ON THE LAW OF THE ITERATED LOGARITHM

RICHARD MAI

1. Abstract

In real-life, randomness is everywhere. However, some arbitrary things turn out to follow rules nicely. The Law of the iterated logarithm bounds the partial sum of the random variables when n goes to infinity. In this paper, our primary objective is to illustrate a formal proof of the law of the iterated logarithm. To achieve this, we'll first walk through the basic definitions in probability theory. We'll also provide famous and interesting limit theorems, which lays the foundation of the proof for the LIL. Lastly, we'll discuss the two lemmas that are turn out to be the last building blocks to the proof, and we'll formally give an intuitive and less-obvious proof of the law of the iterated logarithm. Notably, compared to lots of proofs out there, our proof doesn't involve heavy measure theory.

This is both an interesting and important theorem to prove because it is a major result in probability theory and allows for applications in finance, machine learning, and the study of statistics. After reading the paper, the reader will have a solid grasp of the fundamental limit theorems and be able to understand the intuitive steps to prove the law of the iterated logarithm.

2. Introduction

If you're not an NPC, you've probably had moments when life feels completely random, like getting soaked right after forgetting your umbrella, watching the stock market bounce for no reason, or seeing your favorite NBA team lose to one that's clearly tanking for Cooper Flagg. At first glance, a random walk seems to follow that same chaotic logic. But what's truly fascinating is that even this kind of randomness has its limits.

The law of the iterated logarithm lies upon probability theory, along with some statistics. Probability theory is the study that focuses on quantifying the uncertainty and analyzing the random events. Unlike the "naive" probability that we learn in high school, as Blitzstein referred to, the study of probability aims to analyze much more complex and uneven probability, which are sometimes very counterintuitive.

Here we'll illustrate two interestingly counterintuitive examples that can be extrapolated by probability: the Birthday Paradox and the Monty Hall problem.

We have 365 days in a year, therefore it seems unlikely that in a small group of people, two would share the same birthday. However, according to the Birthday Paradox, it turns out that with just 23 people, the probability that at least two share the same birthday is greater than 50%. Another result from the Monty Hall problem states as follows: on the American show "Let's Make a Deal", there are three doors, one with a car behind and two with a goat behind. After the contestant selects a door, the host, Monty Hall then removes one of the other two door, guaranteed that a goat is behind it. The contestant then chooses

Date: July 14, 2025.

whether to switch doors, in order to maximize his chance of winning the car. It turns out that switching doors doubles the chance of winning the car, which is counter-intuitive. While people may think that either switching or non-switching has a $\frac{1}{2}$, the probability after switching is actually $\frac{2}{3}$.

These two illustration provides a brief background and understanding of why we care about probability theory, and in fact, it has far more interesting applications in our real-life.

In real life, things don't just average out cleanly. Sometimes they drift far from what we expect. The Law of the Iterated Logarithm (LIL) helps us understand just how far things can go, and how often. Whether it's modeling financial markets, training AI, or even estimating how long you'll be waiting in a Starbucks line, understanding the limits of fluctuation matters.

3. Background

Definition 3.1 (Random Variables)**.** A random variable is a function that maps random events to a number; it is a way of quantifying events to study them.

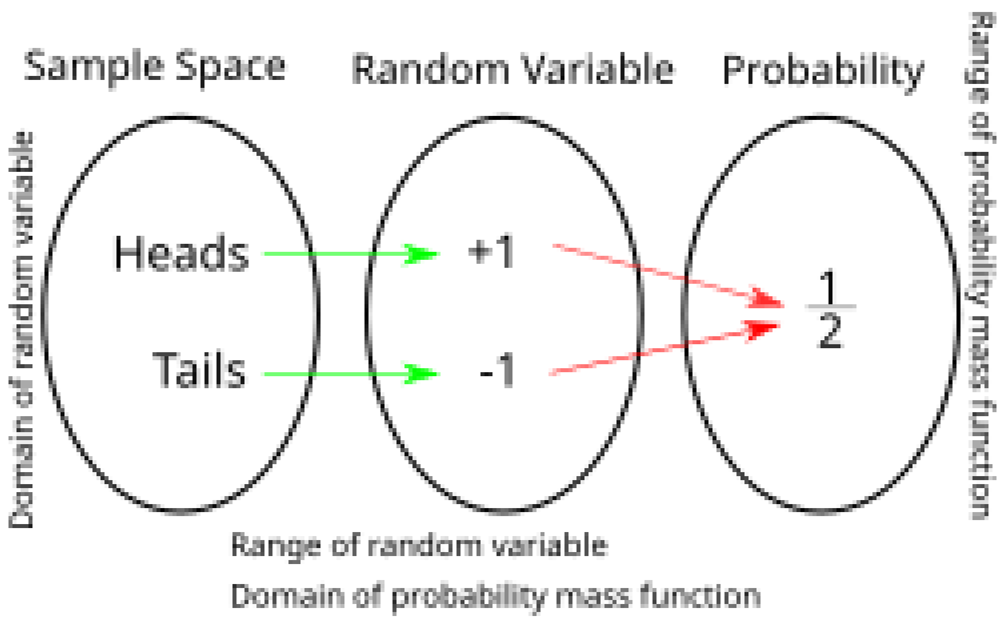

Figure 1. A graph of Random Variables [1]

For example, from tossing the coin, there are two possible outcomes: Heads(H) and Tails(T)(1). This is the domain of the function. According to the function, we can map Heads to the random variable +1 and Tails to the random variable −1. This is the range of the function. Keep in mind, that the randomness occurs when choosing the possible outcomes from the sample space, which in this case are heads and tails, while the process of mapping the possible outcomes to a specific number is fixed.

Specifically, an **indicator random variable** (I_A or $\mathbf{1}_A$) is a simple but powerful concept: it takes the value 1 if an event A occurs, and 0 otherwise.

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

An interesting property is that its **expected value is precisely the probability of the event**: $E[I_A] = P(A)$. This allows us to convert statements about event probabilities into numerical expectations, often simplifying calculations. In advanced probability, indicator variables are crucial for transforming abstract event-based arguments into tangible numerical forms.

