Martingales: History and a Limit Theorem

Jian Kang and Bhaumik Mehta

University of Washington Statistics Directed Reading Program

June 14, 2025

Contents

The First Limit Theorem (c. 1713)

- **Jacob Bernoulli (1713)** studied an **independent** urn model:

$$X_i = \begin{cases} 1 & \text{(white ball, probability } p\text{)}, \\ 0 & \text{(black ball, probability } 1-p\text{)}. \end{cases}$$

Jacob Bernoulli
(1654–1705)

The First Limit Theorem (c. 1713)

- **Jacob Bernoulli (1713)** studied an **independent** urn model:

$$X_i = \begin{cases} 1 & \text{(white ball, probability } p\text{)}, \\ 0 & \text{(black ball, probability } 1-p\text{)}. \end{cases}$$

- He proved:

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \varepsilon\right) \longrightarrow 0.$$

Jacob Bernoulli
(1654–1705)

The First Limit Theorem (c. 1713)

- **Jacob Bernoulli (1713)** studied an **independent** urn model:

$$X_i = \begin{cases} 1 & \text{(white ball, probability } p\text{)}, \\ 0 & \text{(black ball, probability } 1-p\text{)}. \end{cases}$$

- He proved:

$$\Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - p \right| > \varepsilon \right) \longrightarrow 0.$$

- Prototype for Weak Law of Large Numbers (WLLN)

Jacob Bernoulli (1654–1705)

The First Limit Theorem (c. 1713)

- **Jacob Bernoulli (1713)** studied an **independent** urn model:

$$X_i = \begin{cases} 1 & \text{(white ball, probability } p\text{)}, \\ 0 & \text{(black ball, probability } 1-p\text{)}. \end{cases}$$

- He proved:

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \varepsilon\right) \longrightarrow 0.$$

- Prototype for Weak Law of Large Numbers (WLLN)
- First ever limit theorem of probability theory.

Jacob Bernoulli (1654–1705)

De Moivre and Laplace: The Proto-CLT (c. 1738–1810)

- **Abraham De Moivre (c. 1738)** observed that for $X_i = \pm 1$ (fair coin), the distribution of $S_n = \sum_{i=1}^{n} X_i$ approximates a bell curve when n is large.

De Moivre and Laplace: The Proto-CLT (c. 1738–1810)

- **Abraham De Moivre (c. 1738)** observed that for $X_i = \pm 1$ (fair coin), the distribution of $S_n = \sum_{i=1}^{n} X_i$ approximates a bell curve when n is large.
- **Pierre-Simon Laplace (1810)** made this precise for **independent** Bernoulli(p) trials:

$$\Pr\Big(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\Big) \;\longrightarrow\; \Phi(x), \quad n \to \infty,$$

where Φ is the standard normal CDF.

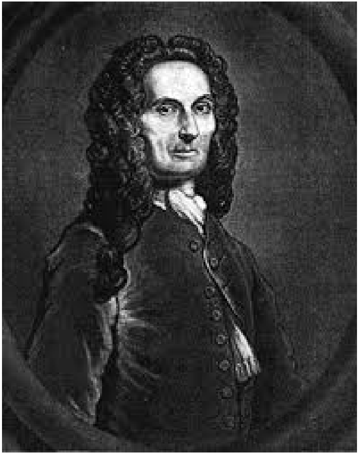
De Moivre and Laplace: The Proto-CLT (c. 1738–1810)

- **Abraham De Moivre (c. 1738)** observed that for $X_i = \pm 1$ (fair coin), the distribution of $S_n = \sum_{i=1}^{n} X_i$ approximates a bell curve when n is large.
- **Pierre-Simon Laplace (1810)** made this precise for **independent** Bernoulli(p) trials:

$$\Pr\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right) \longrightarrow \Phi(x), \quad n \to \infty,$$

  where Φ is the standard normal CDF.
- First glimpse of the Central Limit Theorem (CLT).

