsatisfied with the use of Strong Markov Property (like I am), we can get around with that and use simply intigration: Let $\tau_n = \sum_{k=1}^{\infty} \frac{k}{2^n} 1_{\frac{k-1}{2^n} \leq \tau < \frac{k}{2^n}}(\omega) + \infty 1_{\tau=\infty}(\omega)$, and directly compute $\mathbb{P}\left[\tau_a \leq t, B_t - B_{\tau_a} \geq b - a\right]$, let $t_n = \max\{k : \frac{k}{2^n} \leq t\}$*

$$\begin{aligned}\mathbb{P}\left[\tau_a \leq t, B_t - B_{\tau_a} \geq b - a\right] &= \lim_{n\to\infty} \sum_{k=1}^{t_n} \mathbb{P}\left[B_t - B_{\tau_n} \geq b - a, \tau_n = \frac{k}{2^n}\right] \\ &= \lim_{n\to\infty} \sum_{k=1}^{t_n} \mathbb{P}\left[B_t - B_{\frac{k}{2^n}} \geq b - a\right] \mathbb{P}\left[\tau_n = \frac{k}{2^n}\right] \\ &= \lim_{n\to\infty} \sum_{k=1}^{\infty} \mathbb{P}\left[-B_t \geq b - a - B_{\tau_n}, \tau \leq t\right] \mathbb{P}\left[\frac{k-1}{2^n} \leq \tau < \frac{k}{2^n}\right]\end{aligned}$$

take the limit $n \to \infty$, by DCT we see the last expression is

$$\int_0^\infty \mathbb{P}\left[\tau \leq t, -B_t \geq b - 2a\right] d\mathbb{P}\left[\tau \leq s\right] = \mathbb{P}\left[B_t \leq 2a - b, \tau \leq t\right]$$

Remark 2.3.11. *A particular case of above is when $a = b$, here we have $\mathbb{P}\left[\tau_a \leq t, B_t \leq a\right] = \mathbb{P}\left[B_t \geq a\right]$. However,*

$$\begin{aligned}\mathbb{P}\left[\max_{0\leq s\leq t} B_s \geq a\right] &= \mathbb{P}\left[\max_{0\leq s\leq t} B_s \geq a, B_t > a\right] + \mathbb{P}\left[\max_{0\leq s\leq t} B_s \geq a, B_t < a\right] \\ &= 2\mathbb{P}\left[\max_{0\leq s\leq t} B_s \geq a, B_t < a\right] \\ &= 2\mathbb{P}\left[B_t \geq a\right] \\ &= \mathbb{P}\left[B_t \geq a\right] + \mathbb{P}[B_t \leq a] = \mathbb{P}[|B_t| \geq a]\end{aligned}$$

where the last equality is due to "union of disjoint sets".

We orgnize the above into a theorem which is also called the Reflection Principle

Theorem 2.3.20 (Reflection Principle 3)**.** *Let B_t be a Brownian motion, then $\max_{0\le s\le t} B_t =_d |B_t|$.*

Remark 2.3.12. *We'll see by Wald's identity that expectation of hitting time of Brownian motion is infinite.*

Having the reflection principle on hand, we can calculate the distribution of Brownain stopping time:

Lemma 2.3.2. *Let B be a Brownain motion and $a \ge 0$, and let $b \le a$, then the joint density function of $(\max_{0\le s\le t} B_s, B_t)$ is*

$$g(a,b) = \frac{\sqrt{2}(2a-b)}{\sqrt{\pi t^3}} \exp\left(-\frac{(2a-b)^2}{2t}\right)$$

Proof. Let $M_t = \max_{0\le s\le t} B_t$, then $\mathbb{P}\left[M_t \ge a, B_t \le b\right] = \mathbb{P}\left[B_t \ge 2a-b\right]$, then the joint density is

$$\begin{aligned} g(a,b) &= \frac{\partial^2}{\partial b \partial a}\left(1 - \frac{1}{\sqrt{2\pi t}}\int_{2a-b}^{\infty} \exp\left(-\frac{x^2}{2t}\right)dx\right) \\ &= \frac{\partial}{\partial b}\sqrt{\frac{2}{\pi t}} \exp\left(-\frac{(2a-b)^2}{2t}\right) \\ &= \frac{\sqrt{2}(2a-b)}{\sqrt{\pi t^3}} \exp\left(-\frac{(2a-b)^2}{2t}\right) \end{aligned}$$

□

From Remark 55, the next lemma is immediate

Lemma 2.3.3. *Let B be a BM and $\tau_a = \inf\{t \ge 0 : B_t = a\}$, then it has the distribution*

$$f_{\tau_a}(t) = \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2t}\right)$$

Proof.

$$\mathbb{P}\left[\tau_a \le t\right] = \mathbb{P}\left[\max_{0\le s\le t} B_t \ge a\right] = 2\mathbb{P}\left[B_t \ge a\right] = 2\mathbb{P}\left[B_1 \ge \frac{a}{\sqrt{t}}\right]$$

so

$$f_{\tau_a}(t) = \frac{\partial}{\partial t}\mathbb{P}\left[\tau_a \le t\right] = \frac{\partial}{\partial t}\left(1 - \int_{\frac{a}{\sqrt{t}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)dx\right) = \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2t}\right).$$

□

Wald's Identity for Brownain Motion

Here is an important result coming from optional sampling theorem for Brownain motion:

Theorem 2.3.21 (Wald's Identity for BM)**.** *Let B be a Brownain motion and τ is a stopping time, the if either*

1. $\mathbb{E}[\tau] < \infty$ *or*

2. *B^τ is uniformly integrable,*

Then we have $\mathbb{E}[B_\tau] = 0$ and if (1) holds, then we have $\mathbb{E}[B_\tau^2] = \mathbb{E}[\tau]$.

Proof. We first do (2): $B_t^\tau = B^{\tau\wedge t}$ itself forms a true martingale, so unifromly integrability implies a.e. convergence (Doob's mtg convergence theorem), moreover, it converges in L^1 as well, so $\lim_{n\to\infty} \mathbb{E}[B_{\tau\wedge n} = \mathbb{E}[B_\tau] = 0$.

For (1) we show (1) $\Rightarrow$ (2): let's consider the quadratic variation of B^τ:

$$\langle B^\tau \rangle_t = \langle B \rangle_{t\wedge\tau} = t \wedge \tau$$

Note that $\langle B^\tau \rangle_t$ is dominated by τ which is in L^1, we have L^1 convergence:

$$\lim_{t\to\infty} \mathbb{E}\left[\langle B^\tau \rangle_t\right] = \mathbb{E}\left[\lim_{t\to\infty} t \wedge \tau\right] = \mathbb{E}[\tau] < \infty$$

hence B^τ is uniformly integrable mtg and the above calculation also shows that $\mathbb{E}[B_\tau^2] = \mathbb{E}[\tau]$ under (1). □

Zero Set of Brownian Motion

Let's define the pathwise zero set of Brownain motion to be

$$\mathcal{L}_\omega \triangleq \{t \in \mathbb{R}^+ : B_t(\omega) = 0\}$$

Remark 2.3.13. *Here we study the zero sets, but you can actually change the 0 in above to any number and the following result would still be true.*

Theorem 2.3.22. *For $\mathbb{P}$-a.e. the zero set $\mathcal{L}_\omega$:*

1. *Has lebesgues' measure zero.*

2. *is closed and unbounded.*

Proof. (1) Here we note lebesgues' measure of the zero set is nonnegative, and if the expectation of a nonnegative function is zero, then the function itself is zero, and we will use this and Fubini to prove (1):

$$\mathbb{E}[\lambda(\{t \in \mathbb{R}^+ : B_t = 0\})] = \mathbb{E}\left[\int_0^\infty 1_{\{B_t=0\}}(t)\right] = \int_0^\infty \mathbb{E}\left[1_{\{B_t=0\}}(t)\right] dt = \int_0^\infty \mathbb{P}\left[B_t = 0\right] dt = 0.$$

The closedness is due to the fact that $\mathcal{L}_\omega = B_\cdot(\omega)^{-1}(\{0\})$ and Brownian motion sample pathes are continous. The unboundedness comes from the time inversion property, namely,

$$\widetilde{B}_t = \begin{cases} tB_{\frac{1}{t}} & t \neq 0 \\ 0 & t = 0 \end{cases}$$

is an Brownian motion, and since the Brownain motion hits zero infinitely many times near $t = 0$, hence $B_{\frac{1}{t}}$ hits zero infinitely many times near $t = 0$, that is, B_t hits zero infinitely many times near infinity. □

The Law of Iterated Logarithm

We first recall the upper and lower bound of the tail probability of standard normal distribution:

Lemma 2.3.4.

$$\frac{u}{u^2+1}e^{-\frac{u^2}{2}} \leq \int_u^\infty e^{\frac{-x^2}{2}}dx \leq u^{-1}e^{-\frac{u^2}{2}} \tag{2.11}$$

Theorem 2.3.23 (The Law of Iterated Logarithm). *Let B be a Brwonain motion, then*

$$\overline{\lim}_{t\to\infty}\frac{B_t}{\sqrt{2\log\log\frac{1}{t}}} = \overline{\lim}_{t\to\infty}\frac{B_t}{\sqrt{2\log\log t}} = 1$$

Remark 2.3.14. *One can see that this can also be used to prove $\tau_a < \infty$ a.e.. Also, we only need to prove one of them by $tB_{\frac{1}{t}}$ is also a Brwonain motion.*

The following proof is based on [KK97].

Remark 2.3.15 (Strategy of the Proof). *Here again, we are trying to get a.e. and limiting result from distributional property, so Borel-Cantelli is likely to kick in. Also, one can observe that it is hard to prove this directly without some "time change", since double logarithm is hard to work with. However, one can use r^n for $r > 1$ and $n \to \infty$ instead of $t \to \infty$, since $\log\log r^n = \log(n\log(r))$ which is easier to work with. Here we show the limit sup both less than and greater than 1.*

Proof. In this proof, we adapt to the Harmonic anlysis' notation, using the same letter C for different constants. First we show

$$\mathbb{P}\left[\overline{\lim}_{t\to\infty}\frac{B_t}{\sqrt{2\log\log t}} > c\right] = 0$$

for all $c > 1$. We first recall that the running max of the Brownian motion has the same distribution as the absolute value of Brownian motion: $\mathbb{P}[\max_{0\leq s\leq t} B_t \geq a] = \mathbb{P}[|B_t| \geq a] = 2\mathbb{P}[B_t \geq a]$. So, from the estimate for tail of normal distirbutions, we have

$$\mathbb{P}\left[\max_{0\leq s\leq t} B_s \geq u\sqrt{t}\right] = 2\mathbb{P}\left[B_t \geq u\sqrt{t}\right] = 2\mathbb{P}\left[\frac{1}{\sqrt{t}}B_t \geq u\right] = 2\mathbb{P}[B_1 \geq u] \leq Cu^{-1}e^{-\frac{u^2}{2}}$$

Let $r > 1$, and let $c > 0$ to be chosen later, and let $h(t) = \sqrt{2t\log\log t}$, and denote the running max to be $M(t)$ and do a direct calculation

$$\begin{aligned}\mathbb{P}[M(r^n) \geq ch(r^n)] &= 2\mathbb{P}\left[M(r^n) \geq \sqrt{r^n}\left(c\cdot\sqrt{2\log\log r^n}\right)\right] && (2.12)\\ &\leq C\frac{1}{c\sqrt{2\log(n\log r)}}\exp\left(-c^2\log(n\log r)\right) && (2.13)\\ &\leq C(\log n)^{-\frac{1}{2}}n^{-c^2} && (2.14)\end{aligned}$$

where the last constant C only depens on c, r. Note that the last expression is summable for any $c > 1$, hence

$$\mathbb{P}[M(r^n) \geq h(r^n) \text{ infinitely often}] = 0$$

which proves the case for upper bound.

Before we go into the lower bound, we'll need the case for which we change $h(r^n)$ to $h(r^{n-1})$ above, we will see in the end why (one comes back and add something to the proof becasue it is needed later, this is normal):

$$\begin{aligned}
\mathbb{P}\left[M(r^n) \geq ch(r^{n+1})\right] &= 2\mathbb{P}\left[B(r^n) \geq \sqrt{r^n}\left(cr^{\frac{1}{2}} \cdot \sqrt{2\log\log r^{n+1}}\right)\right] \\
&\leq C\frac{1}{c\sqrt{2\log\left(n\log r\right)}}\exp\left(-c^2 r\log\left(n+1\right)\log r\right)\right) \\
&\leq C\left(\log n\right)^{-\frac{1}{2}} n^{-c^2 r}
\end{aligned}$$

here we choose $c > \frac{1}{\sqrt{r}}$ then it is also summable. Hence $\mathbb{P}\left[\overline{\lim}_{n\to\infty}\frac{M(r^n)}{h(r^{n-1})} > c\right] = 0$ for all $c > 1$, so $\overline{\lim}_{n\to\infty}\frac{B(r^n)}{h(r^{n-1})} \leq 1/\sqrt{r}$. I'll refer this result as the "comeback" later.