Back to the coin toss as illustrated in 1, let A be the even that a fair coin lands on heads. Simply define the indicator random variable

$$I_A = \begin{cases} 1 & \text{if the coin lands on heads} \\ 0 & \text{otherwise} \end{cases}$$

Then $\mathbb{E}[I_A] = \mathbb{P}[A] = \frac{1}{2}$.

Definition 3.2 (o-notation)**.** Suppose that $g(x) \neq 0$ for all $x \neq a$ in some open interval containing a. Then $f(x)$ is little-oh of $g(x)$ or $f(x)$ is of smaller order than $g(x)$, denoted by

$$f(x) = o(g(x)) \text{ as } x \to a,$$

if

$$lim_{x \to a} \frac{f(x)}{g(x)} = 0$$

Definition 3.3 (O-notation)**.** Suppose that $g(x) \neq 0$ for all $x \neq a$ in some open interval containing a. $f(x)$ is big-oh of $g(x)$ or $f(x)$ is the same order as $g(x)$, denoted by

$$f(x) = O(g(x)) \text{ as } x \to a,$$

if

$$\lim_{\epsilon \to 0} \sup_{\{x \in \mathbb{R} : |x-a| < \epsilon\}} \left| \frac{f(x)}{g(x)} \right| < \infty.$$

Definition 3.4 (Kolmogorov's Axioms)**.** Kolmogorov's Axioms are the basic rules in Probability Theory. It's like in Euclidean space, it is agreed upon that two parallel lines don't concur. On the other hand, projective geometry sSignificanceee two parallel lines to concur at the infinity point.

Let Ω be the sample space (set of all possible outcomes), and let $\mathcal{F}$ be a σ-algebra of subsets of Ω, representing the set of all events. A probability measure is a function $P : \mathcal{F} \to [0, 1]$. The three Kolmogorov axioms are as follows:

(1) **Non-negativity:** For any event $A \in \mathcal{F}$, its probability is non-negative:

$$P(A) \geq 0.$$

(2) **Normalization:** The probability of the entire sample space (the certain event) is 1:

$$P(\Omega) = 1.$$

(3) **Countable Additivity:** For any sequence of pairwise disjoint events $A_1, A_2, A_3, \cdots \in \mathcal{F}$ (i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$), the probability of their union is the sum of their individual probabilities:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

[7]

Definition 3.5 (Moment Generating Function (MGF)). Let X be a real-valued random variable. The *moment generating function* of X, denoted $M_X(t)$, is defined as

$$M_X(t) = \mathbb{E}[e^{tX}]$$

for all values of $t \in \mathbb{R}$ such that the expectation exists and is finite. The MGF, when it exists in an open interval around 0, uniquely determines the distribution of X, and its n-th derivative at $t = 0$ yields the n-th moment of X, i.e.,

$$M_X^{(n)}(0) = \mathbb{E}[X^n].$$

Definition 3.6 (Cumulant Generating Function (CGF)). The *cumulant generating function* of a random variable X is defined as the natural logarithm of its moment generating function:

$$K_X(t) = \log M_X(t) = \log \mathbb{E}[e^{tX}]$$

for all $t \in \mathbb{R}$ where the MGF exists and is finite. The n-th derivative of the CGF at $t = 0$ yields the n-th cumulant of X, denoted κ_n:

$$\kappa_n = K_X^{(n)}(0).$$

Cumulants provide alternative characterizations of a distribution, with the first cumulant representing the mean, the second the variance, and higher-order cumulants capturing aspects such as skewness and kurtosis.

Theorem 3.7 (Markov's Inequality). *For a nonnegative random variable X and $a > 0$,*

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

The proof for this is surprisingly simple. Let $Y = X/a$, therefore we need to show that $P(Y \geq 1) \leq E(Y)$. Here is where the magic comes in. Note that

$$I(Y \geq 1) \leq E(Y).$$

If $I(Y \geq 1) = 0$ then the inequality reduces to $Y \geq 0$, which is true based on the assumption; and if $I(Y \geq 1) = 1$ then $Y \geq 1$, because the indicator says so. Taking the expectation of both sides, we have Markov's inequality.

Markov's inequality is an interesting application of indicator random variables and makes things easier afterwards.

Theorem 3.8 (Chebyshev's inequality). *Let μ and σ^2 be the mean and variance of X, respectively. Then for any nonnegative a,*

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

By Markov's inequality,

$$P(|X-\mu| \geq a) = P((X-\mu)^2 \geq a^2) \leq \frac{E(X-\mu)^2}{a^2} = \frac{\sigma^2}{a^2}.$$

Substituting $c\sigma$ for a, we have the equivalent form described as:

$$P(|X-\mu| \geq c\sigma) \leq \frac{1}{c^2}.$$

Interpreting this inequality by the common use of language, this gives an upper bound on the probability of an random variable being more than c standard deviations away from its mean. For example, there can't be more than 25% chance of being 2 or more standard deviations from the mean.

Theorem 3.9 (Chernoff's inequality)**.** *For any random variable X and constants $a > 0$ and $t > 0$,*

$$P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}.$$

By Markov's inequality again,

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX}}{e^{ta}}.$$

Chernoff's bound offers more than Markov's because the right-hand side can be optimized over t to give the tightest upper bound by taking the derivative.

Remarkably, without any additional operations, just simply by derivation from Markov's inequality, both Chebyshev and Chernoff's inequalities give a more precise bound than Markov's. [6]

Definition 3.10. Let $(E_n)_{n=1}^{\infty}$ be a sequence of events. The **limit supremum** (or **limit superior**) of the sequence of events, denoted by $\limsup_{n\to\infty} E_n$, is the event that infinitely many of the events (E_n) actually occur.