De Moivre and Laplace



*Abraham De Moivre
(1667–1754)

*Pierre-Simon Laplace
(1749–1827)

Classical Limit Theorems for Independent Sums

- **Law of Large Numbers (LLN).** If $\{X_i\}$ **i.i.d.** with $\mathbb{E}[|X_1|] < \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X_1].$$

Classical Limit Theorems for Independent Sums

- **Law of Large Numbers (LLN).** If $\{X_i\}$ **i.i.d.** with $\mathbb{E}[|X_1|] < \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X_1].$$

- **Central Limit Theorem (CLT).** If $\{X_i\}$ **i.i.d.** with $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 < \infty$,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

Classical Limit Theorems for Independent Sums

- **Law of Large Numbers (LLN).** If $\{X_i\}$ **i.i.d.** with $\mathbb{E}[|X_1|] < \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X_1].$$

- **Central Limit Theorem (CLT).** If $\{X_i\}$ **i.i.d.** with $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 < \infty$,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

- By the mid-19th century, foundational results like the LLN and CLT were only proven under i.i.d. assumptions.

The Centrality of Independence

- From Bernoulli to Laplace to Chebyshev, classical probability focused on sums of **independent** random variables and their asymptotic behavior.

The Centrality of Independence

- From Bernoulli to Laplace to Chebyshev, classical probability focused on sums of **independent** random variables and their asymptotic behavior.
- **Independence was the dominant assumption** in probability.

Pavel Nekrasov: Independence and Ideology

- **Pavel Nekrasov** (1853–1924):
  theologian turned probabilist, later Dean at Moscow University.

* Pavel A. Nekrasov
(1853–1924)

Pavel Nekrasov: Independence and Ideology

- **Pavel Nekrasov** (1853–1924): theologian turned probabilist, later Dean at Moscow University.
- Falsely claimed independence is a **necessary and sufficient** condition for LLN to hold.

* Pavel A. Nekrasov
(1853–1924)

Pavel Nekrasov: Independence and Ideology

- **Pavel Nekrasov** (1853–1924): theologian turned probabilist, later Dean at Moscow University.
- Falsely claimed independence is a **necessary and sufficient** condition for LLN to hold.
- Using this false claim, argued LLN is proof of human **free will**. (???)

* Pavel A. Nekrasov (1853–1924)

Andrey Markov: From Independence to Dependence

- **Andrey A. Markov** (1856–1922): Big-time probabilist. Student of Chebyshev, another big-time probabilist.

*Andrey A. Markov (1856–1922)

Andrey Markov: From Independence to Dependence

- **Andrey A. Markov** (1856–1922): Big-time probabilist. Student of Chebyshev, another big-time probabilist.
- Deeply opposed Nekrasov's theological framing of probability and his philosophical insistence on independence.

*Andrey A. Markov (1856–1922)

Andrey Markov: From Independence to Dependence

- **Andrey A. Markov** (1856–1922): Big-time probabilist. Student of Chebyshev, another big-time probabilist.
- Deeply opposed Nekrasov's theological framing of probability and his philosophical insistence on independence.
- In 1906, introduced what we now call **Markov chains**, showing LLN and CLT can hold under certain types of dependence.

*Andrey A. Markov
(1856–1922)

Markov's Refutation (1906)

- "The unique service of P. A. Nekrasov, in my opinion, is namely this: he brings out sharply his delusion, shared, I believe, by many, that independence is a necessary condition for the law of large numbers. **This prompted me to explain... that the [LLN] and [CLT] can apply also to dependent variables."**

Definition: Markov Chains (Markov, 1906)

Definition

A process $\{X_n\}_{n \geq 0}$ is a *Markov chain* if, for all n, i, j,

$$\Pr(X_{n+1} = j \mid X_n = i,\, X_{n-1}, \ldots, X_0) \;=\; \Pr(X_{n+1} = j \mid X_n = i),$$

Definition: Markov Chains (Markov, 1906)

Definition

A process $\{X_n\}_{n\geq 0}$ is a *Markov chain* if, for all n, i, j,

$$\Pr(X_{n+1} = j \mid X_n = i,\, X_{n-1}, \ldots, X_0) \;=\; \Pr(X_{n+1} = j \mid X_n = i),$$

"In this way a construction of a highly general character was actually arrived at, which P. A. Nekrasov can not even dream about."
–Markov

Markov's Law of Large Numbers (1906)

Markov Chain LLN (Markov, 1906): Suppose $\{X_n\}$ irreducible and aperiodic, with stationary distribution π. Then, for bounded f,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{\mathrm{P}} \mathbb{E}_\pi[f].$$

Markov's Law of Large Numbers (1906)

Markov Chain LLN (Markov, 1906): Suppose $\{X_n\}$ irreducible and aperiodic, with stationary distribution π. Then, for bounded f,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{\mathrm{P}} \mathbb{E}_\pi[f].$$

First limit theorem for a sequence of **dependent** random variables.

