Martingale Convergence Theorem and Stochastic Optimization

Grace Yin

December 11, 2020

1 Background

The topic of this project is to explore Martingale Convergence Theorem and its application in Stochastic Optimization problems. Recently, research on martingales becomes heated, since martingales play a crucial role in stochastic optimization, especially in the proof of convergence rates for stochastic approximation and stochastic gradient. Moreover, since martingales have many good properties, they are useful in exploring time series problems, network analysis algorithms and other machine learning extensions as well. Therefore, my project focuses on martingales and their related theorems.

The main result I proved in this project is the proof of Martingale Convergence Theorem (MCT). To understand it better, I also proved the related theorems including Doob's Upcrossing Inequality and Doob's Martingale Inequality. Then I introduced Robbins-Siegmund Theorem and conducted some literature review to discuss how can Martingale Convergence Theorem be applied in stochastic optimization algorithms.

Before exploring deeply, I am going to provide the basic ideas and definitions in martingales[Dur19] [Blo19].

Definition 1. *Consider the probability space $(\Omega, \mathcal{F}, P)$. Let $T = \mathbb{Z}_+ = \{0, 1, 2, \cdots\}$. A family of σ-algebras $\mathcal{F} = (\mathcal{F}_t)_{t\in T}$ is a filtration if it satisfies that $\forall t \in T, t_1 \leq t_2 \implies \mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}$. In other words, $(\mathcal{F}_t)_{t\in T}$ is an increasing collection of sub σ-algebras.*

$X = \bar{X} = (X_t)_{t\in T}$ is a Stochastic Process if $\forall t \in T$, X_t is $\mathcal{F}$-measurable. X is $(\mathcal{F}_t)$-adapted if $\forall t \in T$, X_t is $\mathcal{F}_t$-measurable. Now we can define a martingale now.

Definition 2. *An adapted family $(X_t, \mathcal{F}_t)_{t\in T}$ is called a **martingale** if:*

1. $\forall t \in T$, $\mathbb{E}(|X_t|) < \infty$
2. $\forall s < t, \mathbb{E}(X_t|\mathcal{F}_s) = X_s$ *almost surely.*

Note that we also say (X_t) is an $\mathcal{F}$- or $(\mathcal{F}_n)$-martingale.

An adapted family $(X_t, \mathcal{F}_t)_{t\in T}$ is called a **supermartingale** if (1) $\forall t \in T$, $\mathbb{E}(|X_t|) < \infty$ (2) $\forall s < t, \mathbb{E}(X_t|\mathcal{F}_s) \leq X_s$ **a.s.**.

An adapted family $(X_t, \mathcal{F}_t)_{t\in T}$ is called a **submartingale** if (1) $\forall t \in T$, $\mathbb{E}(|X_t|) < \infty$ (2) $\forall s < t, \mathbb{E}(X_t|\mathcal{F}_s) \geq X_s$ **a.s.**.

Here are two examples of martingales in [Dur19]:

Example 1. *Let $\xi_1.\xi_2, \cdots$ be independent and identically distributed (i.i.d) random variables. Let $S_n = S_0 + \xi_1 + \cdots + \xi_n$ where S_0 is a constant. $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$ for $n \geq 1$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then if $\mu = \mathbb{E}\xi_i = 0$, S_n is a martingale with respect to $\mathcal{F}_n$*

Proof. $S_n \in \mathcal{F}_n, \mathbb{E}S_n < \infty$. By the linearity of conditional expectation, $\mathbb{S}_{\ltimes+\nVdash}|\mathcal{F}_{\ltimes} = \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = S_n + \mathbb{E}(\xi_{n+1}) = S_n$ □

Example 2. *Let Y_n be i.i.d random variables with $P(Y_n = 0) = P(Y_n = 2) = \frac{1}{2}$. Then $X_n = \prod_{j=1}^n Y_j, n \geq 1, X_0 = 1$ is a martingale with respect to $\mathcal{F}_n$.*

Proof. $X_n \in \mathcal{F}_n, \mathbb{E}X_n < \infty$.

$$\begin{aligned}\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(X_n Y_{n+1}|\mathcal{F}_n) &= X_n \mathbb{E}(Y_{n+1}\mathcal{F}_n) \text{ (because } X_n \text{ is } \mathcal{F}_n\text{-measurable)} \\ &= X_n \mathbb{E}(Y_{n+1}) \text{ (because } \mathcal{F}_n \& Y_{n+1} \text{ are independent)} \\ &= X_n \left(0 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2}\right) \\ &= X_n\end{aligned}$$

So X_n is $\mathcal{F}_n$-martingale. □

One of the most important theorems with martingales is Martingale Convergence Theorem. I will introduce and prove it in the following section.