Explicitly, it is defined as:

$$\limsup_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k$$

Lemma 3.11 (The First Borel-Cantelli Lemma)**.** *The Lemma states that if the sum of the probabilities of the events E_n is finite:*

$$\sum_{n=1}^{\infty} Pr(E_n) < \infty,$$

then the probability that infinitely many of them occur is 0, that is,

$$Pr(\limsup_{n\to\infty} E_n) = 0.$$

Lemma 3.12 (The Second Borel-Cantelli Lemma)**.** *Conversely, if the sum of the probabilities of the events E_n diverges,*

$$\sum_{n=1}^{\infty} Pr(E_n) = \infty,$$

then the probability that infinitely many of them occur is 1,

$$Pr(\limsup_{n\to\infty} E_n) = 1.$$

[5]

Theorem 3.13 (Central limit theorem)**.** *some sequence of random variables with the same distribution* $X_1, X_2, \ldots$ *with mean* μ *and variance* σ^2*.* $\overline{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$*.*

The Central limit theorem, or the CLT, states that for large n, the distribution of the random variable $\bar{X}_n$ *after standardization, meaning centered at zero, and divide by the standard deviation of* $\bar{X}_n$*, converges to the normal distribution* $\mathcal{N}(0,1)$*.*

In other words, as $n \to \infty$*,*

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \to \mathcal{N}(0,1).$$

Proof. Let

$$Y_i = \frac{X_i - \mu}{\sigma} \text{so that } \mathbb{E}[Y_i] = 0,\ Var(Y_i) = 1.$$

Then $Z_n = \sqrt{n}\bar{Y}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$*.*

We now compute the moment generating function (MGF) of Z_n*:*

$$M_{Z_n}(t) = \mathbb{E}\left[e^{tZ_n}\right] = \mathbb{E}\left[e^{t \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i}\right]$$

Since the Y_i *are independent, the expectation of a product of independent variables is the product of expectations:*

$$M_{Z_n}(t) = \prod_{i=1}^n \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} Y_i}\right] = \left(M_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

where $M_Y(t) = \mathbb{E}[e^{tY}]$ *is the MGF of a single standardized variable* Y_i*.*

Letting $n \to \infty$*, we get the indeterminate for* 1^∞*, therefore we should take the limit of the logarithm, which gives:*

$$\begin{aligned}
\lim_{n\to\infty} n \log M\left(\frac{t}{\sqrt{n}}\right) &= \lim_{y\to 0} \frac{\log M(yt)}{y^2} && \text{where } y = \frac{1}{\sqrt{n}} \\
&= \lim_{y\to 0} \frac{tM'(yt)}{2yM(yt)} && \text{by L'Hôpital's rule} \\
&= \frac{t}{2} \lim_{y\to 0} \frac{M'(yt)}{M(yt)} \cdot \frac{1}{y} \\
&= \frac{t}{2} \lim_{y\to 0} \frac{M'(yt)}{y} && \text{since } M(yt) \to 1 \\
&= \frac{t^2}{2} \lim_{y\to 0} M''(yt) && \text{by L'Hôpital's rule} \\
&= \frac{t^2}{2}.
\end{aligned}$$

As a result, $\left(M(\frac{t}{\sqrt{n}})\right)^n$*, the MGF of* $\sqrt{n}\bar{X}_n$ *(or in this case* $\bar{Y}_i$*), approaches* $e^{\frac{t^2}{2}}$*, the* $\mathcal{N}(0,1)$ *MGF.*

More generally,

$$\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}).$$

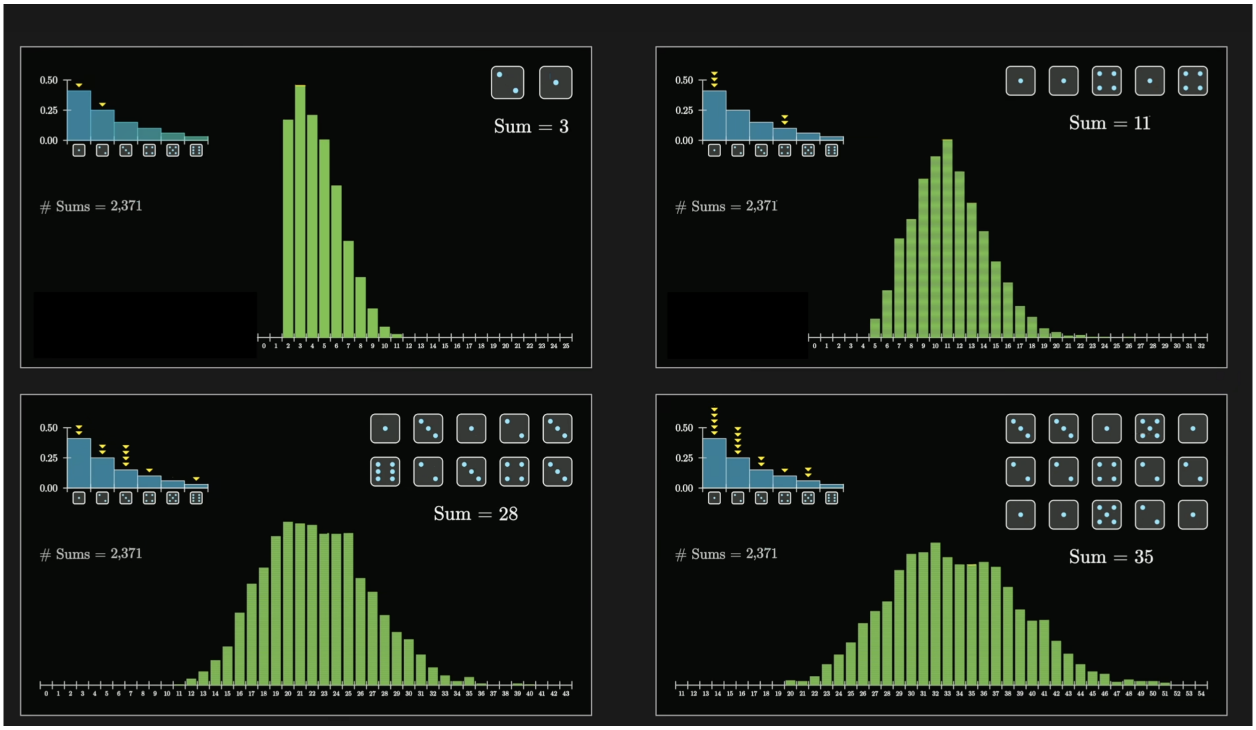

Figure 2. An illustration of the CLT [4]

[6]

What this is saying is that when there is a sufficiently large number n of random variables, regardless of the distribution(i.e., Poisson, Binomial, etc.), the distribution form a bell curve, more formally, it is referred as the normal distribution, denoted as $\mathcal{N}$. According to 3Blue1Brown's video on the CLT2, when we roll 2,5,10,15 dices, correspondingly, and sum the values up, we discover that the distribution apporaches the bell-curved normal distribution. This is pretty counter-intuitive especially because when n is sufficiently large, it doesn't matter what the distribution is, or in other words, the dice can be uneven, the distribution still approaches the normal distribution.