Markov Chains: A Simple Dependent Example

- **Markov Chain LLN (Markov, 1906):** Suppose $\{X_n\}$ irreducible and aperiodic, with stationary distribution π. Then, for bounded f,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{\mathrm{P}} \mathbb{E}_\pi[f].$$

Markov Chains: A Simple Dependent Example

- **Markov Chain LLN (Markov, 1906):** Suppose $\{X_n\}$ irreducible and aperiodic, with stationary distribution π. Then, for bounded f,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{\mathrm{P}} \mathbb{E}_\pi[f].$$

- In that proof, Markov effectively wrote

$$\mathbb{E}[\,f(X_k) \mid X_{k-1}] = \text{“predictable part” at step } k,$$

and then controlled the residuals.

Markov Chains: A Simple Dependent Example

- **Markov Chain LLN (Markov, 1906):** Suppose $\{X_n\}$ irreducible and aperiodic, with stationary distribution π. Then, for bounded f,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{\text{P}} \mathbb{E}_\pi[f].$$

- In that proof, Markov effectively wrote

$$\mathbb{E}[\, f(X_k) \mid X_{k-1}] \;=\; \text{“predictable part” at step } k,$$

  and then controlled the residuals.
- Thus, rather than summing i.i.d. terms, he subtracted a one-step conditional mean and showed the remainder converged.

Markov Chains: A Simple Dependent Example

- **Markov Chain LLN (Markov, 1906):** Suppose $\{X_n\}$ irreducible and aperiodic, with stationary distribution π. Then, for bounded f,

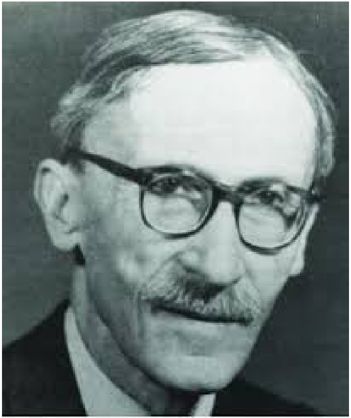
$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{\mathrm{P}} \mathbb{E}_\pi[f].$$

- In that proof, Markov effectively wrote

$$\mathbb{E}[\, f(X_k) \mid X_{k-1}] = \text{"predictable part" at step } k,$$

  and then controlled the residuals.
- Thus, rather than summing i.i.d. terms, he subtracted a one-step conditional mean and showed the remainder converged.
- Lévy recognized this as a template for handling *any* sequence where a predictable conditional mean is known.

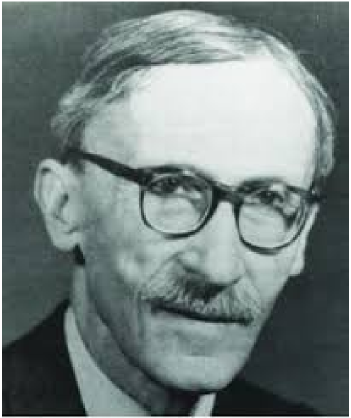
Paul Lévy



*Paul Lévy (1886–1971)

- **Paul Lévy** (1886–1971): Big-time probabilist.

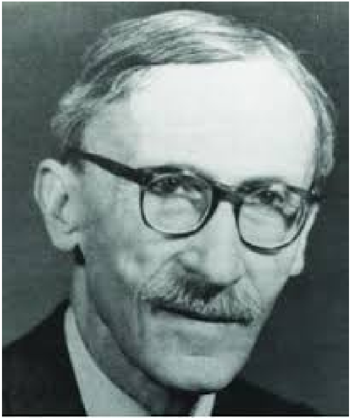
Paul Lévy



*Paul Lévy (1886–1971)

- **Paul Lévy** (1886–1971): Big-time probabilist.
- Contributions: Early martingales, characteristic functions, stable laws, early stochastic processes, etc.

Paul Lévy



*Paul Lévy (1886–1971)

- **Paul Lévy** (1886–1971): Big-time probabilist.
- Contributions: Early martingales, characteristic functions, stable laws, early stochastic processes, etc.
- Wanted a general approach to extending limit theorems to dependent sequences.

Lévy's Approach

- Suppose you have a sequence of random variables $X_1, X_2, \ldots$, not necessarily identical.

Lévy's Approach

- Suppose you have a sequence of random variables $X_1, X_2, \ldots$, not necessarily identical.
- At step k, you know everything that happened in the past: $X_1, \ldots, X_{k-1}$.