2 Martingale Convergence Theorem

2.1 Martingale Convergence Theorem 1

[Blo19]

Theorem 1. *Martingale Theorem 1 Let $X = (X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ be a submartingale. If $\mathbb{E}X_n^+ < \infty$, then X_n converges almost surely to a limit X_∞ with $\mathbb{E}|X_\infty| < \infty$.*

Note that the fact that X is bounded in L^1 and its limit $X_\infty \in L^1$ does not imply that $X_n \to X_\infty$ in L^1. The proof of the Martingale Convergence Theorem demands more sophisticated analysis. Before proving it, I will introduce the concepts of **predictable process, discrete stochastic integral, stopping time** and discuss the proof of **Doob's Upcrossing Inequality**. These are constructive tools for proving the Martingale Convergence Theorem.

2.2 Predictable Process

Definition 3. *Let $\mathcal{F}_n$ be a filtration. A process $(H_n : n \geq 1)$ is $(\mathcal{F}_n)-$ **predictable** if $\forall n \in N$, H_n is $\mathcal{F}_n$-measurable $(H_n \in \mathcal{F}_{n-1})$.*

It can be interpreted to that the value of H_n can be predicted based on the previous information before time $n - 1$.

Definition 4. *If $(H_n : n \in \mathbb{N})$ and $(X_n : n \in \mathbb{Z}_{+kk})$ are stochastic process, then*

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

This is also called discrete stochastic integral $\int_0^t HdX$.

Theorem 2. *Let $H \geq 0$ be bounded and $(\mathcal{F}_n)$-predictable. If $(X_n, \mathcal{F}_n)_{n \geq 0}$ be a supermartingale (respected with martingale of submartingale), then so is $(H \cdot X)$.*

The core step of the proof is:

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

2.3 Stopping Time

Definition 5. *For a probability space $(\Omega, \mathcal{F}, P)$, a random variable T which maps from $\Omega \to \mathbb{Z}_+ \cup \{0\}$ is a $\mathcal{F}_n$-stopping time if $\{T \leq n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{Z}_+$.*

Note that the stopping time can be explained as "you know it when T occurs". To be detailed, for a random process $(X, \mathcal{F})$ with $X = \{X_n\}_{n\geq 0}$. T is a stopping time. Then for all $n \leq T$, $X_n^T = X_{n\wedge T} = X_n$while for $n > T$, $X_n^T = X_T$. In other words, when we stop the process, the random variables do not be changed any more.

Corollary 1. *If T is an $(\mathcal{F}_n)$-stopping time and (X_n) is a $(\mathcal{F})_n-$supermatingale, then so is $X_{n\wedge T}$.*

$(X_n, \mathcal{F}_n)$ is also called a "stopped process"; its proof is based on applying Theorem 2.

Proof. Define $H_n = \mathbb{1}_{\{n\leq T\}}$, $\forall n \in \mathbb{N}$. Then

$$\{n \leq T\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$$

By Theorem 2, H is predictable implies that $(H \cdot X)$ is a submartingale.

$$\begin{aligned}(H \cdot X)_n &= \sum_{k=1}^{n} \mathbb{1}_{\{k\leq T\}}(X_k - X_{k-1})\\ &= \sum_{k=1}^{\infty} \mathbb{1}_{\{k\leq n\}}\mathbb{1}_{\{k\leq T\}}(X_k - X_{k-1})\\ &= \sum_{j=1}^{n\wedge T}(X_k - X_{k-1}) = X_n^T - X_0^T \text{ is a supermartingale}\end{aligned}$$

Therefore, $X_n^T = X_{n\wedge T}$ is a supermatingale. □

Understanding the proof of Corollary 1 is beneficial for understanding the predictive process and stopping time. I am going to introduce Doob's Upcrossing Inequality, which helps prove the Martingale Convergence Theorem. This concept is scintillating and beautiful.