For the lower bound, we use the lower estimate from the tail of standard normal random variables to bound the probability below, but we need to bound the probability of the increment below since we want to show the maximum greater than $h(t)$ infinitely many often with probability one and we can only get that from the second Borel-Cantelli.

$$\begin{aligned}
\mathbb{P}\left[B_{r^n} - B_{r^{n-1}} \geq ch(r^n)\right] &= \mathbb{P}\left[\left(r^{n-1}\right)^{-\frac{1}{2}} B_{r^{n-1}(r-1)} \geq c\sqrt{r}\sqrt{2\log\left(n\log r\right)}\right] \\
&= \mathbb{P}\left[B_{r-1} \geq c\sqrt{2r\log\left(n\log r\right)}\right] \\
&= \mathbb{P}\left[\frac{B_{r-1}}{\sqrt{r-1}} \geq c\sqrt{\frac{2r\log\left(n\log r\right)}{r-1}}\right] \\
&= \mathbb{P}\left[B_1 \geq c\sqrt{\frac{r\log\left(n\log r\right)}{r-1}}\right] \\
&\geq \frac{c\sqrt{r\frac{2\log(n\log r)}{r-1}}}{c^2\frac{r\log(n\log r)}{r-1}+1}\exp\left(-\frac{c^2 r\log\left(n\log r\right)}{r-1}\right) \\
&= C\left(\log n\right)^{-\frac{1}{2}} n^{-\frac{c^2 r}{r-1}}.
\end{aligned}$$

We let $0 < c \leq \sqrt{\frac{r-1}{r}}$ to make it not summable, and note that for each n, the increment in the beginning are independent, so by Borel-Cantelli, we have

$$\overline{\lim}_{n\to\infty}\frac{B(r^n - B(r^{n-1}))}{h(r^n)} \geq \frac{\sqrt{r}}{\sqrt{r-1}}.$$

However, by the previous "comeback" we see $\overline{\lim}_{n\to\infty}\frac{-B(r^{n-1})}{h(r^n)} \leq 1/\sqrt{r}$. Combining the results we have

$$\overline{\lim}_{n\to\infty}\frac{B(r^n)}{h(r^n)} \geq \sqrt{\frac{r}{r-1}} + \frac{1}{\sqrt{r}}$$

take r to be large to see the limsup is at least 1. □

Brownain Motion as a Martingale

Finally, as we noted before, Brownian motion is a Martingale. It has quadratic variation t:

Theorem 2.3.24. *Let B_t be a standard Brownain motion, then $\langle B \rangle_t = t$ a.e. and if Π_n is a sequence of partition of $[0,t]$ with $\lim_{n\to\infty} \|\Pi_n\| = 0$, then*

$$\lim_{n\to\infty} \sum_{\Pi_n} \left(B_{t_{n+1}} - B_{t_n}\right)^2 \to t \quad \textit{in probability (measure).}$$

Proof. As we noted before, if a process B_t has independent increments and has mean zero, then $B^2 - \mathbb{E}\left[B_t^2\right]$ is a martingale. So in the case of Brownian motion, $\mathbb{E}\left[B_t^2\right] = t$ and by the uniqueness of Doob Meyer's decompsoition, we see t is the quadratic variation. SO by similar theorem in the continous Mtg theory, the limit converges to its quadratic variation in probability.

However, we did not prove either of the theorem used in this proof, so we might as well prove this special case here, we note by independent increment, we see that it converges to t in L^2:

$$\int_\Omega \left(\sum_{\Pi_n} (B_{t_n+1} - B_{t_n})^2 - (t_{n+1} - t_n)\right)^2 d\mathbb{P} \le \sum_{m,n} \mathbb{E}\left[(B_{t_{n+1}} - B_{t_n})^2 (B_{t_{m+1}} - B_{t_m})^2\right]$$
$$-2t \sum_{\Pi_n} \mathbb{E}\left[\left(B_{t_{n+1}} - B_{t_n}\right)^2\right] + t^2$$

where the second term is $-2t^2$, so we need to show the first term converges to t^2 as well:

$$\sum_{m,n} \mathbb{E}\left[(B_{t_{n+1}} - B_{t_n})^2 (B_{t_{m+1}} - B_{t_m})^2\right] = \sum_{m\neq n} \mathbb{E}\left[(B_{t_{n+1}} - B_{t_n})^2 (B_{t_{m+1}} - B_{t_m})^2\right] + \sum_{\Pi_n} \mathbb{E}\left[\left(B_{t_{n+1}} - B_{t_n}\right)^4\right]$$

The second term on the right hand side goes to zero, one can either check this directly, or recall if p-variation exists and strictly positive, then $p+\epsilon$-variation goes to zero. Now for the first term, by independent increment, one has

$$\sum_{m\neq n} \mathbb{E}\left[(B_{t_{n+1}} - B_{t_n})^2 (B_{t_{m+1}} - B_{t_m})^2\right] = \sum_m (t_{m+1} - t_m)\left(t - (t_{m+1} - t_m)\right)$$
$$= t^2 - \sum_{\Pi_n} (t_{m+1} - t_m)^2$$

and we see the second term goes to zero, hence we have the desried result. □

Donsker's Invariance Principle (Functional CLT)

Here we have two important theorems, both has names. One tells us that any square integrable random variable can be viewed as a sampling of Brownian motion at a certain stopping time (that is amazing!) and the other tells use that random walks (after rescaling) converges to Brownian motion in functional space $C([0,T])$, the sapce of continuous functions on compact interval with uniform convergence topologies.

For the first embeding theorem we want the targetting X to be centered ($\mathbb{E}(X) = 0$) where we can use optional sampling thoerem from martinagle thoery. However, we immediately run into difficulties since there is no tool for us to construct a such a stopping time directly, so we have to look at the simple cases and see what we can do with them.

We first note that if X is a random variable that only takes $\{a, b\} \subset \mathbb{R}$ with $a < 0 < b$, then it is easy to find a stopping time such that $B_\tau =_d X$. $\tau == \inf\{t \geq 0 : B_t \in \{a, b\}\}$ is a promising candidate, since for a centered X whose distribution is supported on $\{a, b\}$, we must have

$$\begin{cases} \mathbb{P}[X = a] + \mathbb{P}[X = b] = 1 \\ a\mathbb{P}[X = a] + b\mathbb{P}[X = b] = 0 \end{cases}$$

and there is only one solution to this, namely, $\mathbb{P}[X = a] = \frac{-a}{b-a}$ and $\mathbb{P}[X = b] = \frac{b}{b-a}$. It is easy to see that B_τ has the same distribution since it must satisfy the same set of eqautions.

Since our goals is to write $X = B_\tau$ for some stopping time τ for any centered square integrable random variable, it would be nice if we can decompose X in some way into random variables that is supported on only two points on the real line. There is a way, and there is a spectial term for such:

Definition 2.3.9. *We say $\{X_n, \mathcal{F}_n\}$ is a binary splitting martingale if X_{n+1} conditioned on the event $\{X_0 = x_0, \ldots, X_n = x_n\}$, when it has nonzero probability, is supported on at most two points of the real line.*

The following lemma gives us a decompsoition of general X into a limit of binary splitting martingale:

Lemma 2.3.5. *Let X be a square integrable martingale, then there exists a binary splitting martingale $\{X_n, \mathcal{F}_n\}$ such that $X_n \to X$ both in L^2 and a.e..*

Proof. Here we can construct such martingale explicitely: Let $X_0 = \mathbb{E}[X]$ and define ξ_i recurssive by

$$\xi_n = \begin{cases} 1 & X \geq X_n \\ -1 & X < X_n \end{cases}$$

then we define $X_n \triangleq \mathbb{E}[X|\sigma(\xi_0, \ldots, \xi_{n-1})]$, which is obviously a martingale. Also note that $\sigma(\xi_0)$ splits the underlying probability space into two parts: $\{X \geq \mathbb{E}[X]\}$ and $\{X < \mathbb{E}[X]\}$, and $\sigma(\xi_0, \ldots, \xi_n)$ splits each partition of $\sigma(\xi_0, \ldots, \xi_{n-1})$ into two parts. So $\sigma(\xi_0, \ldots, \xi_{n-1})$ splits Ω into 2^n partitions and each partition is of the form $\{X_0 = x_1, \ldots, X_{n-1} = x_{n-1}\}$ therefore it is a binary splitting martingale.

We note $|\mathbb{E}[X|\mathcal{F}]| \leq \mathbb{E}[|X||\mathcal{F}]$, so $\{X_n\}$ is uniformly integrable and bounded in L^1, so by Doob's convergence and the theorem quickly following that, we have $X_n \to \mathbb{E}[X|\mathcal{F}_\infty]$ where $\mathcal{F}_\infty \triangleq \sigma(\xi_0, \ldots, \xi_n, \ldots)$.

Now our takes is to show $X_\infty = X$ a.e. which can be acheived by showing $\mathbb{E}(|X_\infty - X|) = 0$. However, we note that

$$\lim_{n\to\infty} \xi_n (X_{n+1} - X) = |X_\infty - X|$$

since on the set $X_\infty = X$, the above is true by simply a.e. convergence, on the set $\{X_\infty > X\}$, then for fixed ω, for large enough n, we have $X_n > X$, so $\xi_n = 1$ so the left is $X_\infty - X$, and on the set $\{X_\infty < X\}$ we have for all ω, there is $n(\omega)$ with all $n \geq n(\omega)$ that $X_n > X$, so $\xi_n = -1$ and the above holds true as well.

Also, since the left hand side is bounded by $2|X|$, by DCT we have L^1 convergence, so

$$\mathbb{E}[|X_\infty - X|] = \lim_{n\to\infty} \mathbb{E}[\xi_n(X_{n+1} - X)] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[\xi_n(X_{n+1} - X)] | \mathcal{F}_{n+1}] = 0$$

□

Now we are ready for the first theorem:

Theorem 2.3.25 (Skorokod's embeding theorem). *Let $X \in L^2(\Omega)$ be a centered random variable and let B be a Bronwian motion, then there is a stopping time τ such that $B_\tau = X$ and $\mathbb{E}[\tau] = \mathbb{E}[B_\tau^2] = \mathbb{E}[X^2]$.*

Proof. Let $X_n \to X$ be a binary splitting martingale. Define $\tau_1 = \inf\{t : B_t = X_1\}$, looking back at the construction of such binary splitting martiangle, when $\mathbb{E}[X] = 0$, then the support of X is $\{a, b\}$ where one is positive and the other is negative, so the definiton of τ_1 makes sense. Now we can utilize the strong markov property and define τ_n with similar method with a Brownian motion starting at τ_{n-1} (partition Ω into $\tau_{n=1} = x_1^{n-1}, \dots, x_k^{n-1}$ and define τ_n on those partitions seperately). So in the end we get a sequence of increasing stopping times $\tau_n \uparrow \tau$ for some τ. However, we do have $\mathbb{E}[\tau_n] = \mathbb{E}[X_n^2]$, so by dominated convergence theorem on the left and previous lemma on the right, we have the desired convergence result, namely,

$$\mathbb{E}\tau = \mathbb{E}X^2.$$

□

The object simple random walk is defined as follows: let ξ_i be mean zero and $\mathbb{E}\xi^2 = 1$ i.i.d random variables, and let $S_n = \sum_{i=1}^n \xi_i$ and let S^* be the linear interpolation of S, namely,

$$S^*(t) = S_{[t]} + (t - [t]) \left(S_{[t+1]} - S_{[t]} \right) = S_{[t]} + (t - [t])\xi_{[t]+1}$$

where $[t]$ is the integer part of t.

We now extend the Skorokod's theorem to a sequence of random variables

Proposition 2.3.7. *Let S_n be defined as above, and let B be a Brownain motion, then there is $\{\tau_i\}$ such that $S_1, S_2, \dots =_d B(\tau_1), B(\tau_2), \dots$.*

Proof. Here we let τ_1 be the stopping time such that $B_{\tau_1} = \xi_1$ as provided by Skorokod's embeding thoerem. For τ_2, we know that $B_t^1 = B_{t+\tau_1} - B_{\tau_1}$ is a Brownian motion by Strong Markov Property, and it is $\perp$ to $\mathcal{F}_{\tau_1}$. So we define τ_2 be the stopping time such that $B_{\tau_2}^1 =_d \xi_2 = \xi_1$.

It is easy to see that $\tau_1 \perp \tau_2$ and $\tau_1 =_d \tau_2$ and $\mathbb{E}[\tau_i] = 1$. Also, $B_{\tau_2}^1 + B_{\tau_1} =_d \xi_1 + \xi_2 = B_{\tau_1+\tau_2}$. Now we can define τ_n inductively for $n \geq 3$. □

Then we are ready to build the functional CLT, and the first step is to embed the scaled random walk into a Brownain motion, and the following theorem tells us exactly how to do that:

Theorem 2.3.26. *Let $S_n = \sum_{i=1}^n \xi_i$, where ξ_i are i.i.d with mean zero and $\mathbb{E}\xi^2 = 1$, then there exists triangular array $\{\tau_i^n\}_{i=1}^n$ such that*

1. *$\mathbb{E}[\tau_i^n] = 1$ for all n, i,*
2. *for each n, $\{\tau_i^n\}_{1 \leq i \leq n}$ are i.i.d.,*
3. *the following two maps have the same distribution*

$$\left\{ \frac{S_k}{\sqrt{n}}; 1 \leq k \leq n \right\} \quad \text{and} \quad \left\{ B\left(\frac{\tau_1^n + \cdots + \tau_k^n}{n} \right); 1 \leq k \leq n \right\}$$

Remark 2.3.16. *From first glance, we would want to use one of the scaling property to move $\frac{1}{\sqrt{n}}$ term inside of the time component of the Brownain motion. But since τ_i's are random times, there is no theorem tells us we can do that, so we have to take a slightly longer route.*

Proof. Let $B_n = \sqrt{n}B\left(\frac{t}{n}\right)$ which is still a Brownain motion and is of the form of the right hand side of (3). Now apply the previous theorem to S_k for $k = 1, \dots, n$ and get a "row" of stopping times $\tau_1^n, \dots, \tau_n^n$ so that

$$\frac{1}{\sqrt{n}} B_n \left(\frac{\tau_1^n + \cdots + \tau_k^n}{n} \right) =_d \frac{S_k}{n} \quad \forall 1 \le k \le n$$

but the left hand side, by construction, is $B\left(\frac{\tau_1^n + \cdots + \tau_k^n}{n}\right)$.

(1) and (2) is a result of the construction by the previous theorem. □

Remark 2.3.17. *We also note that since τ_i^n and τ_i^m are both constructed from Brownian motions in the exactly the same way, so they are equal in distribution but not neccessarily independent. So there is a sequence of stopping times $\{\tau_i\}$ for which $\tau_i = \tau_i^n$ for all n, and τ_i's are independent.*

Now we are ready to state and prove the big theorem of this subsection

Theorem 2.3.27 (Donsker's Invariance Principle)**.** *Let $S_n = \sum_{i=1}^n \xi_i$ where $\{\xi_i\}$ is a sequence of i.i.d. where with mean zero and $\mathbb{E}[\xi_i^2] = 1$, then*

$$\frac{S_{[tn]}}{\sqrt{n}} \to B_t \quad \textit{in distribution}$$

in $C[0,1]$ with uniform convergence topology.