4. Law of the Iterated Logarithm

4.1. **Two Preliminaries.** To prove the Law of the Iterated Logarithm, we reqiure two preliminary theorems:

Lemma 4.1 (Large Deviation Bound)**.** *Let $S_n = X_1 + X_2 + \cdots + X_n$, where $\{X_j\}_{j=1}^{\infty}$ are independent and identically distributed (i.i.d.) random variables with zero mean ($E[X_j] = 0$) and unit variance ($E[X_j^2] = 1$). (And implicitly all moments of X_j exist and can be determined by taking the derivative of the moment generating function). Suppose that the positive constant sequence $a_1, a_2, \ldots, a_n, \ldots$ satisfies $a_n \to \infty$ and $\frac{a_n}{\sqrt{n}} \to 0$.*

Then there exists a sequence $\zeta_1, \zeta_2, \ldots$ with $\zeta_n \to 0$ such that

$$P[S_n \geq a_n \sqrt{n}] = e^{-a_n^2 (1+\zeta_n)/2}.$$

This theorem provides a more accurate bound for the tail probabilities compared to the central limit theorem. However, the proof of this is rigorous, and we will not prove it here.

The proof involves clever manipulations the moment generating function and the cumulant generating function, the application of the Chernoff bound(which is the Markov inequality when optimized), and some asymptotic analysis involving the o-notation mentioned in previously.

A key insight is that $S_n/\sqrt{n}$ converges in distribution to zero-mean unit-variance Gaussian(Central Limit Theorem). So

$$P\left[S_n \geq a\sqrt{n}\right] \to \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-s^2/2}\, ds$$

$$= \frac{1}{\sqrt{2\pi}a} e^{-a^2/2} \left(1 - \frac{1}{a^2} + \frac{1\cdot 3}{a^4} - \frac{1\cdot 3\cdot 5}{a^6} + \cdots\right),$$

where the last equality holds for $a > 0$. In this theorem, we divide S_n by a little larger quantity than $\sqrt{n}$. For a precise and full proof of the Large Deviation bound, refer to Billingsley's book on probability and measure theory.

Lemma 4.2 (Maximal Inequality). *The second preliminary, the Maximal inequality, is described as follows:*

Let $S_n = X_1 + \cdots + X_n$, where $\{X_j\}_{j=1}^\infty$ are independent and identically distributed (i.i.d.) random variables with mean 0 and variance 1. (And all moments of X_j exist, which can be determined by taking the derivatives of its moment generating function). Also, let $Pr[S_0 = 0] = 1$. Then for $\alpha \geq \sqrt{2}$,

$$Pr\left[\max\{S_0, S_1, \ldots, S_n\} \geq \alpha\sqrt{n}\right] \leq 2Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right].$$

The maximal inequality provides a powerful bound: it states that the probability of the maximum sum reaching a high value is bounded by twice the probability of the final sum (S_n) exceeding a slightly smaller threshold ($\alpha - 2$). This is significant because probabilities involving only the final sum S_n are generally easier to estimate compared the the maximum function.

The proof for this preliminary is relatively straightforward compared to the previous one, we will illustrate the proof in the following.

Proof. Denote $M_n = \max\{S_0, S_1, \ldots, S_n\}$. Then M_n is non-negative and non-decreasing in n.

$$\begin{aligned} \Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] &= \Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha \wedge \frac{M_{n-1}}{\sqrt{n}} < \alpha \vee \cdots \vee \frac{M_1}{\sqrt{n}} \geq \alpha \wedge \frac{M_0}{\sqrt{n}} < \alpha\right] \\ &= \Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha \wedge \frac{M_{n-1}}{\sqrt{n}} < \alpha \wedge \cdots \wedge \frac{M_1}{\sqrt{n}} < \alpha \wedge \frac{M_0}{\sqrt{n}} < \alpha\right] \\ &= \sum_{j=1}^n \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha \wedge \frac{M_{j-1}}{\sqrt{n}} < \alpha\right]. \end{aligned}$$

Using this result, we can further derive:

$$\begin{aligned}
\Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] &= \Pr\left[\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \wedge \left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right)\right] \\
&\quad + \Pr\left[\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right] \\
&= \Pr\left[\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \wedge \left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right)\right] \\
&\quad + \sum_{j=1}^{n} \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right] \\
&\leq \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] \\
&\quad + \sum_{j=1}^{n} \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right].
\end{aligned}$$

Since $\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}$ implies $\frac{S_j}{\sqrt{n}} \geq \alpha$, we have:

$$\begin{aligned}
&\Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right] \\
&\leq \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n - S_j}{\sqrt{n}} < \alpha - \sqrt{2} - \frac{S_j}{\sqrt{n}}\right)\right] \\
&= \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_j - S_n}{\sqrt{n}} > \sqrt{2}\right)\right] \\
&= \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \Pr\left[\frac{S_j - S_n}{\sqrt{n}} > \sqrt{2}\right]
\end{aligned}$$

(M_j and M_{j-1} only depend on $X_1, \ldots, X_j$; $S_j - S_n$ only depends on $X_{j+1}, \ldots, X_n$. So the above two events are independent.)

$$\begin{aligned}
&\leq \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \Pr\left[\left|\frac{S_j - S_n}{\sqrt{n}}\right| > \sqrt{2}\right] \\
&\leq \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \frac{\mathrm{Var}[S_n - S_j]}{2n} \quad \text{(by Chebyshev's ineq.)} \\
&= \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \frac{(n-j)}{2n} \cdot 1 \\
&= \frac{n-j}{2n} \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right].
\end{aligned}$$

Consequently,

$$\begin{aligned}\Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] &\leq \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] \\ &\quad + \sum_{j=1}^{n} \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right] \\ &\leq \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] + \frac{1}{2}\sum_{j=1}^{n} \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \\ &= \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] + \frac{1}{2}\Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right].\end{aligned}$$

■

After so many preliminaries, we come to the final boss.