2.4 Doob's Upcrossing Inequality

Definition 6. *$T_0 = 0$, for $j \geq 0$, let $S_{j+1} = \min\{n > T_j : X_n \leq a\}$ be the first time after T_j that $X_n \leq a$, and $T_{j+1} = \min\{n > S_{j+1} : X_n \geq b\}$ be the first time after S_{j+1} that $X_n \geq b$. Both S_j and T_j are stopping times. Then the number of upcrossings of (a, b) can be defined as*

$$U_n(a, b) = \min\{k : T_k \leq n\} = \sum_{k=1}^{\infty} \mathbb{1}_{(0,n]} \circ T_k$$

Figure 1 in [Cin11] demonstrates the upcrossing time of (a, b) are $T_1, T_2, \cdots$

Theorem 3. *Doob's Upcrossing Inequality If $(X_n, \mathcal{F}_n)$is a submartingale, $a < b$ and $U_n(a, b)$ is the upcrossing number of $[a, b]$, then*

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

Proof. Define $Y_n = a + (X_n - a)^+$, then Y_n is a submartingale, and

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

Here we replace X_n with Y_n. Notice that

$$Y_n - Y_0 = (H \cdot Y)_n + ((1 - H) \cdot Y)_n$$

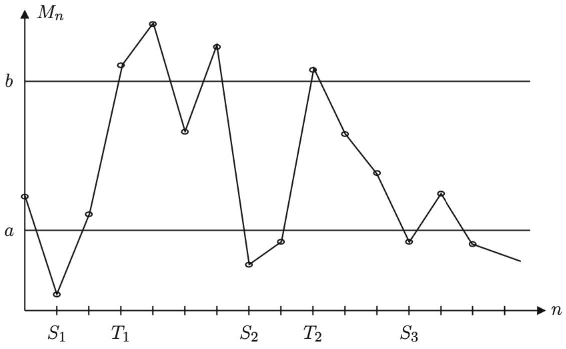

Figure 1: Upcrossing Time of (a,b)

From Theorem 2,

$$\mathbb{E}[((1-H)\cdot Y)_n] \geq \mathbb{E}[((1-H)\cdot Y)_0] \implies \mathbb{E}(H\cdot Y)_n \leq \mathbb{E}(Y_n - Y_0),$$

which completes the proof. □

here are two things that we need to be meticulous about. First, Y_n is a submartingale because $x \mapsto (x-a)^+$ is convex and non-decreasing. Then we can apply Jesen's inequality:

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

In addition, the reason to replace X_n by Y_n is that Y_n upcrosses (a,b) the same number of times as that X_n does, and the final incomplete upcrossing is always non-negative. Therefore, we can use Y_n instead of X_n in the proof.

Hence, we can use Upcrossing Inequality as a tool to prove the Martingale Convergence Theorem 1.

2.5 Proof of MCT 1

Proof. Since $X_n \in L^1$,

$$\mathbb{E}[X_n^+] \leq \mathbb{E}[|X_n|] = 2\mathbb{E}[X_n^+] - \mathbb{E}[X_n] \leq 2\mathbb{E}[X_n^+] - \mathbb{E}[X_0]$$

and

$$\sup_n \mathbb{E}[X_n^+] < \infty \iff \sum_n \mathbb{E}[|X_n|] < \infty$$

Define

$$\begin{aligned} A &= \{\omega \in \Omega : X_n(\omega) \nrightarrow X(\omega)\} \\ &= \{\omega \in \Omega : \lim\inf_{n\to\infty} X_n(\omega) < \lim\sup_{n\to\infty} X(\omega)\} \\ &= \bigcup_{a,b\in\mathbb{Q},a<b} \{\omega \in \Omega : \lim\inf_{n\to\infty} X_n(\omega) < a < b < \lim\sup_{n\to\infty} X(\omega)\} \end{aligned}$$

Also, note that

$$\{\omega \in \Omega : \lim\inf_{n\to\infty} X_n(\omega) < a < b < \lim\sup_{n\to\infty} X(\omega)\} \subset \{\omega \in \Omega : U_\infty(a,b)(\omega) = \infty\}$$

where $U_\infty(a,b) = \lim_n U_n(a,b)$ and $U_n(a,b)$ is the number of upcrossings of(a,b). Apply Monotone Convergence Theorem and Doob's Upcrossing Inequality,

$$\mathbb{E}[U_\infty(a,b)] = \lim_n \mathbb{E}[U_n(a,b)] \le \frac{\sup_n \mathbb{E}[(X_n - a)^+]}{b-a} \le \frac{\sup_n \mathbb{E}[X_n^+] + |a|}{b-a} < \infty$$

Also, note that

$$\mathbb{E}[U_\infty(a,b)] < \infty \implies P(U_\infty(a,b) = \infty) = 0$$

Observe that the sequence of sets $\{\omega \in \Omega : \liminf_{n\to\infty} X_n(\omega) < a < b < \limsup_{n\to\infty} X(\omega)\}$ is countable,

$$P(A) \le \sum_{a<b, a,b\in\mathbb{Q}} P\{\omega \in \Omega : \lim \inf_{n\to\infty} X_n(\omega) < a < b < \lim \sup_{n\to\infty} X(\omega)\} = 0$$