Remark 2.3.18. *Here, it is pretty clear on what to do, the previous theorem provided us something that is easier to work with with sup norm, so we can turn the rescaled partial sums into Brownain motions. Since Bronwian motion is continous, so uniformly continous, we might as well try to show convergence in probability instead of in distribution.*

Proof. Since $\frac{S_{[tn]}}{\sqrt{n}} =_d Y_n(t) \triangleq B\left(\frac{\tau_1^n + \cdots + \tau_{[tn]}^n}{n}\right)$, we can show $Y_n(t) \Rightarrow B(t)$ in $C([0,1])$ in probability instead. Also, by the previous remark, we can replace τ_i^n's by simply τ_i, i.i.d. Finally, let $T_n(t) = \frac{\tau_1 + \cdots + \tau_{[nt]}}{n}$. To show the convergence in probability in the space of continous functions on the unit interval is really to control the following:

$$\mathbb{P}\left[\sup_{0 \le t \le 1} |Y_n(t) - B(t)| \ge \epsilon\right] = \mathbb{P}\left[\sup_{0 \le t \le 1} |B(T_n(t)) - B(t)| \ge \epsilon\right]$$

Let since each Brownain path is uniformly continous on the unit inteval, so for all $\omega \in \Omega$ (we get rid of the null set for which Brownian path is not uniformly conitnous), and for all $\epsilon > 0$, there is $\delta(\omega)$ with $|B_t(\omega) - B_s(\omega)| < \epsilon$ for all $|t - s| < \delta(\omega)$. For notations, we denote $B_\omega(t)$ by the Brownian path ω at time t. Then the above probability on the right hand side is really dominated by

$$\mathbb{P}\left[\omega \in \Omega : \sup_{0 \le t \le 1} |T_n(t)(\omega) - t| \ge \delta(\omega)\right]$$

so if we can show $\sup_{t \in [0,1]} |T_n(t) - t| \to 0$, then we are done, but this is really

$$\begin{aligned} \sup_{t \in [0,1]} \left| \frac{\sum_{i=1}^{[tn]} \tau_i}{n} - t \right| &\le \frac{1}{n} + \sup_{1 \le k \le n} \left| \frac{\sum_{i=1}^k \tau_i}{n} - \frac{k}{n} \right| \\ &\le \frac{1}{n} + \sup_{1 \le k \le n} \frac{k}{n} \left| \frac{\sum_{i=1}^k \tau_k}{k} - 1 \right|. \end{aligned}$$

We run into some trouble at this point, since the above expression sugguests strong law of large number, but we are taking $n \to \infty$ and take sup of k for each n. But there is a trick that helps us to split this into two parts,

$$\sup_{1\le k\le n} \frac{k}{n}\left|\frac{\sum_{i=1}^{k}\tau_k}{k}-1\right| \le \sup_{1\le k\le \epsilon n} \frac{k}{n}\left|\frac{\sum_{i=1}^{k}\tau_k}{k}-1\right| + \sup_{k>\epsilon n}\left|\frac{\sum_{i=1}^{k}\tau_i}{k}-1\right|\frac{k}{n}$$
$$\le \epsilon \sup_{1\le k\le \epsilon n}\left|\frac{\sum_{i=1}^{k}\tau_k}{k}-1\right| + \sup_{k>\epsilon n}\left|\frac{\sum_{i=1}^{k}\tau_i}{k}-1\right|\frac{k}{n}$$

now we can use the strong law of large number to see the second term converges to zero a.e. Also note that for a.e ω, $\frac{\sum_{1\le i\le k}\tau_i(\omega)}{k} \to 1$, so for each such ω, the sup over k is finite a.e. so we take $n \to \infty$, and $\epsilon \downarrow 0$ we see the convergence. □

2.4 Stochastic Integration and It's Properties

We have seen in the Gaussian measure section that we can make sense of integration of an $L^2(\mathbb{R})$ function with respect to a Gaussian process, hence with respect to a brownain motion, and the end result would be a Gaussian process. It is not hard to see that if we integrate a L^2 continous function with respect to a Brownian motion, then the end result is also an continous Gaussian process. In this section, we will extend this idea to the case where the itnegrand is an "L^2 bounded" random function. We will shortly see what this means.

First we recall couple of definitions and theorems from before:

Definition 2.4.1. *Let $\{X_t\}_{t\in\mathbb{R}^+}$ be a continous time stochastic process, we say it is progressively measurable if*

$$X_\cdot(\cdot) : (\omega, t) \to X_t(\omega)$$

is measurable with respect to $\mathcal{F}_t \otimes \mathcal{B}([0,t])$ for all $t \ge 0$, we denote this class of function by $\mathcal{P}$

Theorem 2.4.1. *Let $\mathbb{H}^2$ be the collection of martingales M such that $\mathbb{E}\left[\langle M\rangle_\infty\right] < \infty$. Then $\mathbb{H}$ is an Hilbert space whose Hilbert product is defined to be $\langle M, N\rangle_{\mathbb{H}^2} \triangleq \mathbb{E}\left[\langle N, M, \rangle_\infty\right]$.*

Here is a new definition:

Definition 2.4.2. *Let $M \in \mathbb{H}^2$, and denote $L^2(M)$ by the progressively measurable processes H s.t.*

$$\mathbb{E}\left[\int_0^\infty H_s d\langle M\rangle_s\right] < \infty. \tag{2.15}$$

It is easily seen that $d\langle M\rangle d\mathbb{P}$ defines a measure on $(\mathbb{R}^+ \times \Omega, \mathcal{P})$. So in reality, $L^2(M) \equiv L^2\left(\mathbb{R}^+ \times \Omega, \mathcal{P}, d\langle M\rangle\right)$. So $L^2(M)$ itself is again a Hilbert space with inner product defined by

$$\langle H, K\rangle_{L^2(M)} = \mathbb{E}\left[\int_0^\infty H_s K_s d\langle M\rangle_s\right]. \tag{2.16}$$

Definition 2.4.3. *An Elementary Process is a progressively measurable process of the following form:*

$$H(t) = \sum_{n=0}^{\infty} H_n 1_{(t_n, t_{n+1}]}(t) \tag{2.17}$$

where $H_n \in \mathcal{F}_{t_n}$ are ***bounded*** *random variables does not depend on time, and $H_n \ne 0$ for finitely many n.*

As one might have guessed, the construction of stochastic integral is not much differnet from usual calculus:

Theorem 2.4.2. *The class of simple process is dense in $L^2(M)$ for all $M \in \mathbb{H}^2$.*

Proof. Since $L^2(M)$ is a Hilbert space, we only have to show that if $H \in L^2(M)$ is that

$$\mathbb{E}\left[\int_0^\infty H_s K_s d\langle M\rangle_s\right] = 0$$

for all elementary processes K, then $H \equiv 0$ in $L^2(M)$.

But before that, we observe: since $\mathbb{E}\left[\langle M\rangle_\infty\right] < \infty$, the measure $d\langle M\rangle d\mathbb{P}$ is actually a finite measure on the space $\mathbb{R}^+ \times \Omega$. Therefore, by Jessen's inequality (with some normalization on the measure), we have

$$\mathbb{E}\left[\int_0^\infty |H_s| d\langle M\rangle_s\right] < \infty.$$

In other words, the process

$$X_t = \int_0^t H_s d\langle M\rangle_s \tag{2.18}$$

is in $L^1(\mathbb{P})$ for all $t \geq 0$. Moreover, the process X_t has finite variation.

Let $s \geq 0$ be arbitrary, and let $A \in \mathcal{F}_s$. Then by assumption,

$$E\left[(X_t - X_s)\,1_A\right] = \mathbb{E}\left[1_A \int_s^t H_u d\langle M\rangle_u\right] = \mathbb{E}\left[\int_s^t H_u K_u d\langle M\rangle_u\right] = 0$$

where $K_u = 1_A 1_{(s,t]}$ which is an elementary process. Also, since H is progressively measurable, we see $X_t \in \mathcal{F}_t$ for all t. So the above calculation shows that X_t is a mtg, but with finite variation, so it has to be zero. That means, $H_u \equiv 0$ with respect to the measure $d\langle M\rangle_t$. □

Now we define stochastic integration of elementary processes:

Definition 2.4.4. *Let H_t be an elementary process s.t. $H(t) = \sum_{n=1}^N H_n 1_{(t_n,t_{n+1}]}$, then we define the stochastic integration of H with respect to $M \in \mathbb{H}^2$ to be*

$$\int_0^t H_s dM_s = (H \cdot M)_t \triangleq \sum_{n=1}^N H_n\left(M_{t_{n+1}\wedge t} - M_{t_n \wedge t}\right).$$

Remark 2.4.1. *The above definition should look familiar, it is the Martingale Transform we introduced to prove Optional Sampling before. Such a fancy name for stochastic version of simple function integration.*

Here are some properties of stochastic integration of elementary processes:

Theorem 2.4.3. *Let $H = \sum_{i=1}^n H_i 1_{(t_i,t_{i+1}]}$ be a elementary process, and let $M \in \mathbb{H}^2$, then*

1. *$\int_0^t H_s dM_s$ is a true martingale.*

2. *The map $H \to \int_0^\cdot H_s dM_s$ is an isometry from $L^2(M)$ to $\mathbb{H}^2$.*

3. *$\int_0^\cdot H_s dM_s$ is a unique element in $\mathbb{H}^2$ such that*

$$\langle \int_0^\cdot H_s dM_s, N_s \rangle = \int_0^t H_s d\langle M, N \rangle_s \quad \forall N \in \mathbb{H}^2.$$

4. *In particular, $\langle \int_0^\cdot H_s dM_s, \int_0^\cdot H_s dM_s \rangle = \int_0^t H_s^2 d\langle M \rangle_s$.*

5. *If τ is a stopping time, then*

$$\int^{t\wedge\tau} H_s dM_s = \int_0^t 1_{[0,\tau]}(s) H_s dM_s. \tag{2.19}$$

or in Le Gal's notation in ([LG16])

$$\left(1_{[0,\tau]} H \cdot M\right) = (H \cdot M) = (H \cdot M^\tau)$$

Proof. (1) Is straight forward: let $0 \le s \le t$ and consider

$$\mathbb{E}\left[\sum_{i=1}^n H_i \left(M_{t\wedge t_{i+1}} - M_{t\wedge t_i}\right) | \mathcal{F}_s\right]$$

We note that s divies $\{t_i\}$ into two parts, let $i_0 = \min\{i : t_i \ge s\}$, then the above conditional expectation is the same as

$$\begin{aligned}
&\mathbb{E}\left[\ldots \mathbb{E}\left[\sum_{i=1}^n H_i \left(M_{t\wedge t_{i+1}} - M_{t\wedge t_i}\right) | \mathcal{F}_{t_n}\right] \ldots | \mathcal{F}_{t_{i_0}} | \mathcal{F}_s\right] \\
=&\mathbb{E}\left[\sum_{i=1}^{i_0-1} H_i \left(M_{t_{i+1}} - M_{t_i}\right) | \mathcal{F}_s\right] \\
=&\sum_{i=1}^n H_i \left(M_{t_{i+1}\wedge s} - M_{t_i \wedge s}\right).
\end{aligned}$$

This mtg is obvious in $\mathbb{H}^2$ since for $s, t \ge t_n$, $(H \cdot M)_t = (H \cdot M)_s$ henece the quadratic variation will be constant after t_n.

(2)

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^n H_i \left(M_{t_{i+1}} - M_{t_i}\right)\right]^2 =& \sum_{i\neq j} \mathbb{E}\left[H_i H_j \left(M_{t_{i+1}} - M_{t_i}\right)\left(M_{t_{j+1}} - M_{t_j}\right)\right] \\
&+ \sum_{i=1}^n \mathbb{E}\left[H_i^2 \left(M_{t_{i+1}} - M_{t_i}\right)^2\right]
\end{aligned}$$

The first sum is obviously zero which can bee seen by conditioning on the $\mathcal{F}_s$ where $s = t_i \wedge t_j$. Now we look at each term in the second sum:

$$\begin{aligned}
\mathbb{E}\left[H_i^2 \left(M_{t_{i+1}} - M_{t_i}\right)^2\right] &= \mathbb{E}\left[\mathbb{E}\left[H_i^2 \left(M_{t_{i+1}} - M_{t_i}\right)^2 | \mathcal{F}_{t_i}\right]\right] \\
&= \mathbb{E}\left[H_i^2 \mathbb{E}\left[\left(M_{t_{i+1}} - M_{t_i}\right)^2 | \mathcal{F}_{t_i}\right]\right]
\end{aligned}$$

However, we note that $M_t^2 - \langle M \rangle_t$ is a martingale, so

$$\mathbb{E}\left[M_{t_{i+1}}^2 - M_{t_i}^2 | \mathcal{F}_{t_i}\right] = \mathbb{E}\left[\langle M \rangle_{t_{i+1}} - \langle M \rangle_{t_i} | \mathcal{F}_{t_i}\right]$$

so plug this into above we see the isometry property.

(3) We first observe that each term of the integration defines a martingale, that is,

$$M_\cdot^i = H_i \left(M_{t_{i+1}\wedge\cdot} - M_{t_i\wedge\cdot}\right)$$

is a martingale and it is identically zero on the interval $[0, t_i]$, and constant (random variable) on $[t_{i+1}, \infty)$. So we see

$$\left\langle \int_0^\cdot H_s dM_s, N \right\rangle_t = \sum_{i=1}^n \langle M^i, N \rangle_t$$

However,

$$\langle M^i, N \rangle_t = \langle H_i \left(M_{t_{i+1}\wedge\cdot} - M_{t_i\wedge\cdot}\right), N \rangle_t = H_i \left(\langle M, N \rangle_{t\wedge t_{i+1}} - \langle M, N \rangle_{t\wedge t_i}\right)$$

so

$$\left\langle \int_0^\cdot H_s dM_s, N \right\rangle_t = \sum_{i=1}^n \langle M^i, N \rangle_t = \sum_{i=1}^n H_i \left(\langle M, N \rangle_{t\wedge t_{i+1}} - \langle M, N \rangle_{t\wedge t_i}\right) = \int_0^t H_s d\langle M, N \rangle_s$$

Now suppose there is some Y such that $\left\langle \int_0^\cdot H_s dM_s, N \right\rangle_t = \int_0^t H_s \langle M, N \rangle_s = \langle Y, N \rangle$ for all $N \in \mathbb{H}^2$. Then we have

$$\left\langle \int_0^t H_s dM_s - Y, N \right\rangle_t = 0 \quad \forall N \in \mathbb{H}^2$$

and let $N = \int_0^t H_s dM_s - Y$ we see that $\int_0^t H_s dM_s - Y_t$ is identically zero.

(4) is a direct consequence of (3).

