Probability Lecture Notes

December 07

1 Introduction and The Main Goal

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables (R.V.) on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define the partial sum process:

$$S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i \qquad \forall n \geq 1 \tag{1}$$

1.1 The Objective

The main goal is to understand the behavior of the average $\frac{S_n}{n}$ as $n \to \infty$. This is studied under appropriate assumptions on the process, usually for large enough fixed n.

The most convenient assumption is that the variables (the process) are **Independent and Identically Distributed (i.i.d.)**.

- Recall: For a R.V. $X : \Omega \to \mathbb{R}$, its distribution μ_X is a probability measure on $\mathbb{R}$ defined by $\mu_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$.
- $X \overset{d}{=} Y \iff \mu_X = \mu_Y$. This determines characteristics like Mean $\mu = \mathbb{E}(X)$ and Variance.
- Independence implies: $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (variables are unrelated).

2 Laws of Large Numbers (LLN)

2.1 Weak Law of Large Numbers (WLLN)

If $\mathbb{E}(X^2) < \infty$, then $\frac{S_n}{n} \to \mathbb{E}(X_1)$ in probability. In fact, for any $\epsilon > 0$, we have the bound:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}X_1\right| > \epsilon\right) \leq \frac{1}{n}\frac{\sigma^2}{\epsilon^2} \tag{2}$$

Main Question 1. *Is $(\frac{1}{n})$ rate of convergance the best we can get?*

This question leads to the theory of **Large Deviations Estimates (LDE)**.

2.2 Strong Law of Large Numbers (SLLN)

If $\mathbb{E}|X_1| < \infty$, then $\frac{S_n}{n} \to \mathbb{E}X_1$ almost surely (a.s.). By subtracting the mean (let $\mathbb{E}X_1 = \mu, \mathbb{E}X_1' = 0$), we generally analyze the case where $\frac{S_n}{n} \to 0$.

2.3 Central Limit Theorem (CLT)

Assuming S_n is centered, S_n is usually much smaller than n ($o(n)$).

Main Question 2. *What is the "correct" size of S_n?*

It is roughly $\sqrt{n}$. In an appropriate sense, this describes the Central Limit Theorem.

3 Dynamical Systems: Dependence

Main Question 3. *Is independence really necessary to prove limit laws?*

Many processes are not independent.

Example 1 (Measure Preserving Dynamical Systems (MPDS))**.** *Let Ω be a compact metric space (e.g., Borel probability space). Let $f : \Omega \to \Omega$ be a continuous function. We define the iterations of x under f as $f^n(\omega) = f(f(\dots f(\omega)\dots))$.*

*Let μ be a probability measure on Ω that is **invariant** under f (i.e., $\forall E \subset \Omega, \mu(f^{-1}(E)) = \mu(E)$). This is a Measure Preserving Dynamical System (MPDS).*

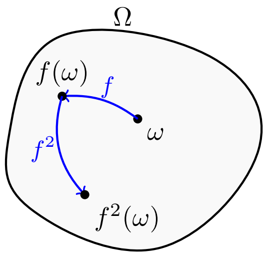


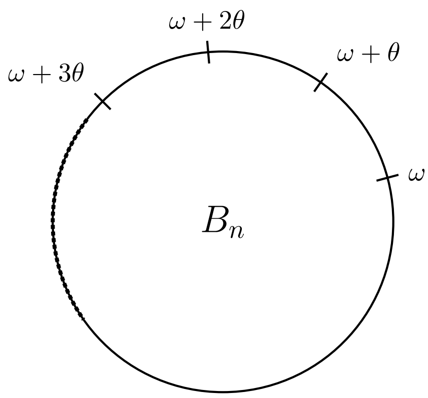
Figure 1: Visualization of Dynamical System Iterations on an Irregular Space Ω

Let $\varphi : \Omega \to \mathbb{R}$ be an $L^1(\mu)$ function (called an **observable**). Consider the random process:

$$X_0(\omega) = \varphi(\omega) \tag{3}$$
$$X_1(\omega) = \varphi(f(\omega)) \tag{4}$$
$$X_2(\omega) = \varphi(f \circ f(\omega)) \tag{5}$$
$$X_n(\omega) = \varphi(\underbrace{f \circ f \dots f}_{\text{n times}}(\omega)) = \varphi(f^n(\omega)) \tag{6}$$

This process is identically distributed but **not independent**.

Example 2. *Fixed rotations by θ. X_n is completely dependent on the previous point.*



The SLLN still holds for MPDS, but LDE and CLT require other hypotheses, specifically **Decay of Correlations** (Mixing).

$$|\mathbb{E}(X_n X_m) - \mathbb{E}(X_n)\mathbb{E}(X_m)| < Ce^{-\epsilon|n-m|} \tag{7}$$

This implies the system loses memory fast (e.g., Markov processes).