4.2. **Proof of the Law of the Iterated Logarithm.**

Proof. Formally, let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent, identically distributed random variables with $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1$, and let $S_n = X_1 + X_2 + \cdots + X_n$. Then the Law of the Iterated Logarithm states:

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1 \quad \text{almost surely.}$$

Recall

$\limsup_{n\to\infty} a_n = a$ if, and only if, $(\forall \epsilon > 0)$ $(\exists N)$ $(\forall n > N)$ $a_n - a < \epsilon$ and $(\forall \epsilon > 0)$ $(\exists N)$ $(\forall n > N)$ $a_n - a > -\epsilon$.

$\limsup_{n\to\infty} a_n = \lim_{N\to\infty} \sup\{a_N, a_{N+1}, a_{N+2}, \ldots\}$

Hence, to prove

$$\Pr\left[\limsup_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{2n\log\log n}} = 1\right] = 1$$

it suffices to prove that for every positive ϵ, (where ϵ is understood as "countable" in its range),

$$P\left(\left\{x \in \mathbb{R}^{\infty} : (\exists N)(\forall n > N)\frac{x_1 + x_2 + \cdots + x_n}{\sigma\sqrt{2n\log\log n}} - 1 < \epsilon\right\}\right) = 1$$

and

$$P\left(\left\{x \in \mathbb{R}^{\infty} : (\forall N)(\exists n > N)\frac{x_1 + x_2 + \cdots + x_n}{\sigma\sqrt{2n\log\log n}} - 1 > -\epsilon\right\}\right) = 1$$

Equivalently (refer back to **Definition 3.10**),

$$P\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{x \in \mathbb{R}^{\infty} : \frac{x_1 + x_2 + \cdots + x_n}{\sigma\sqrt{2n\log\log n}} - 1 < \epsilon\right\}\right) = 1$$

and

$$P\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{x \in \mathbb{R}^{\infty} : \frac{x_1 + x_2 + \cdots + x_n}{\sigma\sqrt{2n\log\log n}} - 1 > -\epsilon\right\}\right) = 1.$$

Or equivalently, by De Morgan's law,

$$P\left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{x\in\mathbb{R}^{\infty}:\frac{x_1+x_2+\cdots+x_n}{\sigma\sqrt{2n\log\log n}}-1\geq\epsilon\right\}\right)=0$$

and

$$P\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{x\in\mathbb{R}^{\infty}:\frac{x_1+x_2+\cdots+x_n}{\sigma\sqrt{2n\log\log n}}-1>-\epsilon\right\}\right)=1.$$

Intuitively, the first inequality we aim to prove is related to the First Borel-Cantelli Lemma, and the second inequality is related to the Second Borel-Cantelli Lemma.

Or equivalently (as written in Billingsley's book),

$$\Pr\left[X_1+X_2+\cdots+X_n\geq(1+\epsilon)\sigma\sqrt{2n\log\log n}\text{ i.o.}\right]=0$$

and

$$\Pr\left[X_1+X_2+\cdots+X_n>(1-\epsilon)\sigma\sqrt{2n\log\log n}\text{ i.o.}\right]=1.$$

> To reduce the confusion, here we distinguish between the "*event*" and the "*set*". That is why we put "Pr[event]" and use "P(set)", where the event is defined through some random variables, and the probability of a set is measured by probability measure P. Billingsley's book sometimes mixes the two together, which may be confusing.

Assumption Without loss of generality, we assume that the variance σ^2 is unity.

We'll first prove the first statement. Choose a positive θ such that $1<\theta^3<1+\epsilon$.

Let $n_k\triangleq\lfloor\theta^k\rfloor>\theta^k-1$, and

for $k\geq\log_\theta(e^e+1)$ (i.e., $\theta^k-1\geq e^e$, which implies $\log\log(n_k)\geq 1$), let $x_k\triangleq\theta\sqrt{2\log\log(n_k)}$ $(\geq\theta\sqrt{2})\geq\sqrt{2}$.

Then 4.1 and 4.2 give that for $k\geq\log_\theta(e^e+1)$ and for some $\zeta_k\to 0$,

$$\begin{aligned}\Pr\left[\frac{M_{n_k}}{\sqrt{n_k}}\geq x_k\right]&\leq 2\Pr\left[\frac{S_{n_k}}{\sqrt{n_k}}\geq x_k-\sqrt{2}\right] \quad 4.2\\&=2\exp\left\{-\frac{1}{2}(x_k-\sqrt{2})^2(1+\zeta_k)\right\} \quad 4.1\\&=2\exp\left\{-\theta^2\log(k)\frac{(x_k-\sqrt{2})^2(1+\zeta_k)}{2\theta^2\log(k)}\right\}.\end{aligned}$$

Since

$$\lim_{k\to\infty}\frac{(x_k-\sqrt{2})^2(1+\zeta_k)}{2\theta^2\log(k)}=\lim_{k\to\infty}\frac{\left(\theta\sqrt{2\log\log(\theta^k)}-\sqrt{2}\right)^2}{2\theta^2\log(k)}\quad(\text{Because }\zeta_k\downarrow 0)$$

$$=\lim_{k\to\infty}\frac{(\theta\sqrt{\log(\theta^k)}-1)^2}{\theta^2\log(k)}=\lim_{k\to\infty}\frac{\left(\theta\sqrt{\log(k)+\log\log(\theta)}-1\right)^2}{\theta^2\log(k)}=1,$$

there exists K such that for $k \geq K$,

$$\frac{(x_k - \sqrt{2})^2(1+\zeta_k)}{2\theta^2 \log(k)} \geq \frac{1}{\theta}.$$

Accordingly, for $k \geq K_0 \triangleq \max\{K, \log_\theta(e^\epsilon + 1)\}$,

$$\Pr\left[\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k\right] \leq 2\exp\{-\theta \log(k)\} = \frac{2}{k^\theta}.$$