For (5), we use the characterization from (3), let $N \in \mathbb{H}^2$, we calculate

$$\langle (1_{[0,\tau]} H \cdot M), N \rangle = \left(1_{[0,\tau]} H\right) \cdot \langle M, N \rangle = H \cdot \langle M, N \rangle^\tau = H \cdot \langle M^\tau, N \rangle$$

where the second to the last equality is due to the nature of lebesgues-Stieltjes integral. Now we consider

$$\langle (H \cdot M)^\tau, N \rangle = \langle H \cdot M, N \rangle^\tau = (H \cdot \langle M, N \rangle)^\tau = \left(1_{[0,\tau]} H\right) \cdot \langle M, N \rangle$$

where the last equality is also due to the nature of lebesgues-Stieltjes integral. □

Since stochastic integral with respect to a fixed $\mathbb{H}^2$ space element is an (partial) isometry from a dense subspace of $L^2(M)$ to $\mathbb{H}^2$, hence there is a natural extension to make the map continous. So we use this as the definition of stochastic integral:

Definition 2.4.5. *Let $H \in L^2(M)$ where $M \in \mathbb{H}^2$, let $H^n \to H$ in $L^2(M)$, then we define the stochastic integral of H with respect to M to be*

$$(H \cdot M)_\cdot = \int_0^\cdot H_s dM_s = \lim_{n\to\infty} \int_0^\cdot H_s^n dM_s$$

Remark 2.4.2. *It can be check with not much effort that all properties in Theorem 68 holds for general integrand (Kunita-Watanabe Inequalities and density argument).*

Theorem 2.4.4. *Let $M \in \mathbb{H}^2$ and $H \in L^2(M)$, then all the properties in Thoerem 68 holds.*

Here we prove a few of them. Note that the characterization of (3) plays an important role in the proof of Thoerem 68. This will also be the case in the following proof.

Proof. Since $\mathbb{H}^2$ is an Hilbert space and $\mathcal{E}$, the space of elementary processes, is dense in $\mathbb{H}^2$, and the stochastic integration is an isometry from $\mathcal{E} \subset L^2(M)$ to $\mathbb{H}^2$. So the continous extention of this map is an isometry from $L^2(M)$ to $\mathbb{H}$, hence $(H \cdot M) \in \mathbb{H}^2$. This solves (1) and (2) (well, this is just stating the obvious anyway).

For (3), let $H^n \to H$ in $L^2(M)$, then consider

$$\lim_{n\to\infty} \langle (H^n \cdot M), N \rangle = \lim_{n\to\infty} H^n \cdot \langle M, N \rangle$$

We note that by Kunita-Watanabe along with Cauchy-Schwartz's Inequalities, we have

$$\begin{aligned}\mathbb{E}\left[\int_0^\infty |H_s^n - H_s|\, d\langle M, N\rangle_s\right] &\leq \mathbb{E}\left[\sqrt{\int_0^\infty (H_s^n - H_s)^2\, d\langle M\rangle_s}\sqrt{\langle N\rangle_\infty}\right] \\ &\leq \sqrt{\mathbb{E}\left[\int_0^\infty (H_s^n - H_s)^2\, d\langle M\rangle_s\right]}\sqrt{\mathbb{E}\left[\langle N\rangle_\infty\right]}\end{aligned}$$

which tends to zero. So

$$\lim_{n\to\infty} \langle (H^n \cdot M), N \rangle = \lim_{n\to\infty} H^n \cdot \langle M, N \rangle = H \cdot \langle M, N \rangle$$

where the last convergence is in $L^1(\mathbb{P})$. Now the only thing left to show is:

$$\langle H \cdot M, N \rangle = \lim_{n\to\infty} \langle (H^n \cdot M), N \rangle$$

However, this is just a consequence of the isometric property:

$$\mathbb{E}\left[\langle (H - H^n) \cdot M, N\rangle\right] \leq \sqrt{\mathbb{E}\left[\langle (H^n - H) \cdot M\rangle_\infty\right]}\sqrt{\mathbb{E}\left[\langle N\rangle_\infty\right]}$$

by the isometric property, we see that

$$\mathbb{E}\left[\langle (H^n - H) \cdot M\rangle_\infty\right] = \|H^n - H\|_{L^2(M)} \to 0.$$

so $\langle H^n \cdot M, N\rangle \to \langle H \cdot M, N\rangle$ in $L^1(\mathbb{P})$.

The last two properties depens only on the above characterization and not the properties of elementary process, so they are automatically true due to (3). □

In calculus, for two functions f, g and a measure μ, if we denote $G(x) = \int_0^x g(y)\mu(dy)$, then we have the following

$$\int_0^t f(x) dG(x) = \int_0^t f(x) g(x) \mu(dx)$$

Similar things also hold in stochastic integrals

Theorem 2.4.5. *Let K, H be progressively measurable processes and $M \in \mathbb{H}^2$. Then $HK \in L^2(M)$ if and only if $H \in L^2((K \cdot M))$. If the latter holds, then*

$$\int_0^{\cdot} H_s d\,(KM)_s = \int_0^{\cdot} H_s K_s dM_s \tag{2.20}$$

or more intuitively,

$$\int_0^t H_s d \int_0^s K_u dM_u = \int_0^t H_s K_s dM_s$$

Proof. We prove the integrability of the stochastic integral first: by definition

$$HK \in L^2(M) \Longleftrightarrow \mathbb{E}\left[\int_0^{\infty} H_s^2 K_s^2 d\langle M\rangle_s\right] < \infty.$$

Also,

$$H \in L^2\left((K \cdot M)\right) \Longleftrightarrow \mathbb{E}\left[\int_0^{\infty} H_s^2 d\langle (K\cdot M)\rangle_s\right] < \infty.$$

However,

$$\langle (K \cdot M)\rangle_t = \int_0^t K_s^2 d\langle M\rangle_s$$

and this is just Lebesgue-Stieltjes integral, so

$$\mathbb{E}\left[\int_0^{\infty} H_s^2 d\langle (K\cdot M)\rangle_s\right] = \mathbb{E}\left[\int_0^{\infty} K_s^2 H_s^2 d\langle M\rangle_s\right].$$

So the integrability part is proven.

To prove the identity in (2.20), by Thoerem 68(3), we only need to show that for all $N \in \mathbb{H}^2$, the left and right hand sides, call them L, R, satisfies

$$\langle L, N\rangle = \langle R, N\rangle$$

but this is obvious. □

In the above setting, $\int_0^{\cdot} H_s dM_s \in \mathbb{H}^2$, so it is square integrable and since for general square integrable mtg that starts with zero, the expectation of the square equals to the expectation of the quadratic variation, we have

$$\mathbb{E}\left[\int_0^t H_s dM_s\right]^2 = \mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right].$$

and its expectation is

$$\mathbb{E}\left[\int_0^t H_s dM_s\right] = 0 \quad \forall t \geq 0.$$

Stochastic Integration for Mtg not in $\mathbb{H}$

Mostly we integrate progressively measurable processes with respect to Brownian motion, which is a true martingale that is not in $\mathbb{H}$. Although we cannot define such process globally, we can still define it locally. Here is what I mean:

Let M be a true sqaure integrable martingale that is not neccessarily in $\mathbb{H}^2$, meaning $\mathbb{E}\left[\langle M\rangle_t\right]$ might not be bounded. Let H be a progressively measurable process such that

$$\mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right] < \infty \quad \forall t \geq 0$$

Here we can define the stochastic integration "pointwise": let $T > 0$ and for all $t \in [0, t]$, we define $(H \cdot M)_t$ to be

$$(H \cdot M)_t = \int_0^t H_s dM_s^T$$

where the process M^T is

$$M_t^T = M_{t \wedge T}.$$

In which case, $M^T \in \mathbb{H}^2$ for all $T \geq 0$, so the above definition makes sense. In the case for Brownian motion, it does not make any difference wheather we use a fixed number T, or ues the stopped process B^{τ_T} where $\tau_T = \inf\{t : \langle B\rangle_t = T\}$ since $\langle B\rangle_t = t$. The introduction of stopping time seems extraneous, but it will serve us well when we want to define integral with respect to local mtg's.

In the case of integration with respect to true square integrable mtg, we see that $\mathbb{E}\left[\langle M\rangle_t\right] = \mathbb{E}\left(M_t^2\right) < \infty$ for all $t \geq 0$, so we can drop the dependent of the stochastic integral on T and just write

$$(H \cdot M)_t = \int_0^t H_s dM_s$$

for $t \in \mathbb{R}^+$, which still makes sense pointwise for each t. We note here that the properties of stochastic integral with respect to the general square integrable Mtgs is the same as the stochastic integral with respect to $\mathbb{H}^2$ elements, due to the following theorem:

Theorem 2.4.6. *Let $\mathcal{M}^c$ be the space of continuous square integrable true martingales, and we define the norm on this space to be*

$$\|M\|_{\mathcal{M}^2} = \sum_{i=1}^{\infty} \frac{1 \wedge \sqrt{\mathbb{E}\left[\langle M\rangle_n\right]}}{2^n}$$

Then $(\mathcal{M}^c, \|\cdot\|)$ forms a Banach space.

We omit the proof, but it utilizes Borel-Cantelli to have uniformly convergence sequence on compact intervals. With above theorem we can actaully define stochastic integration with respect to square integrable Mtg's directly. See ([KS12]) for this appraoch.

Stochastic Integration w.r.t. local Mtg's

Here we look at stochastic integration with respect to local mtg's. With the observation in previous sub-section, and in the square integrable martingale section, we see that we can turn a local martingale into a true square integrable martingale that is in $\mathbb{H}^2$ via stopping times. But before that, we need to specify integrability with respect to a certain local mtg:

Definition 2.4.6. *Let M be a local martingale, we define $L^2(M)$ and $L^2_{loc}(M)$ to be*

$$L^2_{loc}(M) = \left\{ H \text{ progressively measurable} : \int_0^t H_s^2 d\langle M \rangle_s < \infty, \text{a.e. } \omega, \text{for all } t \geq 0 \right\}$$

and

$$L^2(M) = \left\{ H \text{ progressively measurable} : \mathbb{E}\left[\int_0^\infty H_s^2 d\langle M \rangle_s \right] < \infty \right\}$$

Remark 2.4.3. *Where we note locally square integrable is pointwise ω square integrable (almost sure) and square integrable is the same defintion as for $M \in \mathbb{H}^2$, which is $d\langle M \rangle_t d\mathbb{P}$ square integrable.*

Remark 2.4.4. *We note that $L^2(M) \subset L^2_{loc}(N)$, so we can go ahead establish stochastic integration of locally $L^2(M)$ element and it will work for $L^2(M)$ elements.*

First we want to turn M into a $\mathbb{H}^2$ element as well as getting H^2 to be square integrable with respect to $d\langle M \rangle d\mathbb{P}$ to use the stochastic integration we developed in the previous section. To get those two thing together, we use the following stopping time

$$\tau_n = \inf_{t \geq 0} \left\{ \int_0^t 1 + H_s^2 d\langle M \rangle_s > n \right\}$$

and we see that $M^{\tau_n} \in \mathbb{H}^2$ for all $n \geq 1$, since $\langle M^{\tau_n} \rangle_t = \langle M \rangle_{t \wedge \tau_n} \leq n$ for all $t \geq 0$. Also,

$$\mathbb{E}\left[\int_0^\infty H_s^2 \langle M^{\tau_n} \rangle_s \right] = \mathbb{E}\left[\int_0^{\tau_n} H_s^2 \langle M \rangle_s \right] \leq n$$

so $H \in L^2(M^{\tau_n})$ and the stochastic integraiton

$$(H \cdot M^{\tau_n})_t = \int_0^t H_s dM_s^{\tau_n}$$

is defined as in previous section (integration with respect to $\mathbb{H}^2$ element), so for any $m \geq n$ we have

$$(H \cdot M^{\tau_n})^{\tau_m} = \left(1_{[0,\tau_m]} H \right) \cdot M^{\tau_n} = H \cdot M^{\tau_n \wedge \tau_n} = H \cdot M^{\tau_n}.$$

This tells us that $H \cdot M^{\tau_n}$ agrees on $[0, \tau_m]$ for all $n \geq m$. This tells us that there is a unique process $H \cdot M$ such that for all $n \in \mathbb{N}$,

$$(H \cdot M)^{\tau_n} = H \cdot M^{\tau_n}$$

which has continous sample pathes and it is adapted because it is pointwise limit of a seqeunce of adapted processes $((H \cdot M^{\tau_n}))$. So we can define stochatic integral of H with respect to M to be $H \cdot M$ as above, and the resulting integration is a continous local martingale.

Just like in the case of integral with repsect to $\mathbb{H}^2$ element, we can characterize the stochastic integral via quadratic variations. Let H, M be as above, and let N be another local martingale that can vary. Let $\sigma_n = \inf_{t \geq 0}\{\langle N \rangle_t > n\}$, and let $T_n = \tau_n \wedge \sigma_n$, we consider the following:

$$\langle H \cdot M, N \rangle^{T_n} = \langle (H \cdot M)^{\tau_n}, N^{\sigma_n} \rangle = H \cdot \langle M^{\tau_n}, N^{\sigma_n} \rangle = H \cdot \langle M, N \rangle^{T_n} = (H \cdot \langle M, N \rangle)^{T_n}$$

and since $T_n \uparrow \infty$, we see that

$$\langle H \cdot M, N \rangle_t = H \cdot \langle M, N \rangle_t \text{ or } \langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$$

and if X is some other local martinagle such that $\langle X, N\rangle = H\langle M, N\rangle$, then $\langle H\cdot M - X, N\rangle = 0$ and let $N = H\cdot M - X$ we see that $\langle H\cdot M - X\rangle \equiv 0$ hence this property is unique.

It can also be shown that the properties of stochastic integrals with respect to $\mathbb{H}^2$ element also holds here and the proof is very similar to the above discussion.

Now, let M be a local martingale and let H_s be such that

$$\mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right] < \infty \quad \forall t \geq 0.$$

Then for a fixed t, $(H\cdot M)^t = \int_0^{\cdot\wedge t} H_s dM_s$ is a $\mathbb{H}^2$ element. So in this case we have

$$\mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right]^2 = \mathbb{E}\left[\int_0^t H_s dM_s\right]^2.$$

and

$$\mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right] = 0.$$

Integration with respect to Semi-Martingale

Now this section will be short, since if X is a semi-martinagle, then $X = M + A$ where M is a local martiangle and A is an process with finite variation, so we interpret the integration with respect to a semi-martingale as

$$H\cdot X = H\cdot M + H\cdot A$$

where $H\cdot M$ is interpreted as the stochastic integration we established before, and $H\cdot A$ is interpreted as the usual Lebesgue-Stieltjes integrals. This definition makes sense whenever the two integrals make sense. Also, when the integral makes sense, then the properties hold both for the stochatic and the "deterministic" integrals would be true.