Hence,

$$\lim_{n\to\infty} X_n = \lim \sup_{n\to\infty} X_n = \lim \inf_{n\to\infty} X_n = X_\infty \text{ almost surely.}$$

This completes the proof of Martingale Convergence Theorem 1. □

Throughout my analysis, I get to appreciate the beauty within this proof as we can regard "not convergence" as an event, and apply Upcrossing Inequality to demonstrate the probability of every single event is 0 then. I feel that whereas mathematicians whose primary focus is analysis incline to think questions from the perspective of probability distributions, mathematicians with a focus on probability tend to tackle identical questions based on events.

Based on the proof of the Martingale Convergence Theorem (1), I want to leave two exercises, which help understand Martingale Convergence Theorem. The solutions to exercises can be found in the Appendix.

Exercise 1. *Let X_n be a $(\mathcal{F}_n)$-martingale with bounded increments. Martingale with bounded increments mean that there exists $M > 0$ such that $\sup_n |X_{n+1} - X_n| \le M$. Then with probability 1:*

- $\liminf_n X_n = -\infty$, $\limsup_n X_n = +\infty$
- $\lim_n X_n$ *exists and it is finite*

Exercise 2. *Prove the second Borel-Cantelli Lemma: Let $\mathcal{F}_n$ be a filtration. $A_n \in \mathcal{F}_n$ is a sequence of events, Then*

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

I am going to introduce the Martingale Convergence Theorem 2, which furthers Martingale Convergence Theorem 1.

2.6 MCT 2

Theorem 4. *Let $X = (X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be a martingale. If $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable, then $X_n \to X_\infty$ in L^1* ***almost surely****. Moreover, $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ for all $n \in \mathbb{N}$.*

Proof. Recall that a collection of random variables $\{X_i, i \in I\}$ is uniformly integrable if

$$\lim_{M\to\infty} \sup_{i\in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i|>M\}}] = 0$$

and

$$\{X_i\} \text{ is uniformly integrable} \implies \{X_i\} \text{ is } L^1 \text{ bounded}$$

In addition, if X_n is a submartingale, the following statements are equivalent:

- $\{X_n : n \ge 0\}$ is uniformly integrable.

- $X_n \to X_\infty$ almost surely and $X \in L^1$
- $X_n \to X_\infty$ in L^1

I am not gong to prove the above lemmas because we have already leant in class at a certain point. Therefore, to prove MCT 2, the only part we left is to show $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$.
Since $X_n \to X_\infty$ in L^1, we have

$$|\mathbb{E}X_m\mathbb{1}_A - \mathbb{E}X_\infty\mathbb{1}_A| \le \mathbb{E}|X_m\mathbb{1}_A - X_\infty\mathbb{1}_A| \le \mathbb{E}|X_m - X_\infty| \to 0$$

Therefore,

$$\mathbb{E}(X_m\mathbb{1}_A) \to \mathbb{E}(X_\infty\mathbb{1}_A)$$

Since $(X_n, \mathcal{F}_n)$ is a martingale, $\mathbb{E}(X_m|\mathcal{F}_n) = X_n, \forall m > n$, and it implies that

$$\mathbb{E}(X_n\mathbb{1}_A) = \mathbb{E}[X_m\mathbb{1}_A]$$

Recall that $\mathbb{E}(X_m\mathbb{1}_A) \to \mathbb{E}(X_\infty\mathbb{1}_A)$, we can get that

$$\mathbb{E}[X_n\mathbb{1}_A] = \mathbb{E}[X_\infty\mathbb{1}_A], \forall A \in \mathcal{F}$$

Therefore,

$$\mathbb{E}[X_\infty|\mathcal{F}_n] = X_n$$

□

Overall, the proof of second Martingale Convergence Theorem is more about the understanding of martingale properties and the equivalence of the three properties of martingales: (i) uniformly integrable, (ii) convergence almost surely and in L^1, (iii) convergence in L^1.