Lévy's Approach

- Suppose you have a sequence of random variables $X_1, X_2, \ldots$, not necessarily identical.
- At step k, you know everything that happened in the past: $X_1, \ldots, X_{k-1}$.
- You can form a *prediction* $m_k = \mathbb{E}[X_k \mid X_1, \ldots, X_{k-1}]$.

Lévy's Approach

- Suppose you have a sequence of random variables $X_1, X_2, \ldots$, not necessarily identical.
- At step k, you know everything that happened in the past: $X_1, \ldots, X_{k-1}$.
- You can form a *prediction* $m_k = \mathbb{E}[X_k \mid X_1, \ldots, X_{k-1}]$.
- Lévy subtracted this prediction from X_k, then considered the leftover:

$$X_k - m_k.$$

Lévy's Approach

- Lévy then set up the "compensated sum" by summing up the "leftovers",

$$M_n = \sum_{k=1}^{n} \bigl(X_k - m_k\bigr).$$

so each increment is

$$M_k - M_{k-1} \;=\; X_k - m_k.$$

Lévy's Approach

- **Theorem (Lévy, 1934).** If

$$\sup_n \mathbb{E}\big[\,|M_n|\,\big] \;<\; \infty,$$

then M_n converges almost surely to a finite limit M_∞.

- **Theorem (Lévy, 1934).** If

$$\sup_n \mathbb{E}\big[\,|M_n|\,\big] \;<\; \infty,$$

then M_n converges almost surely to a finite limit M_∞.

Very roughly,

  - When $m_k = \mathbb{E}[X_k]$, this recovers the classical LLN.

Lévy's Approach

- **Theorem (Lévy, 1934).** If

$$\sup_n \mathbb{E}\big[\,|M_n|\,\big] \;<\; \infty,$$

then M_n converges almost surely to a finite limit M_∞.

Very roughly,

  - When $m_k = \mathbb{E}[X_k]$, this recovers the classical LLN.
  - When $m_k = \mathbb{E}[X_k \mid X_{k-1}]$, this recovers Markov's LLN.

Lévy's Approach

- **Theorem (Lévy, 1934).** If

$$\sup_n \mathbb{E}\big[\,|M_n|\,\big] \;<\; \infty,$$

  then M_n converges almost surely to a finite limit M_∞.

  Very roughly,
  - When $m_k = \mathbb{E}[X_k]$, this recovers the classical LLN.
  - When $m_k = \mathbb{E}[X_k \mid X_{k-1}]$, this recovers Markov's LLN.
- Lévy's compensated-sum approach unified independent and Markov limit theorems for sums.

First Glimpse at Martingales

Even though Lévy did not make the explicit connection, note that by construction,

$$\mathbb{E}\big[M_k - M_{k-1} \mid \mathcal{F}_{k-1}\big] \;=\; \mathbb{E}\big[X_k - m_k \mid \mathcal{F}_{k-1}\big] \;=\; 0.$$

First Glimpse at Martingales

Even though Lévy did not make the explicit connection, note that by construction,

$$\mathbb{E}\big[M_k - M_{k-1} \mid \mathcal{F}_{k-1}\big] \;=\; \mathbb{E}\big[X_k - m_k \mid \mathcal{F}_{k-1}\big] \;=\; 0.$$

So $\{M_n\}$ satisfies

$$\mathbb{E}[\, M_{n+1} \mid \mathcal{F}_n \,] \;=\; M_n,$$

First Glimpse at Martingales

Even though Lévy did not make the explicit connection, note that by construction,

$$\mathbb{E}\big[M_k - M_{k-1} \mid \mathcal{F}_{k-1}\big] \;=\; \mathbb{E}\big[X_k - m_k \mid \mathcal{F}_{k-1}\big] \;=\; 0.$$

So $\{M_n\}$ satisfies

$$\mathbb{E}[\, M_{n+1} \mid \mathcal{F}_n \,] \;=\; M_n,$$

which is exactly the *martingale* property.

Jean Ville: Defining Martingales in Games of Chance (1939)

- **Jean Ville** (1910–1988) coined the term *martingale* in his 1939 thesis, inspired by Lévy's approach to getting limit theorems for dependent sums.