One of the conditions for integrability is "locally bounded":

Definition 2.4.7. *Let H be a progressively measurable process, then we say it is locally bounded if*

$$\forall t \geq 0, \quad \sup_{0\leq s\leq t} |H_s| < \infty, a.e.$$

Note that if a process is continuous, adapted and progressively measurable, then it is locally bounded (pathwise continous implies path wise bounded). In which case, since A has finite variation, we see

$$\int_0^t |H_s| d|A_s| < \infty \quad a.e.$$

where $|A_t|$ is the total variation of A. Same for $\int_0^t |H_s|^2 d\langle M\rangle_s$ since $\langle M\rangle$ is also of finite variation.

Dominated Convergence Thoerem and Approximation of Stochastic Integral

Here we introduce the stochastic version of the Dominated convergence theorem, which says if a sequence of locally bounded processes that converges "pointwise" t to an "integrable" progressively measurable process, and is dominated by someother locally bounded progressively measurable process, then the stochastic integral converges in probability (which is the thrid best thing we can hope for).

Theorem 2.4.7. *Let $X = M + A$ be a semi martingale where M is local mtg and A is of finite variation, and suppose H^n, H are locally bounded progressively measurable processes, and let K be a progressively measurable process, then if*

1. *$H_t^n \to H_t$ a.e. ω for all $t \geq 0$.*
2. *$|H_t^n| \leq |K_t|$ a.e. for all $t \geq 0$, and*
3. *$\int_0^t K_s^2 d\langle M\rangle_s < \infty$ and $\int_0^t |K_s| d|V_s| < \infty$ a.e. for all $t \geq 0$.*

Then

$$\int_0^t H_s^n dX_s = \int_0^t H_s dX_s \quad \text{in Probability.}$$

Remark 2.4.5. *As discussed just before this subsection, locally bounded is sufficient for stochastic integral to make sense.*

Remark 2.4.6. *For the proof, we notice that there only two things we need to show: (1) the lebesgues-Stieltjes integral converges, which is quite obvious by usual DCT, and the other is convergence of the stochastic integral. And at this point, it should be natrual for us to think that convergence in probability is usually shown by convergence in L^p for some $p > 0$. In this case, since we can turn things into square integrable martingales, $p = 2$ seems prominante.*

Proof. For the lebesgues-Stieltjes part, we actually have a.e. ω convergence since for almost every ω, we have

$$|H_t^n(\omega)| \leq |K_t(\omega)| \text{ and } \int_0^t K(\omega) d|A_t(\omega)| < \infty.$$

and $H_t^n(\omega) \to H_t(\omega)$ for all t for a given ω, so usual DCT kicks in and we are done with this part.

Now for the stochastic integral part, let $\tau_n = \inf_{t\geq 0}\left\{\int_0^t H_s^2 + K_s^2 d\langle M\rangle_s \geq n\right\}$, so τ_n is a stopping time, and the stochatic integral

$$\int_0^{t\wedge\tau_n} Z_s dM_s = \int_0^t Z_s dM_s^{\tau_n}$$

are all $\mathbb{H}^2$ element for $Z = H, K, H^n$. We want L^2 convergence, that is,

$$\mathbb{E}\left[\int_0^{t\wedge\tau_n} (H_s^n - H_s)\, dM_s\right]^2 \to 0.$$

for all τ_n, however, we do have that

$$\mathbb{E}\left[\int_0^{t\wedge\tau_n} (H_s^n - H_s)\, dM_s\right]^2 = \mathbb{E}\left[\int_0^{t\wedge\tau_n} (H_s^n - H_s)^2\, d\langle M\rangle_s\right]$$

Now, the thing inside the expectation converges to zero pointwise by usual DCT, since for almost all ω, $H_s^n(\omega) \to H_s(\omega)$ for almost all t with respect to the measure $d\langle M\rangle_s$ which is absolutely continous with respect to the lebesgues measure. And the thing inside of the expectation is also dominated by $C\int_0^{t\wedge\tau_n} |H_s|^2 + |K_s|^2 d\langle M\rangle_s$ for some C for almost all ω. So another application of DCT gives us the desired L^2 convergence, hence convergence in probability. □

Using the stochastic dominated convergence theorem, we now can see that the there is not much difference between stochastic integral and the integrals we see in real analysis, at least in the case of continuous integrand:

Theorem 2.4.8. *Let X be a continous semi-martingale and let H be a continous progressively measurable continous process. If Π_n are a sequence of partitions of the interval $[0, t]$ whose mesh goes to zero, then we have*

$$\lim_{n\to\infty} \sum_{\Pi_n} H_{t_i}\left(X_{t_{i+1}} - X_{t_i}\right) = \int_0^t H_s dX_s$$

where the convergence is in probability measure.

Proof. We first notice that H is an locally bounded process, and let $K_t = \sup_{0\leq s\leq t} |H_s| < \infty$, then $\int_0^t K_s^2 d\langle M\rangle_s$ and $\int_0^t K_s d|A_t|$ are both finite since the integrators are of finite variations (and absolutely continous with respect to the lebesgues measure for almost every ω). Under this case, we may use the stochastic DCT.

Let $\{t_i\}_{i\leq i\leq n} = \Pi_n$ and define the sequence of progressively measurable processes:

$$H_t^n = \begin{cases} H_{t_i} & t \in (t_i, t_{i+1}] \\ H_0 & t = 0 \\ 0 & \text{elsewhere.} \end{cases}$$

In which case, we have $H_t^n \to H_t$ a.e. ω for all t since H itself is continous, and

$$\int_0^t H_s^n dX_s = \sum_{\Pi_n} H_{t_i}\left(X_{t_{i+1}} - X_{t_i}\right)$$

by definition. So by Stochastic DCT, we are done. □

Remark 2.4.7. *It is important to use the left hand H_{t_i} in the approximation instead of $H_{t_{i+1}}$. To see what can go wrong, see page 113 of [LG16].*

2.4.1 Ito's Formula

Ito's formula is essentially the change of variable formula for stochastic integrations. We will see shortly that it is not much different from usual calculus change of variable formula with the difference being a quadratic variation added as a correction term.

Theorem 2.4.9. *Let $X = \left(X^1, \ldots, X^n\right)$ be a n dimensional semi-martinagle and let $F : \mathbb{R}^n \to \mathbb{R}$ be a twice continous differentiable function, then for all $t \geq 0$ we have the following a.e. equality*

$$F(X_t) = F(X_0) + \sum_{1\leq i\leq n} \int_0^t \frac{\partial F}{\partial x_i}(X_s) dX_s^i + \frac{1}{2} \sum_{1\leq i,j\leq n} \int_0^t \frac{\partial^2 F}{\partial x_i x_j}(X_s) d\langle X^i, X^j\rangle_s.$$

Remark 2.4.8. *Suppose $X = M + A$ and to write the above equation explicitely, it is*

$$F(X_t) = F(X_0) + \sum_{1\leq i\leq n} \int_0^t \frac{\partial F}{\partial x_i} F(X_s) dM_s + \sum_{1\leq i\leq n} \int_0^t \frac{\partial F}{\partial x_i}(X_s) dA_s + \frac{1}{2} \sum_{1\leq i,j\leq n} \int_0^t \frac{\partial^2 F}{\partial x_i x_j}(X_s) d\langle M^i, M^j\rangle_s$$

We can also write the above in differential notation with the knowledge that it is actually an integral equation:

$$dF(X_t) = \sum_{1\leq i\leq n} \partial_{x_i} F(X) dX_t^i + \frac{1}{2} \sum_{1\leq i,j\leq n} \partial_{x_i,x_j} F(X_t) d\langle X^i, X^j\rangle_t$$

Proof. The proof is basically Tylor expansion, here we only prove the case where $n = 1$ and remark that there is no fundamental difference between one dimensional case and higher dimensional case.

Here we assume WLOG that $X = M + A$ where M is a martingale, since we can localize it by $\tau_n = \inf_{t\geq 0}\{|M_t| + |A_t| \geq n\}$, and if the above formula is true for all mgt M, then it will be true for local martingales. Since we can take τ_n to be this form, then $X_t^{\tau_n}$ would be a bounded process, then we might as well just assume that F is compactly supported since it does not make a difference when X is bounded. Say $|M|, |A|$ all bounded by K.

Let Π_n be a sequence of increasing partitions whose mesh goes to zero, and consider

$$\begin{aligned} F(X_t) - F(X_0) &= \sum_{\Pi_n} \left(F(X_{t_i+1}) - F(X_{t_i})\right) \\ &= \sum_{\Pi_n} F'(X_{t_i})\left(X_{t_{i+1}} - X_{t_i}\right) + \frac{1}{2}F''\left(X_{t_i} + \theta_i \Delta_i X\right)\left(X_{t_{i+1}} - X_{t_i}\right)^2 \end{aligned}$$

where $\theta_i(\omega)$ is some number between 0 and 1 and $\Delta_i X = -X_{t_i})$. We notice that F' is continous, so $F'(X_t)$ is locally bounded which dominates the elementary processes $F^n(X_t) = \sum_{\Pi_n} F'(X_{t_i}) 1_{(X_{t_i}, X_{t_{i+1}})}(t)$, so by the approximation theorem above, the first term converges to $\int_0^t F'(X_s)dX_s$ in probability measure. So we only need to show the convergence of the second term to the lebesgues-Stieltjes integral. However, the second term can be written as

$$\sum_{\Pi_n} \frac{1}{2}\left(F''\left(X_{t_i} + \theta_i\Delta_i X\right) - F''(X_{t_i})\right)\left(X_{t_{i+1}} - X_{t_i}\right)^2 + \sum_{\Pi_n}\frac{1}{2}F''(X_{t_i})\left(X_{t_{i+1}} - X_{t_i}\right)^2$$

Call the first term J_1 and second term J_2. Here note that we only need to find a subsequence that converges either in probability or a.e. to the targeting process. We note $\sum_{\Pi_n}\left[X_{t_{i+1}} - X_{t_i}\right]^2 \to \langle M\rangle_t$ in probability, so we take a susbequence such that this converges a.e. Also, by continuity of F'' and the continuity (pathwise) of X_t, and the fact that we are working inside of an compact interval, $F \circ X(\omega)_t$ is uniformly continous in t, hence

$$\begin{aligned} &\sum_{\Pi_n} \frac{1}{2}\left(F''\left(X_{t_i} + \theta_i\Delta_i X\right) - F''(X_{t_i})\right)\left(X_{t_{i+1}} - X_{t_i}\right)^2 \\ \leq &\frac{1}{2}\sup_{t_i}\left|\left(F''\left(X_{t_i} + \theta_i\Delta_i X\right) - F''(X_{t_i})\right|\sum_{\Pi_n}\left(X_{t_{i+1}} - X_{t_i}\right)^2 \end{aligned}$$

however, $\sup_i\left|\left(F''\left(X_{t_i} + \theta_i\Delta_i X\right) - F''(X_{t_i})\right| \to 0$ as $n \to \infty$, and there is a subsequence such that $\sum_{\Pi_n}\left(X_{t_{i+1}} - X_{t_i}\right)^2 \to \langle X\rangle_t$ so the entire J_1 goes to zero a.e. (for some subsequence).

Now we need to show J_2 converges to the targetting thing. We take a look at the square difference term in J_2:

$$\left(X_{t_{i+1}} - X_{t_i}\right)^2 = \left(M_{t_{i+1}} - M_{t_i}\right)^2 + 2\left(A_{t_{i+1}} - A_{t_i}\right)\left(M_{t_{i+1}} - M_{t_i}\right) + \left(A_{t_{i+1}} - A_{t_i}\right)^2$$

and since $F(X_s(\omega)$ is bounded on the interval $[0, t]$ for a.e. ω. So when we sum things up adding the $F(X_t)$ term, the last two terms disapears for a.e. ω (take a.s. convergence subsequence if neccessary). So we only need to worry about the following sum

$$\sum_{\Pi_n} F(X_{t_i})\left(M_{t_{i+1}} - M_{t_i}\right)^2$$

but

$$\mathbb{E}\left[\sum_{\Pi_n} F(X_{t_i})\left(M_{t_{i+1}} - M_{t_i}\right)^2 - \sum_{\Pi_n} F(X_{t_i})\left(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}\right)\right]^2 \tag{2.21}$$

$$= \sum_{i,j} \mathbb{E}\Phi_i\Phi_j \tag{2.22}$$

where $\Phi_i = F(X_{t_i})\left(M_{t_{i+1}} - M_{t_i}\right)^2 - \left(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}\right)$, we look at the case where $i \neq j$, and say $j < i$, and for the sake of display setting let's denote $\xi_i = \left(\left(M_{t_{i+1}} - M_{t_i}\right)^2 - \left(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}\right)\right)$, and we note $\xi_j \in \mathcal{F}_{t_i}$ since $t_{j+1} \leq t_i$.

$$\begin{aligned}
\mathbb{E}\left[\Phi_i\Phi_j\right] &= \mathbb{E}\left[F(X_{t_i})F(X_{t_j})\xi_i\xi_j\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[F(X_{t_i})F(X_{t_j})\xi_i\xi_j|\mathcal{F}_{t_i}\right]\right] \\
&= \mathbb{E}\left[F(X_{t_i})F(X_{t_j})\xi_j\mathbb{E}\left[\xi_i|\mathcal{F}_{t_i}\right]\right]
\end{aligned}$$

Now let's investigate the conditional expectation inside:

$$\begin{aligned}
\mathbb{E}\left[\xi_i|\mathcal{F}_{t_i}\right] &= \mathbb{E}\left[\left(\left(M_{t_{i+1}} - M_{t_i}\right)^2 - \left(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}\right)\right)|\mathcal{F}_{t_i}\right] \\
&= \mathbb{E}\left[M^2_{t_{i+1}}|\mathcal{F}_{t_i}\right] - 2\mathbb{E}\left[M_{t_{i+1}}M_{t_i}|\mathcal{F}_{t_i}\right] + E\left[M^2_{t_i}|\mathcal{F}_{t_i}\right] - E\left[\langle M\rangle_{t_{i+1}}|\mathcal{F}_{t_i}\right] + \langle M\rangle_{t_i} \\
&= \mathbb{E}\left[(M^2_{t_{i+1}} - \langle M\rangle_{t_{i+1}}) - (M^2_{t_i} - \langle M\rangle_{t_i})|\mathcal{F}_i\right]
\end{aligned}$$

but this is zero since $M_t^2 - \langle M\rangle_t$ is a martingale (recall the assumption in the beginning). So (2.21) is really

$$\begin{aligned}
\mathbb{E}\left[\sum_i \Phi_i^2\right] &= \mathbb{E}\left[F(X_{t_i})^2\left(\left(X_{t_{i+1}} - X_{t_i}\right)^2 - \left(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}\right)\right)^2\right] \\
&\leq \|F\|_\infty \mathbb{E}\left[\sum_{\Pi_i}\left(\left(X_{t_{i+1}} - X_{t_i}\right)^2 - \left(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}\right)\right)^2\right] \\
&\leq \|F\|_\infty \mathbb{E}\left[\sum_{\Pi_i}\left(X_{t_{i+1}} - X_{t_i}\right)^4 + \left(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}\right)^2\right]
\end{aligned}$$

The first term goes to zero since

$$\begin{aligned}
\mathbb{E}\left[\sum_{\Pi_i}\left(X_{t_{i+1}} - X_{t_i}\right)^4\right] &\leq \mathbb{E}[\max\left(X_{t_{i+1}} - X_{t_i}\right)^2 \sum_{\Pi_n}\left(X_{t_{i+1}} - X_{t_i}\right)^2] \\
&= \mathbb{E}[\max\left(X_{t_{i+1}} - X_{t_i}\right)^2 \sum\left(X_{t_{i+1}} - X_{t_i}\right)^2]
\end{aligned}$$

where the maximum term goes to zero and the sum goes to $\langle M\rangle_t$ a.e. (take subsequence) and both of them are bounded. The second term goes to zero as well for similar reasons, hence we have the desired convergence. □

Let's do an application of Ito's lemma to something called the exponential (local) martingales

Lemma 2.4.1. *Let M be a continous local martingale and we denote $\mathcal{E}(M)$ the process*

$$\mathcal{E}(M)_t = \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right)$$

and it is an local martingale.