Now for $n \geq e^{\theta^2}$ fixed, there exists k such that $n_{k-1} < n \leq n_k$, and

$$\sqrt{2n \log\log(n)} \geq \sqrt{2(n_{k-1}+1)\log\log(n_{k-1}+1)} > \sqrt{2\left(\frac{n_k}{\theta}\right)\log\log\left(\frac{n_k}{\theta}\right)}.$$

The last strict inequality follows from:

$$n_{k-1} = \lfloor \theta^{k-1} \rfloor > \theta^{k-1} - 1 = \frac{\theta^k}{\theta} - 1 \geq \frac{\lfloor \theta^k \rfloor}{\theta} - 1 = \frac{n_k}{\theta} - 1.$$

which implies that (This is the only step requiring $\theta^3 < (1+\epsilon)$)

$$\begin{aligned}(1+\epsilon)\sqrt{2n\log\log(n)} &> \theta^3\sqrt{2\left(\frac{n_k}{\theta}\right)\log\log\left(\frac{n_k}{\theta}\right)} \\ &= \theta^2\sqrt{2n_k \log\log\left(\frac{n_k}{\theta}\right)} \\ &\geq \theta\sqrt{2n_k\log\log(n_k)} \quad (\text{since } \theta^2 > \theta)\end{aligned}$$

Claim: Given $\theta > 1$, $f_0(x) \triangleq \log\left(\frac{x}{\theta}\right) - \frac{1}{\theta}\log x \geq 0$ for $x \geq e^\theta$.
Proof: The claim can be validated by $f_0'(x) = \frac{(\theta-1)\log(x)+\log(\theta)}{x\theta\log(x)[\log(x)-\log(\theta)]} > 0$ for $x \geq e^\theta$ and $f_0(e^\theta) = \log[e^\theta - \log(\theta)] - 1 > 0$ for any $\theta > 1$. □

Accordingly, $[S_n \geq (1+\epsilon)\sqrt{2n\log\log(n)}]$ implies that

$$M_{n_k} \geq S_n \geq (1+\epsilon)\sqrt{2n\log\log(n)} \geq \theta\sqrt{2n_k\log\log(n_k)},$$

where k is the unique integer satisfying $\lfloor\theta^{k-1}\rfloor < n \leq \lfloor\theta^k\rfloor$ (namely, $k = \lfloor\log_\theta(n)\rfloor$). As a consequence,

$$\Pr\left[S_n \geq (1+\epsilon)\sqrt{2n\log\log(n)} \text{ i.o. in } n\right] \leq \Pr\left[M_{n_k} \geq \theta\sqrt{2n_k\log\log(n_k)} \text{ i.o. in } k\right].$$

By the first Borel-Cantelli lemma,

$$\begin{aligned}\sum_{k=1}^{\infty} \Pr\left[M_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)}\right] &= \sum_{k=1}^{K_0} \Pr\left[M_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)}\right] \\ &\quad + \sum_{k=K_0+1}^{\infty} \Pr\left[M_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)}\right] \\ &\leq K_0 + \sum_{k=K_0+1}^{\infty} \frac{2}{k^\theta} \\ &\leq K_0 + \int_{K_0}^{\infty} \frac{2}{x^\theta} dx \\ &= K_0 + \frac{2}{(\theta-1)K_0^{\theta-1}} \\ &< \infty,\end{aligned}$$

we obtain (The below equality holds without the condition that $\theta^3 < (1+\epsilon)$):

$$\Pr\left[M_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)} \text{ i.o. in } k\right] = 0.$$

This completes the proof of the first part. ■

2. $\Pr\left[S_n > (1-\epsilon)\sqrt{2n\log\log(n)} \text{ i.o.}\right] = 1.$
Choose ξ satisfying $\xi > \max\{1, 9/\epsilon^2\}$. Take $n_k \triangleq \lfloor \xi^k \rfloor$.
For any k, let $m_k = n_k - n_{k-1}$ and $a_k = \left(1 - \frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}$.
Then $a_k \to \infty$ and $\frac{a_k}{\sqrt{m_k}} \to 0$ as $k \to \infty$.

- $a_k \geq \left(1-\frac{1}{\xi}\right)\frac{\sqrt{2n_k\log\log(n_k)}}{\sqrt{n_k}} = \left(1-\frac{1}{\xi}\right)\sqrt{2\log\log(n_k)} \to \infty.$
- $\lim_{k\to\infty} \frac{a_k}{\sqrt{m_k}} = \lim_{k\to\infty} \frac{\left(1-\frac{1}{\xi}\right)\sqrt{2n_k\log\log(n_k)}}{\sqrt{m_k}}$

$$\begin{aligned}&= \lim_{k\to\infty} \frac{\left(1-\frac{1}{\xi}\right)\sqrt{2\xi^k \log\log(\xi^k)}}{\sqrt{\xi^k - \xi^{k-1}}} \\ &= \lim_{k\to\infty} \sqrt{\frac{2\log\log(\xi^k)}{\xi^k}} = 0\end{aligned}$$

4.1 then implies:

$$
\begin{aligned}
&\Pr\left[S_{n_k} - S_{n_{k-1}} \geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right] \\
&= \Pr\left[X_{n_{k-1}+1} + \cdots + X_{n_k} \geq a_k\sqrt{m_k}\right] \\
&= \exp\left\{-\frac{1}{2}a_k^2(1+\zeta_k)\right\} \quad \text{for some } \zeta_k \to 0 \\
&= \exp\left\{-\frac{1}{2}\left(\frac{\left(1-\frac{1}{\xi}\right)\sqrt{2n_k\log\log(n_k)}}{\sqrt{m_k}}\right)^2(1+\zeta_k)\right\} \\
&= \exp\left\{-\frac{(\xi-1)^2[2n_k\log\log(n_k)](1+\zeta_k)}{\xi^2(n_k-n_{k-1})}\right\} \\
&\geq \exp\left\{-\frac{(\xi-1)^2[2\xi^k\log\log(\xi^k)](1+\zeta_k)}{\xi^2(\xi^k-\xi^{k-1})}\right\} \quad (\text{since } \xi^{k-1} < n_k \leq \xi^k \text{ and } n_{k-1} \leq \xi^{k-1}) \\
&= \exp\left\{-\frac{(\xi-1)^2[2\xi^k\log\log(\xi^k)](1+\zeta_k)}{\xi^2(\xi^k-\xi^{k-1})}\frac{1}{1-\frac{1}{\xi}}\right\} \\
&= \exp\left\{-\frac{(\xi-1)^2[2\log\log(\xi^k)](1+\zeta_k)}{\xi^2\left(1-\frac{1}{\xi}\right)}\right\} \\
&= \exp\left\{-\frac{(\xi-1)^2[2\log\log(\xi^k)](1+\zeta_k)}{\xi(\xi-1)}\right\} \\
&= \exp\left\{-\frac{(\xi-1)[2\log\log(\xi^k)](1+\zeta_k)}{\xi}\right\} \\
&= \exp\left\{-2\left(1-\frac{1}{\xi}\right)\log\log(\xi^k)(1+\zeta_k)\right\}
\end{aligned}
$$