Jean Ville: Defining Martingales in Games of Chance (1939)

- **Jean Ville** (1910–1988) coined the term *martingale* in his 1939 thesis, inspired by Lévy's approach to getting limit theorems for dependent sums.
- **Martingale definition (Ville):** A sequence $\{M_n\}$ is a martingale if

$$\mathbb{E}\big[M_{n+1} \mid \mathcal{F}_n\big] = M_n,$$

regardless of whether M_n arises from sums, products, or more general operations.

Jean Ville: Defining Martingales in Games of Chance (1939)

- **Jean Ville** (1910–1988) coined the term *martingale* in his 1939 thesis, inspired by Lévy's approach to getting limit theorems for dependent sums.
- **Martingale definition (Ville):** A sequence $\{M_n\}$ is a martingale if

$$\mathbb{E}\big[M_{n+1} \mid \mathcal{F}_n\big] = M_n,$$

regardless of whether M_n arises from sums, products, or more general operations.
- Ville emphasized that **no special form (sum/product) is needed**: any process satisfying the conditional-expectation property qualifies.

Joseph L. Doob: Martingales in a Measure-Theoretic Framework (1940s)

- **Joseph L. Doob** (1910–2004) systematically developed martingale theory using measure theory and filtrations $(\mathcal{F}_n)$.

Joseph L. Doob: Martingales in a Measure-Theoretic Framework (1940s)

- **Joseph L. Doob** (1910–2004) systematically developed martingale theory using measure theory and filtrations $(\mathcal{F}_n)$.

General definition

A sequence $\{M_n\}$ of integrable random variables is a martingale if

$$M_n \text{ is } \mathcal{F}_n\text{-measurable and } \mathbb{E}\left[M_{n+1} \mid \mathcal{F}_n\right] = M_n \quad \text{a.s.}$$

Joseph L. Doob: Martingales in a Measure-Theoretic Framework (1940s)

- **Joseph L. Doob** (1910–2004) systematically developed martingale theory using measure theory and filtrations $(\mathcal{F}_n)$.

General definition

A sequence $\{M_n\}$ of integrable random variables is a martingale if

$$M_n \text{ is } \mathcal{F}_n\text{-measurable and } \mathbb{E}\big[M_{n+1} \mid \mathcal{F}_n\big] = M_n \quad \text{a.s.}$$

- **Doob's Martingale Convergence Theorem (1940):** If $\{M_n\}$ is a martingale with $\sup_n \mathbb{E}[\,|M_n|\,] < \infty$, then M_n converges almost surely.

Ville and Doob

*Jean Ville (1910–1989)

*Joseph L. Doob (1910–2004)

Doob's Convergence Theorem

Theorem

Let M_n be a martingale with

$$\sup_{n \geq 0} \mathbb{E}[|M_n|] < \infty.$$

Then there exists M_∞ such that

$$M_n \xrightarrow{\text{a.s.}} M_\infty.$$

Doob's Convergence Theorem

Theorem

Let M_n be a martingale with

$$\sup_{n \geq 0} \mathbb{E}[|M_n|] < \infty.$$

Then there exists M_∞ such that

$$M_n \xrightarrow{\text{a.s.}} M_\infty.$$

- Proof strategy: bound the number of significant oscillations (upcrossings) to check convergence.

Doob's Convergence Theorem

Theorem

Let M_n be a martingale with

$$\sup_{n\geq 0} \mathbb{E}[|M_n|] < \infty.$$

Then there exists M_∞ such that

$$M_n \xrightarrow{\text{a.s.}} M_\infty.$$

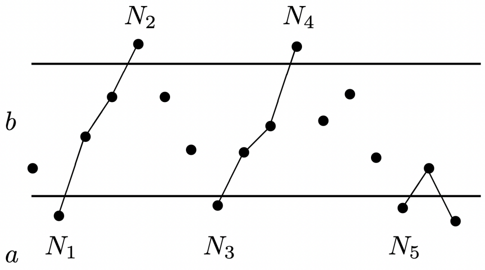
- Proof strategy: bound the number of significant oscillations (upcrossings) to check convergence.
- **Key idea:** Martingales can "buy low, sell high" only finitely many times.

Upcrossings: Formalizing Swings

- Fix two levels $a < b$. An *upcrossing* is one complete swing from at or below a up to at or above b.

Upcrossings: Formalizing Swings

- Fix two levels $a < b$. An *upcrossing* is one complete swing from at or below a up to at or above b.
- Let $U_n(a, b)$ denote the number of upcrossings by time n.