Proof. Note we can view $M - \frac{1}{2}\langle M\rangle$ as a semi-martingale, so apply Ito's rule to the exponential function to get

$$\begin{aligned} d\mathcal{E}(M)_t &= \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right) dM - \frac{1}{2}\exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right) d\langle M\rangle_t + \frac{1}{2}\exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right) d\langle M\rangle_t \\ &= \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right) dM_t \end{aligned}$$

so writting $\mathcal{E}(M)$ explicitely to get

$$\mathcal{E}(M)_t = 1 + \int_0^t \exp\left(M_s - \frac{1}{2}\langle M\rangle_s\right) dM_s$$

and we see that $\tau_n = \inf_{t\geq 0}\{|M_t| + \langle M\rangle_t \geq n\}$ would reduce this to an true martingale. □

Remark 2.4.9. *We can use the same proof to show*

$$\mathcal{E}(\lambda M) = \exp\left(\lambda M - \frac{\lambda}{2}\langle M\rangle\right) \tag{2.23}$$

is a local martignale as well, for $\lambda \in \mathbb{C}$, which would give us a complex martingale, namely, both real and imaginary parts are continous local martingales.o

From Ito's formula, we can derive change of variable formulas for stochastic integration:

Theorem 2.4.10 (Integration by Parts)**.** *Let M, N be two local martingales, then*

$$M_tN_t = N_0M_0 + \int_0^t M_s dN_s + \int_0^t N_s dM_s + \langle M, N\rangle_t$$

Proof. Let $F(x,y) = xy$ where $x, y \in \mathbb{R}$, so F is twice differentiable whose second derivatives are globally bounded (this does not give us anything extra though, but it is nice), so we apply Ito's formula to

$$dM_tN_t = dF(M_t, M_t) = M_t dN_t + N_t dM_t + d\langle M, N\rangle_t$$

□

Levy's Characterization of Brownian Motion

From here we do some "application" of Ito's lemma to show some important theorems. First, we show Brownain motions' characters are unique to Brownian motions

Theorem 2.4.11. *Let B be a continous local martingale, then it is a Bronwian motion if and only if $\langle B\rangle_t = t$.*

More generally, if $(B_t^1, \ldots, B_t^n)$ is a $\mathbb{R}^n$ local martingale, then it is $\mathbb{R}^n$ Brownain motion if and only if $\langle B^i, B^j\rangle_t = \delta_{i,j}t$.

Proof. Here are the things we need to show: (1) $B_t - B_s \sim \mathcal{N}(0,(t-s)I_d)$ where I_d is the identity matrix of $\mathbb{R}^n$; (2) B has independent increments. For (1), we can utilize the uniqueness of characteristic functions of probability measures (or random variables).

(1) Let $\xi \in \mathbb{R}^n$, we consider $\mathcal{E}(i\xi \cdot B)$:

$$\mathcal{E}(i\xi \cdot B)_t = \exp\left(i\xi \cdot B_t + \frac{1}{2}|\xi|t\right)$$

which is still a continous local martingale. We note that for $|\mathcal{E}(i\xi \cdot B)_t| \leq e^{\frac{1}{2}|\xi|t}$ which is bounded, so $\mathbb{E}\langle \mathcal{E}(i\xi \cdot B)_t\rangle_t < \infty$ for all t, hence it is a true martingale. Now we calculate its increment: suppose $0 \leq s \leq t$, then

$$\mathcal{E}(i\xi \cdot B)_t - \mathcal{E}(i\xi \cdot B)_s = \exp\left(i\xi \cdot B_s + \frac{1}{2}|\xi|t\right)\left(\exp\left(i\xi \cdot (B_t - B_s) + \frac{1}{2}|\xi|^2(t-s)\right) - 1\right)$$

conditioning both side on $\mathcal{F}_s$ to see that

$$\mathbb{E}\left[\exp\left(i\xi \cdot (B_t - B_s) + \frac{1}{2}|\xi|^2(t-s)\right)|\mathcal{F}_s\right] = 1 \tag{2.24}$$
$$\Rightarrow \mathbb{E}\left[\exp\left(i\xi \cdot (B_t - B_s)\right)\right] = \exp\left(-\frac{1}{2}|\xi|^2(t-s)\right)$$

so we conclude that $B_t - B_s \sim \mathcal{N}(0, I_d(t-s))$. However, from (2.24) we also note that for all $A \in \mathcal{F}_s$ we have

$$\mathbb{E}\left[1_A \exp\left(i\xi \cdot (B_t - B_s)\right)|\mathcal{F}_s\right] = 1_A \mathbb{E}\left[\exp\left(i\xi \cdot (B_t - B_s)\right)|\mathcal{F}_s\right] = 1_A \exp\left(-\frac{1}{2}|\xi|^2(t-s)\right)$$

Let $\varphi_A(\eta) = e^{i\eta 1_A}$, then we can replace 1_A above by φ_A and ge similar result, taking expectation to see

$$\mathbb{E}\left[e^{i\eta 1_A}\exp\left(i\xi \cdot (B_t - B_s)\right)\right] = \mathbb{E}\left[e^{i\eta 1_A}\right]\exp\left(-\frac{1}{2}|\xi|^2(t-s)\right)$$

so by Kac's theorem for independence, $B_t - B_s \perp 1_A$ for all $A \in \mathcal{F}_s$, so $B_t - B_s \perp \mathcal{F}_t$. □

Continous Mtg as Time Changed Brownian Motions

Here we see that Brownian motion is a fundamental martingale for the following reason:

Theorem 2.4.12 (Dubin-Dambis-Schwartz)**.** *Let M be a continous local martingale with $\langle M\rangle_\infty = \infty$. Then there is a B, brownain motion such that*

$$M_t = B_{\langle M\rangle_t}$$

where the equality is a.e. ω.

Remark 2.4.10. *The result is both obviously suprising and obviously unsuprising, since we already have Levy's characterization of Brownain motion, then we can image that M_{τ_t} where $t = \inf\{s \geq 0 : \langle M\rangle_s = t\}$ is likely to be a Brownain motion and it is likely to satisfy the theorem.*

Proof. Just as in the remark, let $\tau_t = \inf_{s\geq 0}\{\langle M\rangle_s = t\}$, and let $B_t \triangleq M_{\tau_t}$, then we want to check this is a Brownain motion, so by Levy's characterization of Brownian motion, we need to check the following things:

1. There is a filtration $\mathcal{G}_t$ such that B_t is a local martingale.
2. $\langle B \rangle_t = t$, and
3. B_t is continous.

Let's do them one by one.

(1) Let $\mathcal{G}_t \triangleq \mathcal{F}_{\tau_t} = \{A \in \mathcal{F}_\infty : A \cap \{\tau_t \leq s\} \in \mathcal{F}_s\}$. To check this filtration makes B_t is martingale, let $s \leq t \leq T$ then $\tau_s \leq \tau_t \leq \tau_T$ a.e, so we see

$$(M^{\tau_T})^{\tau_t} = M^{\tau_t}$$

and same for τ_s. Furthermore, M^{τ_T} is an uniformly integrable martingale since $\langle M^{\tau_T} \rangle_t \leq T$ is uniformly bounded, therefore, we can apply optional sampling theorem:

$$\mathbb{E}\left[M_{\tau_t} | \mathcal{F}_{\tau_s}\right] = \mathbb{E}\left[M_{\tau_t}^{\tau_T} | \mathcal{F}_{\tau_s}\right] = M_{\tau_s}^{\tau_T} = M_{\tau_s}$$

so B is indeed a martingale adapted to $\mathcal{G}_t$.

(2) is true by construction.

(3) We note first that τ_t is left continous with right limit, where we denote the right limit by $\tau_{t+} \triangleq \lim_{\epsilon \downarrow 0} \tau_{t+\epsilon} = \inf\{s \geq 0 : \langle M \rangle_s > t\}$. So by continuity of M, we see that B is at least left continous. Since $\langle M \rangle_s$ is pathwise continuous and nondecreasing in s, so the only thing that can give us trouble is when $\tau_{t+} > \tau_t$. However, also by such property, we see that $\langle M \rangle_{\tau_{t+}} = \langle M \rangle_{\tau_t}$. In that case, we can show B is continous by showing that for $\tau_{t+} > \tau_t, M_{\tau_t \vee s}^{\tau_{t+}} - M_{\tau_t}^{\tau_{t+}} = 0$ or $M_s^{\tau_{t+}} - M_{\tau_t \wedge s}^{\tau_{t+}} = 0$ pathwise for $s \geq \tau_t$. One of such ways is to show the quadratic variation of such double stopped process is zero, and here we look at the first one:

$$\langle M_{\tau_t \vee s}^{\tau_{t+}} - M_{\tau_t}^{\tau_{t+}} \rangle = \langle M^{\tau_{t+}} \rangle_{\tau_t \vee s} - \langle M^{\tau_{t+}} \rangle_{\tau_t} = \langle M \rangle_{(\tau_t \vee s) \wedge \tau_{t+}} - \langle M \rangle_{\tau_t \wedge \tau_{t+}}$$

but this is zero since $\langle M \rangle$ is pathwise nondecreasing and continuous, so $M_{\tau_t} = M_{\tau_{t+}} = B_t = B_{t+}$ hence B is continous. □

Now, suppose $\langle M \rangle_\infty$ is not identically infinity, then we can still represent it as a time changed Brownain motion:

Theorem 2.4.13. *Let M be a local martiangle, we define*

$$\tau_t = \inf\{s \geq 0 : \langle M \rangle_s = t\}, \quad \mathcal{G}_t = \mathcal{F}_{\tau_t}$$

then there is a Brownain motion w.r.t. a possibly extended filtration $\widehat{\mathcal{G}\ddagger}$ of $\mathcal{G}$ and probability space such that

$$B_t = M_{\tau_t} \quad on\ [0, \langle M \rangle_\infty)$$

and we have following representation:

$$M_t = B_{\langle M \rangle_t}$$

Proof. Here we can make the assumption that there is a Brownain motion W with respect to the filtration $\mathcal{G}_t$ that is independent of M, since if not, we can create in another probabiliy space a independent Brownian motion and use product σ field to create such thing. We define the process B as

$$B_t = M_{\tau_t} + \int_0^t 1_{\tau_s < \infty} dW_s$$

since M and W are independent, the quadratic variation of B is

$$\langle B\rangle_t = \langle M\rangle_{\tau_t} + \int_0^t 1_{\tau_s<\infty} ds$$

where we note that the first term is $t \wedge \langle M\rangle_\infty$, and the second term is

$$\int_0^t 1 - 1_{\tau_s=\infty} ds = \int_0^t 1 - 1_{\langle M\rangle_\infty \leq s} ds = t - \int_0^{\langle M\rangle_\infty \wedge t} ds$$

furthermore, M_{τ_t} is a local martignale with respect to $\mathcal{G}_t$, and so is $\int_0^t 1_{\tau_t<\infty} dW_s$, so Levy tells us B_t is a Brownain motion. Furthermore, on the set $\{s \leq \langle M\rangle_\infty\}$, we have

$$B_{\langle M\rangle_{\tau_s}} = M_{\tau_{\langle M\rangle_t}} = M_t \quad \text{a.e.}$$

□

Two representation theorems of local martingales in terms of integral of Brownian Motions

Our first goal is to represent a local martingale in terms of a Brownain motion, possibly in a extended probability space. We first take a look at the one dimensional case:

Proposition 2.4.1. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ be a filtered probability space, and suppose M is a local martingale on this space with $d\langle M\rangle << dt$. Then there exists a Brownain motion, possibly on an extended probability space $(\widetilde{\Omega}, \widehat{\mathcal{F}}, \widetilde{\mathbb{P}})$ and a Brownain motion, B, with a filtration $\widetilde{\mathcal{F}_t}$, such that there is X_t, measurable adapted process, with the property that*

$$\int_0^t X_s dB_s = M_t \quad a.e.\ \widetilde{\mathbb{P}}.$$

Remark 2.4.11. *Here we will assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough such that no extension is needed. For the construction of the extended space, we can use the product space $\Omega \otimes \widetilde{\Omega}$ equipped with the product σ field, $\mathcal{F} \otimes \widetilde{\mathcal{F}}$ and put in together with the product measure. For the filtration, we can use the augmented product filtration where $\overline{\mathcal{F}_t \otimes \widetilde{\mathcal{F}}_t}$ where overline means includes all subsets of null sets and make it right continuous.*

Remark 2.4.12. *Strategy of the proof: It would be great if for some Brownain motion W we have*

$$dW_t = \lambda_t dM_t$$

form some λ_t. Or the second best thing, for some W such that $dW_t = \frac{1}{\sqrt{\frac{d\langle M\rangle_t}{dt}}} dM_t$. Then it would be very close to be a Brownain motion.