As $\zeta_k \to 0$, there exists K_1 such that $\zeta_k < \frac{1}{2\xi-1}$ for all $k \geq K_1$.
Also, $k \geq \frac{\log(\xi)+\log(2\xi-1)-\log(\xi-1)}{\log(\xi)}$, if and only if $\frac{1}{\xi^k-\xi^{k-1}} \leq \frac{1}{2\xi-1}$.
Hence, for $k \geq K_2 \triangleq \max\left\{K_1, \frac{\log(\xi)+\log(2\xi-1)-\log(\xi-1)}{\log(\xi)}\right\}$,

$$
\begin{aligned}
&\Pr\left[S_{n_k} - S_{n_{k-1}} \geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k\log\log(n_k)}\right] \\
&\geq \exp\left\{-\frac{(\xi-1)\log\log(\xi^k)(1+\zeta_k)}{\xi\left(1-\frac{1}{\xi^k-\xi^{k-1}}\right)}\right\} \\
&\geq \exp\left\{-\frac{(\xi-1)\log\log(\xi^k)\left(1+\frac{1}{2\xi-1}\right)}{\xi\left(1-\frac{1}{2\xi-1}\right)}\right\} \\
&= \exp\left\{-\log\log(\xi^k)\right\} = \frac{1}{k\log(\xi)}.
\end{aligned}
$$

Since $\left\{\left[S_{n_k} - S_{n_{k-1}} \geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right]\right\}_{k=1}^{\infty}$ are independent events, and

$$\begin{aligned}
&\sum_{k=1}^{\infty} \Pr\left[S_{n_k} - S_{n_{k-1}} \geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right] \\
&\geq \sum_{k=K_2}^{\infty} \Pr\left[S_{n_k} - S_{n_{k-1}} \geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right] \\
&\geq \sum_{k=K_2}^{\infty} \frac{1}{k\log(\xi)} = \infty,
\end{aligned}$$

it follows from the second Borel-Cantelli lemma that with probability 1, $S_{n_k} - S_{n_{k-1}} \geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}$ infinitely often in k.

Now we can let $\overline{X}_n = -X_n$, and let $\overline{M}_n$ and $\overline{S}_n$ be respectively the counterparts of M_n and S_n for $\{\overline{X}_n\}$ (and $n_k = \lfloor \xi^k \rfloor$), and apply the proof in the first part with $\theta = \sqrt{2}$ to show that (cf. Slide 9-72)(This holds without $\theta^3 < (1+\epsilon)$):

$$\Pr\left[\overline{M}_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)} \text{ i.o. in } k\right] = 0.$$

Hence, it is with probability 1 that $-\overline{S}_{n_{k-1}} = S_{n_{k-1}} \leq \overline{M}_{n_{k-1}} < \theta\sqrt{2n_{k-1}\log\log(n_{k-1})}$ for all but finitely many k.

Observe that $\theta\sqrt{2n_{k-1}\log\log(n_{k-1})} \leq \frac{2}{\sqrt{\xi}}\sqrt{2n_k \log\log(n_k)}$.

The validity of the above inequality follows:

- Apply $n_{k-1}\left(\triangleq \lfloor \xi^{k-1} \rfloor\right) \leq \xi^{k-1} = \frac{\xi^k}{\xi} \leq \frac{n_k+1}{\xi} \leq \frac{2n_k}{\xi}$ for n_{k-1} outside $\log(\cdot)$.
- Apply $n_{k-1} \leq n_k$ for n_{k-1} inside $\log(\cdot)$.

As a result, it is with probability 1 that $-S_{n_{k-1}} \leq \frac{2}{\sqrt{\xi}}\sqrt{2n_k \log\log(n_k)}$ for all but finitely many k.

To summarize, it is with probability 1 that for infinitely many k,

$$\begin{aligned}
S_{n_k} &\geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)} + S_{n_{k-1}} \\
&\geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)} - \frac{2}{\sqrt{\xi}}\sqrt{2n_k \log\log(n_k)} \\
&= \left(1-\frac{1}{\xi}-\frac{2}{\sqrt{\xi}}\right)\sqrt{2n_k \log\log(n_k)} \\
&\geq \left(1-\frac{1}{\sqrt{\xi}}-\frac{2}{\sqrt{\xi}}\right)\sqrt{2n_k \log\log(n_k)} \\
&= \left(1-\frac{3}{\sqrt{\xi}}\right)\sqrt{2n_k \log\log(n_k)} \\
&\geq (1-\epsilon)\sqrt{2n_k \log\log(n_k)}.
\end{aligned}$$

[8]

To encapsulate, we proved the two statements. Combining leads to the formal statement of the law of the iterated logarithm.