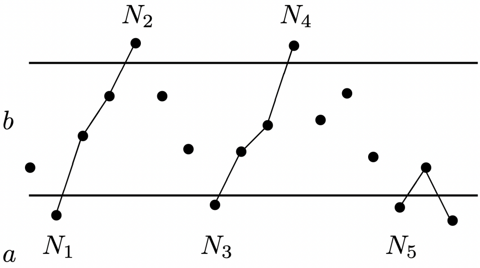


Upcrossings: Formalizing Swings

- Intuitively, we "buy" the stock if the price falls below a, then "sell" once the price reaches above b.

Upcrossings: Formalizing Swings

- Intuitively, we "buy" the stock if the price falls below a, then "sell" once the price reaches above b.
- Number of upcrossings is how many times we "buy low, sell high".



Doob's Upcrossing Lemma

Lemma

Let (M_n) be a martingale. For any $a < b$,

$$(b - a)\,\mathbb{E}[U_n(a,b)] \leq \mathbb{E}[(M_n - a)^-] - \mathbb{E}[(M_0 - a)^-].$$

Doob's Upcrossing Lemma

Lemma

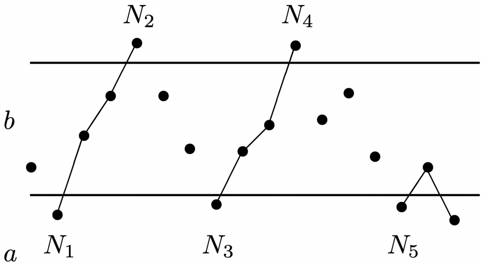
Let (M_n) be a martingale. For any $a < b$,

$$(b-a)\,\mathbb{E}[U_n(a,b)] \leq \mathbb{E}[(M_n - a)^-] - \mathbb{E}[(M_0 - a)^-].$$

- **Key takeaway:** Since $\sup_{n\geq 0} \mathbb{E}[|M_n|] < \infty$ by assumption, RHS is finite, so the expected number of upcrossings is finite.

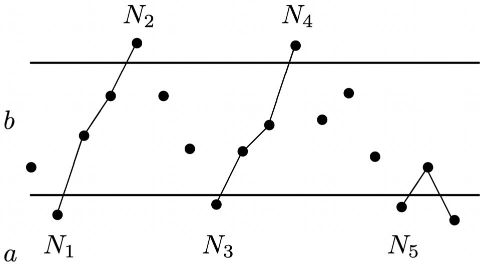
Why Finite Upcrossings Imply Convergence

- For any $a < b$, we showed $\mathbb{E}[U_\infty(a,b)] < \infty$.



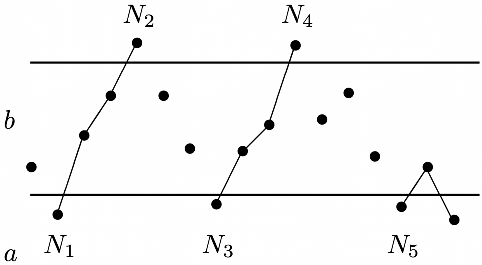
Why Finite Upcrossings Imply Convergence

- For any $a < b$, we showed $\mathbb{E}[U_\infty(a, b)] < \infty$.
- This implies $U_\infty(a, b) < \infty$ a.s.



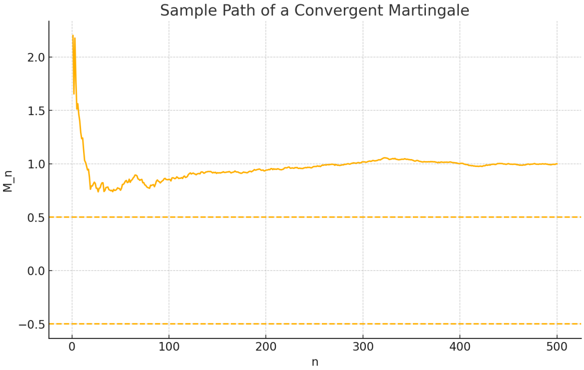
Why Finite Upcrossings Imply Convergence

- For any $a < b$, we showed $\mathbb{E}[U_\infty(a,b)] < \infty$.
- This implies $U_\infty(a,b) < \infty$ a.s.
- So M_n can only cross between **any** $a < b$ finitely often a.s.