Proof. Say $\frac{d\langle M\rangle_t}{dt} = \lambda_t$ and since $\langle M\rangle_t$ is nondecreasing, then λ_t is nonnegative a.e. (ω). Moreover, it is progressively measurable, being the limit of two such things. Let B be an brownain motion independent of M and define

$$dW_t = 1_{\lambda_t>0} \frac{1}{\sqrt{\lambda_t}} dM_t + 1_{\lambda_t=0} dB_t \tag{2.25}$$

which is a local martingale since

$$\int_0^t 1_{\lambda_s>0} \frac{1}{\lambda_s} d\langle M\rangle_s = \int_0^t 1_{\lambda_s\neq 0} ds \leq t.$$

Moreover

$$\langle W\rangle_t = \int_0^t 1_{\lambda_s\neq 0}ds + \int_0^t 1_{\lambda_s=0}ds = t$$

where the first equality is by independence of M and B. So by Levy's Characterization of Brownain motion, we see W is a Brownain motion. Now, multiplying both side of (2.25) by $\sqrt{\lambda_t}$ and integrate to see

$$\int_0^t \sqrt{\lambda_s}dW_s = \int_0^t 1_{\lambda_s\neq 0}dM_t.$$

Here we might seem to be in trouble due to the fact that the integrand on the right hand side is not identically 1, but we are actually done here:

$$M_t = \int_0^t 1_{\lambda_s\neq 0}dM_s + \int_0^t 1_{\lambda_s=0}dM_t \tag{2.26}$$

however,

$$\left\langle \int_0^\cdot 1_{\lambda_s=0}dM_t \right\rangle_t = \int_0^t 1_{\lambda_s=0}\lambda_s dt = 0$$

which means the second term on the right hand side of (2.26) is identically zero. □

Take the idea from this proof, we can easily extend this result into higher dimensional case.

Theorem 2.4.14. *Let $M = (M^1, \dots, M^d)$ be a local martingale such that $\langle M^i, M^j\rangle$ is absolutely continous with repsect to the Lebesgue's measure. Then there exists a Brownain motion W, possibly on a extended probability space, and matrix X with progressively measurable entires such that*

$$M_t = \int_0^t X_s dW_t.$$

Proof. Here we perform the same trick, but a bit of linear algebra is requried to change basis. Let Z_t be defiend as

$$Z^{i,j} = \frac{d}{dt}\langle M^i, M^j\rangle_t$$

which is symmetric. Also, let $x = (x_1, \dots, x_)d) \in \mathbb{R}^d$, then

$$x\cdot Zx = \sum_{1\leq i\leq d}\sum_{1\leq j\leq d} x_i z_t^{i,j} x_j = \left\langle \sum_{i=1}^d xM^i \right\rangle_t \geq 0$$

so it is also nonnegative definite. So there exists Q, Λ where Q is unitary ($Q^T = Q^{-1}$), and Λ is diagonal with entries λ_t^i, such that $Z_t = Q_t\Lambda_t Q_t^T$ where all the entries are progressively measurable, or $Q_t^T Z_t = \Lambda_t Q_t^T$. Now let $dN_t = Q_t^T dM_t$, this makes sense because each row of Q_t^T forms a vector of norm 1. Now consider its quadratic variation:

$$\begin{aligned} d\langle N^i, N^j\rangle_t &= \left\langle \sum_{k=1}^d q_t^{k,i} dM_t^k, \sum_{l=1}^d q^{l,j} dM^l \right\rangle \\ &= \sum_{k=1}^d\sum_{l=1}^d q^{k,i}q^{l,j} d\langle M^k, M^l\rangle_t \\ &= \sum_{1\leq l,k\leq d} q_t^{k,i} z_t^{k,l} q_t^{l,j} dt = \delta_{i,j}\lambda_t^i dt \end{aligned}$$

by diagonalization of Z_t. Again, we have N really wants to be a brownain motion. Now we let $\{B^i\}_1^d$ be d independent Brownain motion and say they are independent of M_t (possibly in an extended probability space) and let

$$dW_t^i = 1_{\lambda_t^i \neq 0} \frac{1}{\sqrt{\lambda_t^i}} dN^i + 1_{\lambda_t^i = 0} dB_t$$

then again W_t is a Brownain motion since

$$\left(1_{\lambda_t^i \neq 0} \frac{1}{\sqrt{\lambda_t^i}} \right) d\langle N^i \rangle_t = 1_{\lambda_t^i \neq 0} dt$$

which takes care of the integrability issue and

$$d\langle W^i \rangle_t = 1_{\lambda_t^i \neq 0} \frac{1}{\lambda_t^i} \lambda_t^i dt + 1_{\lambda_t^i = 0} dt$$

which shows W is brownain motion by Levy's characterization. Now we consider

$$\int_0^t \sqrt{\Lambda_s} dW_s = \int_0^t \sqrt{\Lambda_t} \Lambda' dN_s$$

where $\Lambda' = \left(\delta_{i,j} \frac{1}{\sqrt{\lambda_t^i}} 1_{\lambda_t^i \neq 0} \right)$, so the above equation is really

$$\sqrt{\Lambda_t} dW_t = dN_s$$

but $dN_t = Q^T dM_t$ so $QdN_t = dM_t$, so we obtain

$$Q_t \sqrt{\Lambda_t} dW_t = dM_t$$

□

There is definitely some unsatisfactories of the above representation theorem, even though the martingale is arbitrary, but we have to create a new probability space and new brownain motions. So the natural question to aks is that, if we already have a Brownain motion, and a process is already a martignale with resepct to the Brownain filtration, can we still represent it as integral of some progressively measurable process against the given Brownain motion.

Before answering the question, let's look at a remarkable fact that allow us to identify a concrete dense set of certain L^2 space:

Proposition 2.4.2. *Let B be a Brownain motion and $\mathcal{F}_t$ be its filtration. Then the linear span of the random variables of the following form is dense in $L^2_{\mathbb{C}}(\Omega, \mathcal{F}_\infty, \mathbb{P})$:*

$$\exp\left(i \sum_{j=1}^n \lambda_i \left(B_{t_i} - B_{t_{i-1}} \right) \right)$$

where $0 = t_0 < t_1 < \cdots < t_n$ and λ_i are any real numbers.

Proof. We note that $L^2_{\mathbb{C}}(\Omega, \mathcal{F}_\infty, \mathbb{P})$ is a Hilbert space, so to show the given space is dense, we only have to show that if $Z \in L^2_{\mathbb{C}}(\Omega, \mathcal{F}_\infty, \mathbb{P})$ is that

$$\mathbb{E}\left[Z \exp\left(i \sum_{j=1}^{n} \lambda_i \left(B_{t_i} - B_{t_{i-1}}\right)\right)\right] = 0 \tag{2.27}$$

then $Z = 0$.

For any fixed set of $\{t_i\}_{i=1}^n$, denote $\overline{B}_n = \left(B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}\right)$, we define a measure on $\mathbb{R}^n$, we define a measure ν on $\mathbb{R}^n$ by

$$\nu(F) = \mathbb{E}\left[Z 1_{\overline{B}_n \in F}\right]$$

with some self convincing (possibly by change of notation) we see the left hand side of (2.27) is the fourier transform of ν. By uniqueness of Fourier transform, we see $\nu \equiv 0$ as a measure on $\mathbb{R}^n$, so $Z \equiv 0$ on $\sigma(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$, or $\mathbb{E}[Z1_A] = 0$ on $\sigma(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$ for all finite collections of $\{t_i\}_{1 \le i \le n}$. We make the following definition for the monotone class argument

$$\mathcal{G} = \{A \in \mathcal{F}_\infty : \mathbb{E}[Z1_A] = 0\}$$

then $\mathcal{G}$ contains all σ field of the form $\sigma(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$ which generates $\mathcal{F}_\infty$, so by monotone class thoerem, $Z = 0$ on $\mathcal{F}_\infty$. □

The proof of the following representation theorem shows us why we put the above proposition in this position:

Theorem 2.4.15. *Let B be a Brownain motion with canonical fintration $\mathcal{F}_\infty$ and suppose that $M \in L^2(\mathcal{F}_\infty, \mathbb{P})$. Then there is a unique progressively measurable process, call h_s that is in $L^2(B)$, meaning $\mathbb{E}\left[\int_0^\infty h_t^2 dt\right] < \infty$ such that*

$$M = \mathbb{E}[M] + \int_0^\infty h_s dB_s.$$

Remark 2.4.13. *Now it is clear why the above theorem is useful in here, since we can approximate M with linear combinations of (2.27) which is exponential martingale of a stochastic integral so we can apply Ito's lemma to it.*

Proof. Assume WLOG that M is a centered random variable. Before we forget, let's show uniqueness first. Suppose h, h' both achieve the same result, then

$$\mathbb{E}\left[\left(\int_0^\infty h_t dB_t - \int_0^\infty h'_t dB_t\right)^2\right] = \mathbb{E}\left[\left(\int_0^\infty h_t - h'_t dB_t\right)^2\right] = \mathbb{E}\left[\int_0^\infty \left(h_s - h'_s\right) ds\right] = 0$$

which gives us uniqueness.

For existance, in the light of Proposition 67, we first consider the following: let $h_s^n \triangleq \sum_{i=1}^n \lambda_i 1_{(t_{i-1}, t_i]}$ where $t_0 = 0$ and t_i strictly increasing. Then

$$M_t^n \triangleq \int_0^t h_t^n dB_t = \sum_{i=1}^n \lambda_i \left(B_{t_i} - B_{t_{i-1}}\right)$$

is a true martingale, apply Ito's lemma to the exponential function

$$\exp(M_t^n) = 1 + \int_0^t \exp(M_s^n) h_s^n dB_s + \frac{1}{2}\int_0^t \exp(M_s^n)\left(h_s^n\right)^2 dt$$

so

$$\exp\left(iM_t^n + \frac{1}{2}\int_0^t (h_s^n)^2\, ds\right) = 1 + \int_0^t \exp\left(M_s^n\right) h_s^n dB_s$$

so M_∞ can be written as the required form, namely,

$$\exp(iM_\infty^n) = \exp\left(-\frac{1}{2}\int_0^\infty (h_s^n)^2\, ds\right)\left(1 + \int_0^\infty \exp(M_s^n) h_s^n dB_s\right) \tag{2.28}$$

where we note that h_s^n is deterministic function on $\mathbb{R}^+$. Now, let $M \in L^2_{\mathbb{C}}(\mathcal{F}_\infty, \mathbb{P})$, then by density it can be written as infinite sum of $\exp(iM_\infty^n)$'s that converges in L^2, which is still the above form. □

Since we can represent $L^2(\mathcal{F}_\infty)$ random variable via stochastic integration with respect to Brownian motion, we can also write any L^2 bounded martingales in such term: M_t be a martingale, then since L^2 boudned, hence uniformly integrable hence closed hence there is M_∞ such that $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]$ and let h_t be such that

$$M_\infty = \mathbb{E}[M_\infty] + \int_0^\infty h_s dB_s$$

then the second term on the right hand side is a true martinale, hence

$$M_t = \mathbb{E}[M_\infty|\mathcal{F}_t] = \mathbb{E}[M_\infty] + \int_0^t h_s dB_s.$$

Now, with the help of localization, we can prove the same thing for continous local martingales. We orgnize this into a theorem:

Theorem 2.4.16 (Representation of $\mathcal{M}^{c,loc}$ as integration against BM)**.** *Let B be a Brownain motion with filtration $\mathcal{F}_t$, and let $M \in \mathcal{F}^{c,loc}$ with the same filtration. Then there is a unique progressively measurable process, cal h_t in $L^{loc}(B)$, meaning $\int_0^t h_s^2 ds < \infty$ a.e., such that*

$$M_t = C + \int_0^t h_s dB_s \quad a.e.$$

Girsanov's Theorem

Since we used exponential martingales in the previous subsection extensively, we might as well look at another representation theorem involving exponential martignales and set us up for Girsanov's theorem

Theorem 2.4.17 (Real exponential mtgs)**.** *Let $Z > 0$ be a real valued process, then it is a local martingale if and only if there is another local martignale M such that*

$$Z = \mathcal{E}(M)_t \triangleq \exp\left(M - \frac{1}{2}\langle M\rangle\right)$$

where M is unique. Furthermore, if N is another local martingale, then

$$\langle N, M\rangle_t = \int_0^t \frac{1}{Z_s} d\langle Z, N\rangle_s$$

Proof. For the first assertion, the converse is a simply application of Ito's lemma: let M be a local martingale, then

$$Z_t \triangleq \mathcal{E}(M)_t = \exp(M_t - \frac{1}{2}\langle M\rangle_t) = 1 + \int_0^t \mathcal{E}(M)_s dM_s$$

which is a local martiangle.

Now for the converse, suppose Z is a striclty positive local martingale, then consider

$$\log(Z_t) = \log(Z_0) + \int_0^t \frac{1}{Z_s} dZ_s \tag{2.29}$$

which is also a local martingale by positivity of Z. Take the exponential we'd see the representation.

Now for the last assertion, we note that $M_t = \log(Z_0) + \int_0^t \frac{1}{Z_s} dZ_s$, and now the equility is obviuos. □

We recall that positive local martingales with $M_0 \in L^1$ is a super martingale by conditional Fatou's lemma. Furthermore, if a super martingale is a true martingale if and only if it has constant expectation since

$$\mathbb{E}\left[|M_s - \mathbb{E}\left[M_t|\mathcal{F}_s\right]|\right] = \mathbb{E}\left[M_s - \mathbb{E}\left[M_t|\mathcal{F}_s\right]\right] = \mathbb{E}[M_s] - \mathbb{E}[M_t] = 0.$$

Therefore, a positive local martinagle (including a exponential local martingale) is a true martingale if and only if it has constant expectation.

Now suppose Z is a true positive martingale on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to the filtration $\{\mathcal{F}_t\}_t$. We define a new meausere by the following formula:

$$\mathbb{P}_T(A) = \mathbb{E}\left[1_A Z_T\right] \quad \forall A \in \mathcal{F}_t \text{ for } 0 \le t \le T.$$

We note that this definition is consistance in the sense that for a fixed s and $A \in \mathcal{F}_s$, and for all $t \in [0, T]$, we have

$$\mathbb{P}_t(A) = \mathbb{P}_T(A)$$

by the Martingale property.