2.7 Doob's Martingale Inequality

Theorem 5. *If $X = (X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a martingale in L^p, where $p \ge 1$. Then for $c > 0$,*

$$P\{\max_{k\le n} |X_k| > c\} \le \frac{1}{c^p}\mathbb{E}[|X_n|^p]$$

This is a more graceful extended version of Doob-Kolmogorov inequality, which. is

$$P(\max_{1\le m\le n} X_n \ge c) \le \frac{\mathbb{E}[X_n^2]}{c^2}$$

These results are commonly used in Stochastic Optimization problems because in a time interval $[1, n]$, the worst case deviation of a submartingale is proportional to the value of X_n. Since these inequalities are widely applied in many stochastic optimization algorithms, I feel the necessity to prove it. To simplify the proof process, I will first prove the Doob-Kolmogorov Inequality, then use the result to show Theorem 5.

Proof. Define $A = \{\max_{1\le m\le n} X_m \le c\}$ and $B_m = \{\max_{1\le i\le m_1} X_i \le c, X_m \ge c\}$. This can be explained as B_m is the event that the submartingale X_n touches c for the first time at time m. It is obvious that the whole space $\Omega = A \cup (\cup_{1\le m\le n} B_m)$, and A, B_m are exclusive, so

$$\mathbb{E}[X_n^2] = \mathbb{E}[X_n^2\mathbb{1}_A] + \sum_{1\le m\le n} \mathbb{E}[X_n^2\mathbb{1}_{B_m}] \ge \sum_{1\le m\le n} \mathbb{E}[X_n^2\mathbb{1}_{B_m}]$$

Note that

$$\begin{aligned}\mathbb{E}[X_n^2\mathbb{1}_{B_m}] &= \mathbb{E}[(X_n - X_m + X_m)^2\mathbb{1}_{B_m}] \\ &= \mathbb{E}[(X_n - X_m)^2\mathbb{1}_{B_m}] + 2\mathbb{E}[(X_n - X_m)X_m\mathbb{1}_{B_m}] + \mathbb{E}[X_m^2\mathbb{1}_{B_m}]\end{aligned}$$

Observe that $\mathbb{E}[(X_n - X_m)^2 \mathbb{1}_{B_m}] \geq 0$, and

$$\mathbb{E}[X_m^2 \mathbb{1}_{B_m}] = X_m^2 \mathbb{E}[\mathbb{1}_{B_m}] \geq c^2 P(B_m)$$

Apply tower property,

$$\begin{aligned}\mathbb{E}[(X_n - X_m)X_m \mathbb{1}_{B_m}] &= \mathbb{E}[\mathbb{E}[(X_n - X_m)X_m \mathbb{1}_{B_m}]|\mathcal{F}_m] \\ &= \mathbb{E}[X_m \mathbb{1}_{B_m} \mathbb{E}[(X_n - X_m)|\mathcal{F}_m]]\end{aligned}$$

Since X_n is a submartingale, then $\mathbb{E}[(X_n - X_m)|\mathcal{F}_m]] \geq 0$. Since $X_m \mathbb{1}_{B_m}$ is either 0 or $c > 0$, $X_m \mathbb{1}_{B_m} \geq 0$ So,

$$\mathbb{E}[(X_n - X_m)X_m \mathbb{1}_{B_m}] \geq 0$$

Thus, we can get

$$\mathbb{E}[X^2] \geq \sum_{1 \leq m \leq n} c^2 P(B_m) = c^2 P(\cup_m B_m) = c^2 P(\max_{1 \leq m \leq n} X_m > c)$$

This completes the proof of Doob-Kolmogorov inequality.
Now we can use this result to prove the general version. Since absolute value functions and $f(x) = x^{\frac{p}{2}}$, for $p \geq 2$ are convex functions, applying Jesen's inequality (which can be found in section 2.4), $|X_n|^{\frac{p}{2}}$ is also a submartingale. Thus,

$$P(\max_{1 \leq m \leq n} |X_n| \geq \varepsilon) = P(\max_{1 \leq m \leq n} |X_n|^{\frac{p}{2}} \geq c^{\frac{p}{2}}) \leq \frac{\mathbb{E}[|X_n|^p]}{c^p}$$

□

The idea of this proof is similar to that of Doob's Upcrossing Inequality. The first and foremost step is to construct the events A and B, which work to describe the relationship between submartingale X_n and time m. As subsets of sample space, the expected values can be connected with probability through indicator functions. Jesen's inequality of martingales plays a crucial role in this proof as well. It is a useful tool in martingale proof and optimization problems since it simplifies the originally complicated process of proof, as illustrated when demonstrating the general case of Doob's Martingale Inequality.

3 Related Research: Robbins-Siegmund Theorem

In this section, I am going to introduce Stochastic Optimization briefly and discuss the application of Martingale Convergence Theorem in Robbins-Siegmund Theorem.