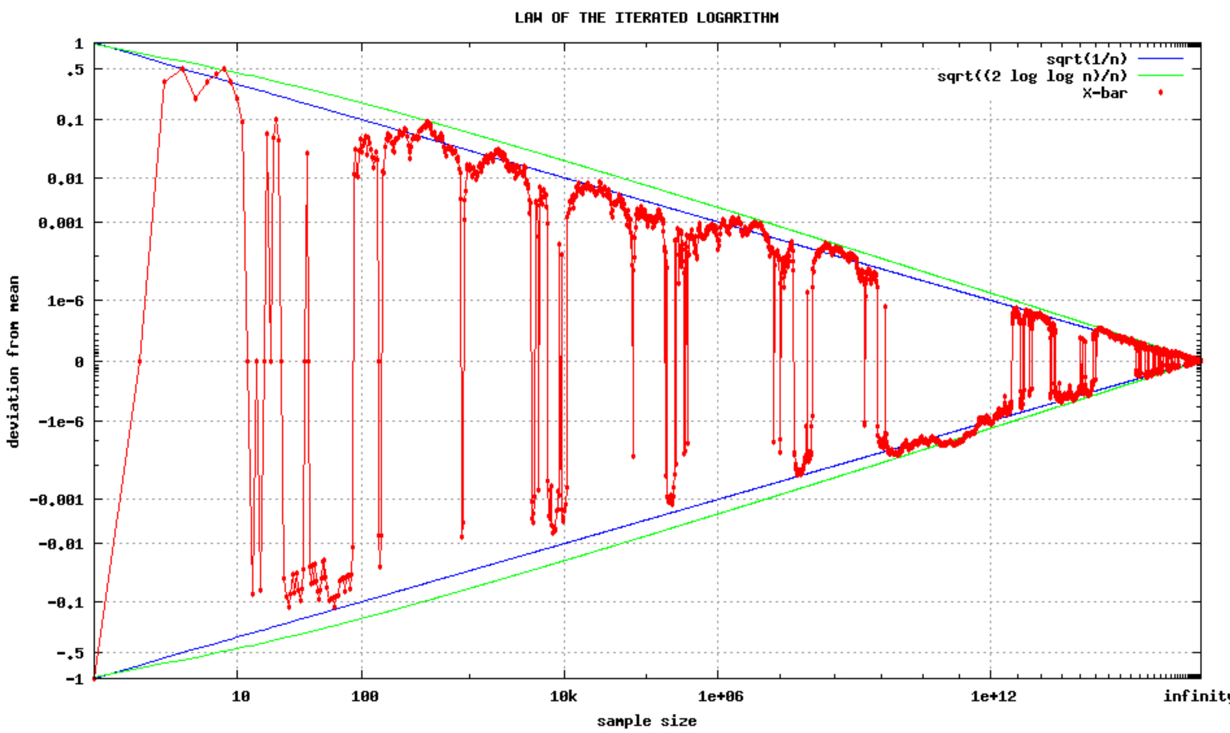

Figure 3. A graph on the law of the iterated logarithm [2]

Remark. The original statement of the law of the iterated logarithm was given by A.Y. Kinchin in 1924. Another statement was given by A. N. Kolmogorov in 1929.

5. Conclusion

It is a special and fascinating phenomenon in probability theory that $\frac{X_1+X_2+\cdots+X_n}{\sigma\sqrt{2n\log\log n}}$ oscillates between ± 1 when n is sufficiently large, as illustrated in Figure 3, regardless of the distribution.

Figure 4 gives a more general graph of the limits, when the partial sums are divided by $1, \sqrt{n\log\log n}, \sqrt{n}, n$, correspondingly.

6. Acknowledgement

I would like to give special thanks to Emma Cardwell, who led me through this process, this paper would be impossible without her guidance. I would also like to thank Simon Rubinstein-Salzedo for making Euler Circle possible, providing the template, and the guidance to writing a comprehensive and self-contained paper.

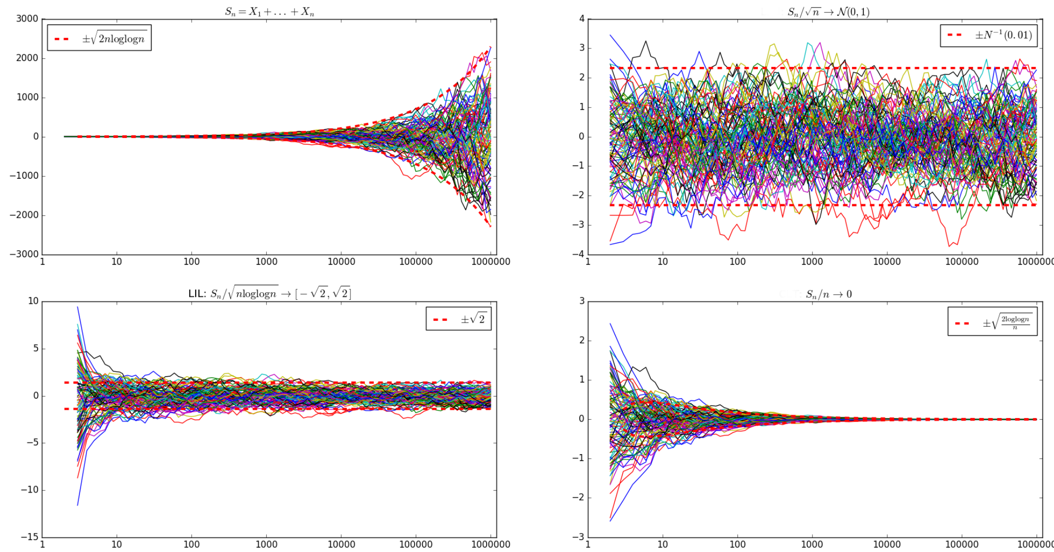

Figure 4. A graph on different limit theorems [3]

References

[1] May 2025.
[2] Jun 2025.
[3] Jun 2025.
[4] 3Blue1Brown. But what is the central limit theorem? `https://www.youtube.com/watch?v=zeJD6dqJ5lo`, March 2020. [Online; accessed 10 July 2025].
[5] Patrick Billingsley. *Probability and measure3rd edition Billingsley, Patrick*. Wiley, 1995.
[6] Joseph K. Blitzstein and Jessica Hwang. *Introduction to probability*. CRC Press, Taylor Francis Group, 2019.
[7] Joseph K. Blitzstein and Carl N. Morris. Probability for statistical science. Draft Version, August 2020, 2020.
[8] Bo-Ning Chen. Advanced probability for communications: Lecture 3 – large deviations and the law of the iterated logarithm. YouTube video, NYCU OpenCourseWare, October 2020. Department of Electrical Engineering, National Yang Ming Chiao Tung University.

Email address: `ricardooorichado@gmail.com`