The following theorem tells us the relation between the measure $\mathbb{P}_T$ and the original measure $\mathbb{P}$:

Lemma 2.4.2. *Let $\mathbb{P}_T, Z$ be defined as above, then for any $Y \in L^1$, then*

$$\mathbb{E}_T\left[Y|\mathcal{F}_s\right] = \frac{1}{Z_s}\mathbb{E}\left[YZ_t|\mathcal{F}_s\right]$$

both in $\mathbb{P}, \mathbb{P}_T$.

Proof. Here we need to show that if we take the expectation $\mathbb{E}_T$ and $\mathbb{E}$ on both sides on any set $A \in \mathcal{F}_s$, they agrees separately. So

$$\mathbb{E}_T\left[1_A \mathbb{E}_T\left[Y|\mathcal{F}_s\right]\right] = \mathbb{E}_T\left[1_A Y\right] = \mathbb{E}\left[YZ_T 1_A\right]$$

and

$$\mathbb{E}_T\left[1_A \frac{1}{Z_s}\mathbb{E}\left[YZ_T|\mathcal{F}_s\right]\right] = \mathbb{E}\left[\frac{Z_T}{Z_s}\mathbb{E}[YZ_T 1_A|\mathcal{F}_s]\right] = \mathbb{E}[YZ_T 1_A]$$

where the last equality is obtained by taking conditional expectation inside conditioned on $\mathcal{F}_s$, so $\mathbb{E}_T$ part is taken care of.

For the second part, we don't actually have to take the expectation $\mathbb{E}$ on both sides (I can't do it anyway), but we observe that they are mutually absolutely continous due to the assumption that $Z_T > 0$, so the equality also holds in $\mathbb{P}$. □

Now we are ready for the first part of our big theorems. But before that, let's recall that since Z_t is a positive martingale, then it can be written as

$$Z_t = \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right) \tag{2.30}$$

Where M is a local martiangle here. However, by the stochastic representation of local martingales in terms of Brownain motions, there exists $\vec{X}_t$ and $\vec{W}_t$ of d dimensional progressively measurable processes and Brownain motion respectively, possibly in an extended probability space. So from now on, we will assume Z_t is of the following form:

$$Z_t = \exp\left(\sum_{i=1}^{d}\int_0^t X_s^i dW_s^i - \frac{1}{2}\int_0^t \|\vec{X}_s\| ds\right)$$

where the norm is the usual $\mathbb{R}^d$ Euclidean norm, and we will assume, for now, Z_t is a true martingale. Also, since the above form depens on X, we denote it as $Z_t(X)$ to record this fact.

Theorem 2.4.18. *Suppose $Z(X)$ is a true martingale of the form in (2.30) with respect to $\mathcal{F}_t$ under the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\widetilde{\mathbb{P}}_T$ be defined as before, $\widetilde{P}_T(A) = \mathbb{E}[1_A Z_T]$ for all $A \in \mathcal{F}_t$ for $0 \leq t \leq T$. Let M be a local martingale with respect to the same filtration, then*

$$\widetilde{M}_t \triangleq M_t - \sum_{i=1}^{d}\int_0^t X_s^i d\langle W^i, M\rangle_s \tag{2.31}$$

is a local martingale under $\widetilde{\mathbb{P}}_T$. Also, if N is another local martiangle and

$$\widetilde{N}_t = N_t - \sum_{i=1}^{d}\int_0^t X_s^i d\langle W^i, N\rangle_s$$

then

$$\langle N, M\rangle_t = \langle \widetilde{N}, \widetilde{M}\rangle_t \quad \text{a.e. } \mathbb{P} \text{ and } \widetilde{\mathbb{P}}_T \tag{2.32}$$

Proof. For the first part, it is pretty obvious what to do now: we apply the previous lemma to look at the conditional expectation:

$$\widetilde{E}_T\left[\widetilde{M}_t|\mathcal{F}_s\right] = \frac{1}{Z_s(X)}\mathbb{E}\left[\widetilde{M}_t Z_t|\mathcal{F}_s\right] \tag{2.33}$$

Now, the conditional expectations are note neccessarily defined since M is only assumed to be a local martingale. So here we assume WLOG that M is a unifromly integrable local martingale (or even bounded). Here we apply the integration by parts formula to $\widetilde{M}_t Z_t$ under the measure $\mathbb{P}$:

$$d\left(\widetilde{M}_t Z_t(X)\right) = Z_t(X)d\widetilde{M}_t + \widetilde{M}_t dZ_t(X) + d\langle M, Z(X)\rangle_t \tag{2.34}$$

but we recall that

$$dZ_t(X) = \sum_{i=1}^{d} Z_t(X)dW_t^i$$

and note that $Z_t(X)d\widetilde{M}_t$ admits a stochastic integral part and a Lebesgues-Stieltjes' integral part, and the latter is

$$-\sum_{i=1}^{d} X_t^i Z_t(X) d\langle M, W^i\rangle_t \tag{2.35}$$

and expand $d\langle M, Z(X)\rangle_t$ to get

$$d\langle M, Z(X)\rangle_t = \sum_{i=1}^{d} Z_t(X) X_t^i d\langle W^i, M\rangle_t$$

which cancels with (2.35), so the remaining part of (2.34) are all stochastic integrals:

$$\widetilde{M}_t Z_t(X) = Z_0(X) M_0 + \int_0^t Z_t(X) dM_t + \sum_{i=1}^{d} \int_0^t \widetilde{M}_s Z_s(X) X_s^i dW_s^i$$

which is a local martingale in $\mathbb{P}$. then use the relation in (2.33) to see that

$$\widetilde{E}_T\left[\widetilde{M}_t | \mathcal{F}_s\right] = \widetilde{M}_s \quad \mathbb{P}, \widetilde{\mathbb{P}}_T \text{ a.e.}$$

so $\widetilde{M}$ is a local martiangle under $\widetilde{\mathbb{P}}_T$.

For the second assertion, we can use the fact that the corss-variation of two semi-martingales is equal to the quadratic variation of their local martingale parts. Here, $\widetilde{M}, \widetilde{N}$ are local martingales under $\widetilde{\mathbb{P}}_T$ and are semi-martiangles under $\mathbb{P}$, similar for M, N: local mgt under $\mathbb{P}$ and semi-martingale under $\widetilde{\mathbb{P}}_T$.

Or, one can use the definition of quadratic variation for local martingales and play around like [KS12] did. □

Here comes the theorem with the name of this subsection:

Theorem 2.4.19 (Girsanov's Theorem)**.** *Same setting as above, and suppose $Z(X)$ is a true martingale with $\vec{W}$ as the Brownian motion occuring in $Z(X)$, then the process:*

$$\widetilde{W}_t^i \triangleq W_t^i - \int_0^t X_s^i ds$$

is a Brownain motion under $\left(\Omega, \mathcal{F}, \widetilde{\mathbb{P}}_T\right)$.

Proof. Replacing the M in (2.31) of Theorem 83 with W^i to see

$$\widetilde{W}_t^i = W_t^i - \int_0^t X_s^i ds$$

is a local martiangle, and now we use (2.32) to see that the quadratic variation of $\widetilde{W}$ is t under $\widetilde{\mathbb{P}}_T$. □

Now we take a look at an application of Girsanov's theorem. Recall that the distribution of Brownain hitting time at level $a \neq 0$ is

$$f_{\tau_a}(t) = \frac{|a|}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2t}\right)$$

this is due to symmetry of Brownain motion since when $a > 0$ we don't have the absolute value.

Now let W be a Brownain motion and $\widetilde{W}_t = W_t - \mu t$, then it is a Brownain motion under the measure

$$\mathbb{P}^\mu(A) = \mathbb{E}\left[1_A Z_t\right], \quad A \in \mathcal{F}_t$$

where $Z_t = \exp\left(\mu W_t - \frac{1}{2}\mu^2 t\right)$, then we call $W_t = \widetilde{W}_t + \mu t$ the Brownain motion with drift μ under $\mathbb{P}^\mu$. Here, Girsanov's theorem gives us an easy way to calculate the distribution the hitting time of a Brownain motion with drift:

$$\begin{aligned} \mathbb{P}^\mu\left[\tau_b \le t\right] &= \mathbb{E}\left[1_{\tau_b \le t} Z_t\right] \\ &= \mathbb{E}\left[1_{\tau_b \le t} \mathbb{E}\left[Z_t | \mathcal{F}_{t\wedge\tau_b}\right]\right] \\ &= \mathbb{E}\left[1_{\tau_b \le t} Z_{t\wedge\tau_b}\right] \\ &= \mathbb{E}\left[1_{\tau_b \le t} Z_{\tau_b}\right] \\ &= \mathbb{E}\left[1_{\tau_b \le t} \exp\left(\mu b - \frac{1}{2}\mu^2 \tau_b\right)\right] \end{aligned} \tag{2.36}$$

where we used optional sampling theorem to the bounded stopping time $\tau_n \wedge t$. Writting the last expression in terms of the density function we get

$$\mathbb{P}^\mu(\tau_b \le t) = \int_0^t \exp\left(\mu b - \frac{1}{2}\mu^2 s - \frac{b^2}{2s}\right) \frac{b}{\sqrt{2\pi s^3}} ds$$

So the density of τ_b udner $\mathbb{P}^\mu$ is

$$\mathbb{P}\left[\tau_b \in dt\right] = \frac{d}{dt}\mathbb{P}\left[\tau_a \le t\right] = \frac{b}{\sqrt{2\pi t^3}} \exp\left(-\frac{(\mu t - b)^2}{2t}\right)$$

Now we look at (2.36), and take $t \to \infty$ to see that

$$\begin{aligned} \mathbb{P}^\mu\left[\tau_b < \infty\right] &= \mathbb{E}\left[1_{\tau < \infty} \exp\left(-\frac{1}{2}\mu^2 \tau_b\right)\right] \exp\left(\mu b\right) \\ &= \exp\left(\mu b - |\mu b|\right) \end{aligned}$$

where the last equality is obtained by some calculation (oh, well).

The following Theorem is a direct consequence of (2.8)

Theorem 2.4.20. *Let $B, Z, \mathbb{P}, \mathbb{P}^\mu$ be defined as above, then the neccessary and sufficient condition for the Wald's idenity*

$$\mathbb{E}\left[\exp\left(\mu W_T - \frac{1}{2}\mu^2 T\right)\right] = 1$$

to holds for the stopping time T is

$$\mathbb{P}^\mu\left[T < \infty\right] = 1.$$

In particular, if $b, \mu \in \mathbb{R}$, and $\mu b \ge 0$, then the Wald's identity above holds for the stopping time

$$S_b = \inf\{t \ge 0 : W_t - \mu t = b\}.$$

Here is another condition about Wald's identity for exponential martiangle:

Theorem 2.4.21. *Let τ be a stopping time of the brownain motion B, then*

$$\mathbb{E}\left[e^{\frac{\tau}{2}}\right] < \infty \Rightarrow \mathbb{E}\left[\exp\left(B_\tau - \frac{1}{2}\tau\right)\right] = 1.$$

Proof. By the previous thoerem, we see that $\tau_b = \inf\{s \geq 0 : B_t - t = b\}$ satisfies the Wald's idenity, so let τ satisfies the condition given, then

$$1 = \mathbb{E}\left[\exp\left(B_{\tau\wedge\tau_b} - \frac{1}{2}\tau \wedge \tau_b\right)\right] = \mathbb{E}\left[\exp\left(B_{\tau_b} - \frac{1}{2}\tau_b\right) 1_{\tau \geq \tau_b}\right] + \mathbb{E}\left[\exp\left(B_\tau - \frac{1}{2}\tau\right) 1_{\tau < \tau_b}\right]$$

by the condition given, we see when we take $b \to \infty$, the first term vanishes (also by DCT) so we only have to worry about the last term. However, note that the integrand of the last term is monotone w.r.t. b, so by monotone convergence theorem, we have the desired result. □

Now we state and prove a condition for which $\mathcal{E}(M)$ is a uniformly integrable martingale:

Theorem 2.4.22. *Suppose M is a local martiangle, then*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\langle M\rangle_\infty\right)\right] < \infty \Rightarrow \mathbb{E}\left[\mathcal{E}_t(M)\right] = 1.$$

Proof. By the condition given, $\langle M\rangle_\infty < \infty$ a.e. Also recall that we can represent M with a time changed Brwonain motion, namely,

$$B_t = M_{\tau_t} \quad \text{on } [0, \langle M\rangle_\infty)$$

is a brownain motion with respect to the filtration $\mathcal{G}_t \supset \mathcal{F}_{\tau_t}$ where $\tau_t = \inf\{s \geq 0 : \langle M\rangle_s = t\}$ and $\mathcal{F}_t$ is the original filtration of M. Then we claim $\langle M\rangle_\infty$ is a stopping time with repsect to $\mathcal{G}_t$. Consider

$$\{\langle M\rangle_\infty \leq t\} \cap \{\tau_t \leq s\} = \emptyset \in \mathcal{F}_s$$

$\{\tau_t \leq s\}$ means $\langle M\rangle$ reaches t before s, so the intersection is the empty set. Then use the previous theorem to get the desired conclusion. □

Bibliography

[Bil08] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[CZ01] Kai Lai Chung and Kailai Zhong. *A course in probability theory*. Academic press, 2001.

[DKM+09] Robert C Dalang, Davar Khoshnevisan, Carl Mueller, David Nualart, and Yimin Xiao. *A minicourse on stochastic partial differential equations*, volume 1962. Springer, 2009.

[DPZ14] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[Fol99] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

[KK97] Olav Kallenberg and Olav Kallenberg. *Foundations of modern probability*, volume 2. Springer, 1997.

[KS12] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[Law18] Gregory F Lawler. *Introduction to stochastic processes*. Chapman and Hall/CRC, 2018.

[LG16] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*. Springer, 2016.

[MS13] Camil Muscalu and Wilhelm Schlag. *Classical and Multilinear Harmonic Analysis: Volume 1*, volume 137. Cambridge University Press, 2013.

[RKS+96] Sheldon M Ross, John J Kelly, Roger J Sullivan, William James Perry, Donald Mercer, Ruth M Davis, Thomas Dell Washburn, Earl V Sager, Joseph B Boyce, and Vincent L Bristow. *Stochastic processes*, volume 2. Wiley New York, 1996.