4 Conditional Expectation and Martingales

Martingales generalize sums of i.i.d. random variables.

4.1 Conditional Expectation

Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a sub-σ-algebra $\mathcal{F}_0 \subset \mathcal{F}$. For $X \in L^1$, the conditional expectation $\mathbb{E}(X|\mathcal{F}_0)$ is the unique $\mathcal{F}_0$-measurable random variable such that $\forall E \in \mathcal{F}_0$:

$$\int_E \mathbb{E}(X|\mathcal{F}_0) d\mathbb{P} = \int_E X d\mathbb{P} \tag{8}$$

Properties:

1. If X is $\mathcal{F}_0$-measurable, $\mathbb{E}(X|\mathcal{F}_0) = X$.
2. If X is independent of $\mathcal{F}_0$, $\mathbb{E}(X|\mathcal{F}_0) = \mathbb{E}(X)$.
3. **Tower Property:** If $\mathcal{F}_1 \subset \mathcal{F}_2$, $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(X|\mathcal{F}_1)$.
4. **Taking out what is known:** $\mathbb{E}(XY|\mathcal{F}_0) = X\mathbb{E}(Y|\mathcal{F}_0)$ if X is $\mathcal{F}_0$-measurable.
5. **Jensen's Inequality:** $\varphi(\mathbb{E}(X|\mathcal{F}_0)) \leq \mathbb{E}(\varphi(X)|\mathcal{F}_0)$ for convex φ.

4.2 Martingales

A **filtration** is an increasing sequence of sub-σ-algebras $\mathcal{F}_n$. A sequence (X_n) is **adapted** if X_n is $\mathcal{F}_n$-measurable.

Definition 1. *$X = (X_n)$ is a martingale with respect to $\mathcal{F}_n$ if:*

1. *X_n is adapted.*
2. *$X_n \in L^1$.*
3. *$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$.*

Example 3. *Sums of centered i.i.d. variables ($S_n = \sum X_i$ where $\mathbb{E}(X_i) = 0$) form a martingale w.r.t $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$.*

Example 4 (Doob's Martingale)**.** *Let $\big(\mathcal{F}_n\big)$ be a filteration. For $X \in L^1$, $X_n = \mathbb{E}(X|\mathcal{F}_n)$ is a martingale.*

5 Doob's Theorems and Decomposition

5.1 Doob's Convergence Theorem

Let $X = (X_n)$ be a (sub/super)martingale. Assume $\sup_n \mathbb{E}|X_n| < \infty$ (Uniform L^1 Boundedness). Then there exists $X_\infty \in L^1$ such that $X_n \to X_\infty$ almost surely.

5.2 Doob's Decomposition Theorem

Let $X = (X_n)$ be a random process in L^1 adapted to $\mathcal{F}_n$. There exists a unique decomposition:

$$X_n - X_0 = M_n + A_n \tag{9}$$

where:

- $M = (M_n)$ is a martingale (null at 0, $M_0 = 0$).
- $A = (A_n)$ is a **predictable** process (A_n is $\mathcal{F}_{n-1}$-measurable).

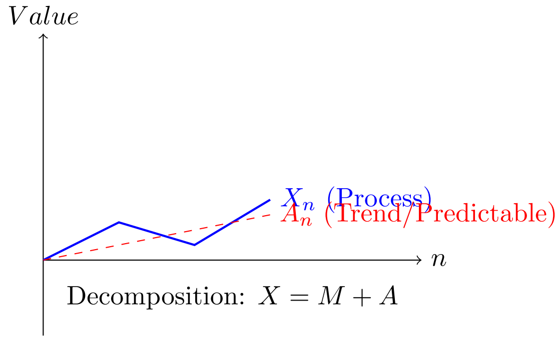
Derivation of the terms: To find A_n, we observe the increments. Since M is a martingale, $\mathbb{E}(M_k - M_{k-1}|\mathcal{F}_{k-1}) = 0$. Taking the conditional expectation of the increment $X_k - X_{k-1}$:

$$\begin{aligned} \mathbb{E}(X_k - X_{k-1}|\mathcal{F}_{k-1}) &= \mathbb{E}((M_k + A_k) - (M_{k-1} + A_{k-1})|\mathcal{F}_{k-1}) \\ &= \mathbb{E}(M_k - M_{k-1}|\mathcal{F}_{k-1}) + \mathbb{E}(A_k - A_{k-1}|\mathcal{F}_{k-1}) \\ &= 0 + (A_k - A_{k-1}) \quad \text{(since } A \text{ is predictable)} \end{aligned}$$

Thus, the increment of the predictable process is defined by the expected conditional drift of X. Summing these up gives the explicit formula:

$$A_n = \sum_{k=1}^{n} \mathbb{E}(X_k - X_{k-1}|\mathcal{F}_{k-1}) \tag{10}$$

X is a submartingale iff A is increasing.