Stochastic Optimization aims to find an optimal solution to a problem by minimizing or maximizing an objective function with randomness. It plays a crucial role in Machine Learning. Martingales are usually applied to Stochastic Optimization algorithms due to their corresponding properties. Robbins-Siegmund Theorem is one of the most valuable results in Stochastic algorithms, because it can be applied to convergence of the stochastic gradient method. Martingale Convergence Theorem plays a significant role in the proof of Robbins-Siegmund Theorem [SGa18].

Theorem 6. *Robbins-Siegmund $(\mathcal{F}_n)_{n \geq 0}$ is a filtration. The four random variables (U_n), (V_n), (α_n) and (β_n) are $(\mathcal{F}_n)$-measurable, non-negative, integrable and satisfy the following properties:*

- *(U_n), (α_n) and (β_n) are $(\mathcal{F}_n)$-predictable*
- *$\sup_{\omega \in \Omega} \prod_{n \geq 1}(1 + \alpha_n(\omega)) < \infty$ and $\sum_{n \geq 0} \mathbb{E}[\beta_n] < \infty$*
- *$\forall n \in \mathbb{N}$, $\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leq V_n(1 + \alpha_{n+1}) + \beta_{n+1} - U_{n+1}$*

Then we can get the following two results:

1. *$V_n \to V_\infty \in L^1$ and $\sup_{n \geq 0} \mathbb{E}[V_n] < \infty$*

2. $\sum_{n\geq 0}\mathbb{E}[U_n]<\infty$ *and* $\sum_{n\geq 0}U_n<\infty$

Proof. Since U_n is predictable and the four variables are all non-neagtive,

$$\begin{aligned}\mathbb{E}\left(V_{n+1}+\sum_{k=1}^{n+1}U_k\mid\mathcal{F}_n\right)&\leq V_n(1+\alpha_{n+1})+\sum_{k=1}^{n}U_k+\beta_{n+1}\\&\leq\left(V_n+\sum_{k=1}^{n}U_k\right)(1+\alpha_{n+1})+\beta_{n+1}\end{aligned}$$

Divide $\prod_{k=1}^{n+1}(1+\alpha_k)$ to both sides:

$$\mathbb{E}\left(\frac{V_{n+1}+\sum_{k=1}^{n+1}U_k}{\prod_{k=1}^{n}(1+\alpha_k)}\mid\mathcal{F}_n\right)\leq\frac{(V_n+\sum_{k=1}^{n}U_k)\,(1+\alpha_{n+1})}{\prod_{k=1}^{n+1}(1+\alpha_k)}+\frac{\beta_{n+1}}{\prod_{k=1}^{n+1}(1+\alpha_k)}$$

We can obtain

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n]\leq S_n+\tilde{\beta}_{n+1}$$

where

$$S_n=\frac{V_n+\sum_{k=1}^{n}U_k}{\prod_{k=1}^{n}(1+\alpha_k)},\ \tilde{\beta}_n=\frac{\beta_n}{\prod_{k=1}^{n}(1+\alpha_k)}$$

Define $B_n=\sum_{k=1}^{n}\tilde{\beta}_k$. Since

$$\sup_{\omega\in.\Omega}\prod_{n\geq 1}(1+\alpha_n(\omega))<\infty,\ \beta_n\text{is integrable,}$$

we can obtain that

$$B_n\to B_\infty\text{ almost surely and }B_\infty\in L^1\implies\mathbb{E}B_\infty<\infty$$

In addition,

$$\sum_{n\geq 0}\mathbb{E}[\beta_n]<\infty\implies\sup_{n\geq 0}\mathbb{E}[S_n]<\infty$$

The next step is to **construct a super-martingale**. Define $\tilde{S}_n=S_n+\mathbb{E}[B_\infty|\mathcal{F}_n]-B_n$. We need to show$\tilde{S}_n$ is a supermartingale. Since $\mathbb{E}[S_{n+1}|\mathcal{F}_n]\leq S_n+\tilde{\beta}_{n+1}$ and $B_n\to B_\infty$, then

$$\begin{aligned}\mathbb{E}[\tilde{S}_n|\mathcal{F}_n]&\leq S_n+\tilde{\beta}_{n+1}+\mathbb{E}[B_\infty|\mathcal{F}_n]-\mathbb{E}[B_{n+1}|\mathcal{F}_n]\\&\leq S_n+\mathbb{E}[B_\infty|\mathcal{F}_n]-B_{n+1}+\tilde{\beta}_{n+1}\\&=\tilde{S}_n\end{aligned}$$