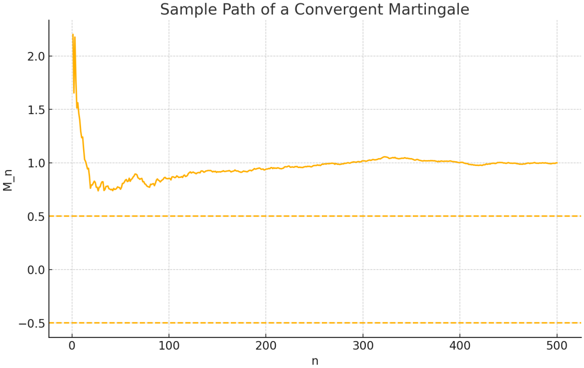
Why Finite Upcrossings Imply Convergence

- That is, eventually M_n stays within some interval.

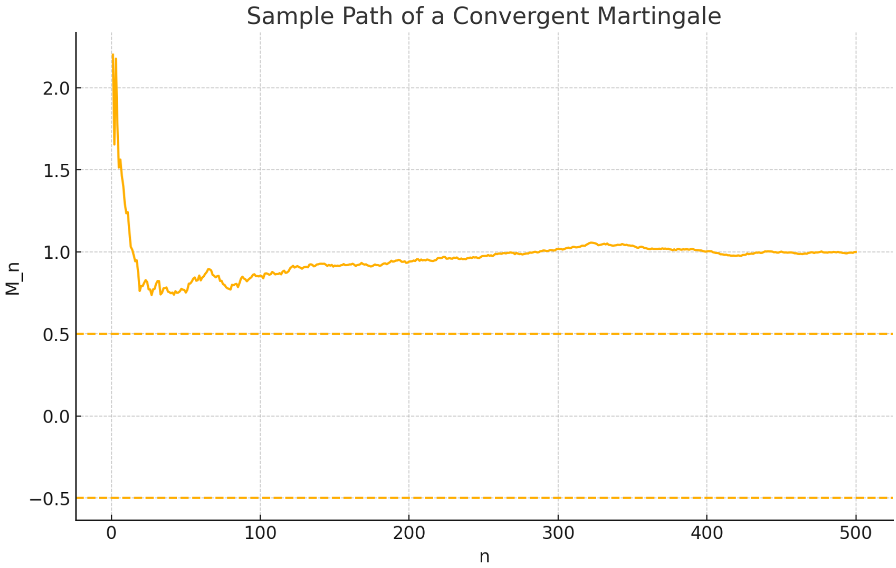


Why Finite Upcrossings Imply Convergence

- That is, eventually M_n stays within some interval.
- Since M_n stays within an arbitrarily small interval after some time, **every** sample path converges, i.e., M_n converges a.s. to some M_∞.



Visual of Martingale a.s. Convergence



Key Ideas Summarized

- **Upcrossings** count significant oscillations.

Key Ideas Summarized

- **Upcrossings** count significant oscillations.
- **Upcrossing lemma** bounds the expected count of these oscillations using boundedness of the martingale.

Key Ideas Summarized

- **Upcrossings** count significant oscillations.
- **Upcrossing lemma** bounds the expected count of these oscillations using boundedness of the martingale.
- **Almost-sure convergence** follows by ruling out infinite oscillations between any two levels.

Martingale Convergence Theorem Revisited

Theorem

Let M_n be a martingale with

$$\sup_{n\geq 0} \mathbb{E}[|M_n|] < \infty.$$

Then there exists M_∞ such that

$$M_n \xrightarrow{\text{a.s.}} M_\infty.$$

Acknowledgements

First and foremost, we want to thank Nila and Leon for being amazing mentors. We learned so much from you guys the past two quarters and had a lot of fun along the way. We also want to thank Ethan for organizing the DRP and making this possible.

References

R. Durrett. *Probability: Theory and Examples*, 4th ed., Cambridge University Press, Cambridge, 2010.

L. Mazliak and G. Shafer, eds. *The Splendors and Miseries of Martingales: Their History from the Casino to Mathematics.* Trends in the History of Science series, Springer, 2022.

E. Seneta. "Statistical Regularity and Free Will: L.A.J. Quetelet and P.A. Nekrasov." *Internat. Statist. Rev.* 71(2):319–334, August 2003.

E. Seneta. "Markov and the Birth of Chain Dependence Theory." *Internat. Statist. Rev.* 64(3):255–263, December 1996.

D. Williams. *Probability with Martingales*, Cambridge University Press, Cambridge, 1991.

Thank you!