6 Stopping Times and Transforms

6.1 Stopping Times

A random variable $T : \Omega \to \mathbb{Z}^+ \cup \{\infty\}$ is a **stopping time** if $\{T = n\} \in \mathcal{F}_n$ for all n.

Example 5. *Hitting time:* $T(\omega) = \inf\{n \geq 0 : X_n(\omega) \in B\}$ *for a Borel set* B.

6.2 Stopped Process

Given a stopping time T, the stopped process X^T is defined by $X_n^T(\omega) = X_{n\wedge T}(\omega)$, where $n \wedge T = \min(n, T)$.

Exercise 1. *Show that if* X *is a (sub/super) martingale or predictable, then* X^T *preserves this property.*

6.3 Martingale Transform (Discrete Stochastic Integral)

Let $C = (C_n)_{n\geq 1}$ be a predictable process and $X = (X_n)_{n\geq 0}$ be a martingale. The transform $C \cdot X$ is:

$$(C \cdot X)_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}) \tag{11}$$

Exercise 2 (Lemma)**.** *Prove that if* X_n *is a martingale, then* $(C \cdot X)_n$ *is a martingale (null at 0).*

7 Martingales in L^2 and Angle Brackets

7.1 Convergence in L^2

If $M = (M_n)$ is a martingale with $\sup_n \mathbb{E}(M_n^2) < \infty$, then $\lim_{n\to\infty} M_n$ exists almost surely and in L^2.

Orthogonality of Increments (Derivation): We wish to compute $\mathbb{E}(M_n^2)$. We can write $M_n = M_{n-1} + (M_n - M_{n-1})$. Squaring both sides:

$$M_n^2 = M_{n-1}^2 + 2M_{n-1}(M_n - M_{n-1}) + (M_n - M_{n-1})^2 \tag{12}$$

Taking expectations:

$$\mathbb{E}(M_n^2) = \mathbb{E}(M_{n-1}^2) + 2\mathbb{E}[M_{n-1}(M_n - M_{n-1})] + \mathbb{E}[(M_n - M_{n-1})^2]$$

Consider the middle term. Using the Tower Property $\mathbb{E}[\cdot] = \mathbb{E}[\mathbb{E}(\cdot|\mathcal{F}_{n-1})]$:

$$\begin{aligned}\mathbb{E}[M_{n-1}(M_n - M_{n-1})] &= \mathbb{E}\left[\mathbb{E}(M_{n-1}(M_n - M_{n-1})|\mathcal{F}_{n-1})\right] \\ &= \mathbb{E}\left[M_{n-1} \underbrace{\mathbb{E}((M_n - M_{n-1})|\mathcal{F}_{n-1})}_{=0 \text{ (Martingale prop)}}\right] = 0\end{aligned}$$

Thus, we arrive at the Pythagorean relation:

$$\mathbb{E}(M_n^2) = \mathbb{E}(M_0^2) + \sum_{k=1}^{n} \mathbb{E}((M_k - M_{k-1})^2) \tag{13}$$

7.2 Angle-Bracket Process $\langle M \rangle$

Let M be a martingale null at 0. M^2 is a submartingale. By Doob's Decomposition, $M^2 = N + A$, where A is predictable and increasing. We denote this A as $\langle M \rangle$.

$$\langle M \rangle_n = \sum_{k=1}^{n} \mathbb{E}((M_k - M_{k-1})^2|\mathcal{F}_{k-1}) \tag{14}$$

7.3 Theorem: Convergence on Finite Variation

Theorem 1. *Let M be a martingale null at 0. On the set $\{\langle M \rangle_\infty < \infty\}$, $\lim_{n\to\infty} M_n$ exists almost surely.*

Proof Sketch: We cannot apply Doob's Convergence directly because $\sup \mathbb{E}|M_n|$ might not be finite. We use a **stopped process**. Define stopping times $S_k = \inf\{n \geq 0 : \langle M \rangle_{n+1} > k\}$.

Exercise 3. *Verify that $(M^{S_k})^2 - \langle M \rangle^{S_k}$ is a martingale.*

Exercise 4. *Show that $\langle M^{S_k} \rangle = \langle M \rangle^{S_k}$.*

Since $\langle M \rangle_{S_k} \leq k$ (by definition of S_k), M^{S_k} is uniformly bounded in L^2. Therefore, $\lim M_n^{S_k}$ exists a.s. by Doob's L^2 consequence. Taking limits as $k \to \infty$, we recover convergence on the set where $\langle M \rangle_\infty < \infty$.