Check $\tilde{S}_n\in L^1$:

$$\mathbb{E}|\tilde{S}_n|\leq\mathbb{E}(|S_n|+|\mathbb{E}[B_\infty|\mathcal{F}_n]-B_n|)\leq\mathbb{E}S_n+\mathbb{E}B_\infty<\infty$$

Therefore, $(\tilde{S}_n)$ is a super-martingale. Apply **Martingale Convergence Theorem**,

$$\tilde{S}_n\to\tilde{S}_\infty\text{ and }\tilde{S}_\infty\in L^1$$

Since $\mathbb{E}[B_\infty|\mathcal{F}_n]-B_n\to 0$, in fact $S_\infty=\tilde{S}_\infty$ and $S_\infty\in L^1$. Also

$$S_n\leq\tilde{S}_n\implies\sup_{n\geq 1}\mathbb{E}[S_n]<\mathbb{E}[\tilde{S}_n]<\infty$$

Recall that

$$S_n=\frac{V_n+\sum_{k=1}^{n}U_k}{\prod_{k=1}^{n}(1+\alpha_k)}\implies\mathbb{E}[V_n]+\mathbb{E}\left(\sum_{k=1}^{n}U_k\right)=\mathbb{E}[\left(\prod_{k=1}^{n}(1+\alpha_K)\right)S_n]$$

Since

$$\mathbb{E}[\left(\prod_{k=1}^{n}(1+\alpha_K)\right)S_n] \leq \left\|\prod_{k=1}^{\infty}(1+\alpha_k)\right\|_{\infty}\mathbb{E}[S_n] < \infty$$

Therefore,

$$\sup_{n\geq 1}\mathbb{E}[V_n] < \infty$$

and

$$\mathbb{E}[\sum_{k\geq 1}U_k] < \infty \implies \sum_{k\geq 1}U_k < \infty \text{ almost surely}$$

Recall that $S_n \to S_\infty$ and $\prod_{k\geq 1}(1+\alpha_k) < \infty$,

$$V_n = S_n\prod_{k\geq 1}^{n}(1+\alpha_k) - \sum_{k\geq 1}U_k \to S_\infty\prod_{k\geq 1}(1+\alpha_k) - \sum_{k\geq 1}U_k = V_\infty \in L^1 \text{ almost surely}$$

This completes the proof. □

The most critical step in this proof of the Robbins-Siegmund Theorem is to construct super martingales with some given properties of random variables, then rearrange them and apply Martingale Convergence Theorem to $\tilde{S}_n, S_n$and B_n several times. This step brings significant convenience when proving the convergence of stochastic optimization algorithms. For instance, the key idea of the proof of De Finetti's Theorem[DGa13] (Given a sequence of exchangeable, tat is, for any $\pi_n \in S_n, (X_1, \cdots, X_n) \overset{\Delta}{=} (X_{\pi_n(1)}, \cdots, X_{\pi_n(n)})$, then conditional on ξ, $X_1, \cdots, X_n$ are i.i.d.) is to construct martingales and appy backward MG convergence theorem.

In [ABE19], Andreas, Krishnakumar and Murat (2019) provided non-asymptotic convergence rates of the Polyak-Ruppert averaged stochastic gradient descent (SGD) to a normal random vector with Martingale Central Limit Theorem, which is an extended version of Martingale Convergence Theorem. This paper [Zho+19]discssed the convergence of mirror descent in a class of stochastic optimization problems which are variational coherence by constructing submartingales combined with a series of submartingale convergence arguments. In general, the usage of Martingale Convergence Theorem brings insights to proving convergence in stochastic optimization problems. By employing this brilliant theory, we can tackle the problems of convergence rate with much ease.

4 Acknowledgement

I want to thank for Anqi Pu for helping me correct my grammatical mistakes and navigate through scientific writing. I also want to thank Yuchong Pan and David Xu who gave me a lot of valuable advice on my project.

References

[ABE19] A. Anastasiou, K. Balasubramanian, and M. A. Erdogdu. “Normal Approximation for Stochastic Gradient Descent via Non-Asymptotic Rates of Martingale CLT”. In: ed. by A. Beygelzimer and D. Hsu. Vol. 99. Proceedings of Machine Learning Research. Phoenix, USA: PMLR, 25–28 Jun 2019, pp. 115–137. URL: http://proceedings.mlr.press/v99/anastasiou19a.html.

[Blo19] B. Bloem-Reddy. *Lecture notes for STAT 547C: Topics in Probability*. Nov. 2019.

[Cin11] E. Cinlar. In: *Probability and Stochastics*. Springer International Publishing, 2011, pp. 193–196. URL: https://www.springer.com/gp/book/9780387878584.

[DGa13] D.Gamarnik. *Lecture notes for Advanced Stochastic Processes*. 2013.

[Dur19] R. Durrett. *Probability: Theory and Examples*. Springer–Verlag New York, 2019.

[SGa18] S.Gadat. *Lecture notes for S10 - Stochastic Optimization algorithms*. 2018.

[Zho+19] Z. Zhou et al. “On the Convergence of Mirror Descent beyond Stochastic Convex Programming”. In: *SIAM J. Optim.* 30.1 (2019), pp. 687–716. URL: https://doi.org/10.1137/17M1134925.

A Exercises

1. Let X_n be a $(\mathcal{F}_n)$-martingale with bounded increments. Note that Martingale with bounded increments mean that there exists $M > 0$ such that $\sup_n |X_{n+1} - X_n| \leq M$. Then with probability 1:

    - $\liminf_n X_n = -\infty$, $\limsup_n X_n = +\infty$
    - $\lim_n X_n$ exists and it is finite

    Proof. Without loss of general, pick $X_0 = 0$ by replacing X_n with $X_n - X_0$. Let $N_k = \inf\{n : X_n \geq k\}, k > 0$, then N_k is a stopping time by checking $\{N_k \leq n\} \in \mathcal{F}_n$. Therefore, by Corollary 1, $X_{n\wedge N_k}$ is a martingale, and

    $$0 \leq (X_{n\wedge N_k})^+ \leq k + M, \ \forall n \in \mathbb{N}$$

    due to the way we define N_k and the definition of bounded on increments. Therefore,

    $$\mathbb{E}[(X_{n\wedge N_k})^+] \leq k + M < \infty \implies \lim_{n\to\infty} X_{n\wedge N_k} \text{ exists and is finite}$$

    Then by Martingale Convergence Theorem,

    $$\lim_{n\to\infty} X_n = \lim_{n\to\infty} X_{n\wedge N_K} \text{ exists almost surely and is finite}$$

    In other words, $X_n \mathbb{1}_{\{N_k=\infty\}}$ converges almost surely. Then for

    $$\omega \in \bigcup_{k\in\mathbb{N}} \{N_k = +\infty\} = \{\lim \sup_{n\to\infty} X_n < \infty\} \text{ (real analysis)}$$

    Thus, $\lim_{n\to\infty} X_n$ exists and is finite. Replacing X with $-X$, then for almost every $\omega \in \{\liminf_{n\to\infty} X_n > -\infty\}$, $\lim_{n\to\infty} X_n$ exists and is finite. Let

    $$C = \{\lim_{n\to\infty} X_n \text{ exists and is finite}\}$$

    ,

    $$D = \{\{\lim \sup_{n\to\infty} X_n < \infty\} \cup \{\lim \inf_{n\to\infty} X_n > -\infty\}\}$$

    We have shown that $D \subset C$, then

    $$P(D \cap C^c) = 0 \implies P(D^c \cup C) = 1 \text{ almost surely}$$

    D^c is the first property and C is the second property, this completes the proof

    □

2. Proof the Second Borel-Cantelli Lemma

    Proof. Construct a martingae

    $$M_n = \sum_{k=1}^{n} (\mathbb{1}_{A_k} - P(A_k|\mathcal{F}_{k-1}))$$

    where $|M_n| \leq n, \forall n$ and $|M_{n+1} - M_n| \leq 1, \forall n$. Using the same notation as what we used in the proof of Exercise 1, On $C = \{\lim_{n\to\infty} X_n \text{ exists and is finite}\}$, we have

    $$\sum_{k=1}^{\infty} \mathbb{1}_{A_k} = +\infty \iff \sum_{k=1}^{\infty} P(A_k|\mathcal{F}k) = \infty$$

    This is because if either

    $$\sum_{k=1}^{\infty} \mathbb{1}_{A_k} < \infty \text{ or } \sum_{k=1}^{\infty} P(A_k|\mathcal{F}_{k-1}) < \infty,$$

then

$$M_n \to \left(\sum_{k=1}^{\infty} \mathbb{1}_{A_k}\right) - \left(\sum_{k=1}^{\infty} P(A_k|\mathcal{F}_{\|-\infty})\right) > \infty$$

On $D^c = \{\{\limsup_{n\to\infty} X_n = \infty\} \cup \{\liminf_{n\to\infty} X_n = -\infty\}\}$, we have

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

This completes the proof. □