Law of the iterated logarithm

Alex Fu

2023-01-19

As promised, we will continue the journey started in *2023-01-02*, *2023-01-03*, and *2023-01-18*.

The primary focus of this note will be on sums of i.i.d. Rademacher random variables $X_1, X_2, \ldots$, which are ± 1 with probability $\frac{1}{2}$ each. While more general i.i.d. sums are featured in the study of Poisson point processes and renewal theory, the sums of i.i.d. Rademachers $S_n = \sum_{i=1}^n X_i$, where $S_0 = 0$, form a classical example of a random process: the symmetric or *simple random walk* on the integers $\mathbb{Z}$. Our methods will also apply to the asymmetric case.

(Recall that Rademachers are convenient to work with because they are *standardized*, i.e. zero-mean unit-variance, random variables. We often want to impose this condition in general by translation and scaling, but for X_i Rademacher, it's already a given.)

From our previous discussions, the Strong Law of Large Numbers states that $S_n \in o(n^{1/2}(\log n)^{1/2+\varepsilon})$ almost surely. The Central Limit Theorem states that $S_n \in O(n^{1/2})$ almost surely, or more specifically $S_n/n^{1/2}$ converges to the standard normal $\mathcal{N}(0,1)$ in distribution. But, despite how close they may seem, there is a vast, untouched borderland of complexity classes between $n^{1/2}$ and $n^{1/2}(\log n)^{1/2+\varepsilon}$, and perhaps the abrupt transition from 0 to the a.s. finite, but unbounded $Z \sim \mathcal{N}(0,1)$ may seem a bit jarring.

Can we be more precise in our understanding of the eventual behavior of the simple random walk S_n? Of course, S_n is a random variable that can fluctuate probabilistically, and we already know that S_n is "typically" $\pm\sqrt{n}$. What we really want is a theorem situated *between* the SLLN and the CLT — a tight asymptotic envelope.

Theorem 1 (Law of the iterated logarithm, or LIL)**.**

Let $X_1, X_2, \ldots$ be i.i.d. Rademacher random variables, and let $S_n = \sum_{i=1}^n X_i$. Then

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1\right) = 1.$$

(For convenience, we will write $f(n) \coloneqq \sqrt{2n\log\log n}$ throughout.)

By symmetry, $\liminf_{n\to\infty} S_n/\sqrt{2n\log\log n} = -1$ almost surely as well. Theorem 1 is really two almost-sure bounds packaged into one:

1. The upper bound $\limsup_{n\to\infty} S_n/f(n) \le 1$. For the limit supremum of a sequence to be bounded above by 1 means that the sequence *eventually* lies below 1, or, in other words, the sequence exceeds 1 only finitely often.

2. The lower bound $\limsup_{n\to\infty} S_n/f(n) \geq 1$. The sequence $S_n/f(n)$ exceeds 1 infinitely often; combined with the previous upper bound, this means $S_n/f(n)$ actually hits or attains 1 infinitely often!

In other words, $f(n)$ is truly an **envelope** for S_n: the "trajectory" of S_n falls within $[-f(n), f(n)]$, or $|S_n| \leq f(n)$, eventually with probability 1. Moreover, this envelope is tight: S_n is eventually within the bounds, but also reaches and *touches* the ceiling and floor $\pm f(n)$ infinitely often.

This is a remarkable result, but it may also seem entirely arbitrary. The namesake of Theorem 1 doubles as its oddest feature, the term $(\log\log n)^{1/2}$, especially since S_n itself admits a very simple formulation. How do we reason about it? Well, for a start, we can convince ourselves that the iterated logarithm is no more odd than $n^{1/2}$. This will be a question we aim to answer through this note: *where does the iterated logarithm come from*?

Now, as we might expect, to make finer distinctions rather than broad strokes in determining asymptotic rate will require more technical tools and prowess. These involve a different kind of intuition, the intuition for technique in proof, reducing complexity, and transforming objectives.

First, let us recall our weapon of choice for tackling statements of the form "eventually almost surely."

Lemma 1 (Borel–Cantelli lemma)**.**

Let $A_1, A_2, \ldots$ be any sequence of events. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$.

We also have a partial converse to Lemma 1 that often proves to be useful when we have independence.

Lemma 2 (Second Borel–Cantelli lemma)**.**

Let $A_1, A_2, \ldots$ be any sequence of mutually independent events. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof. We wish to leverage the monotone continuity of probability and the independence of $(A_k)_{k=1}^{\infty}$, so let us find

$$1 - \mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k^c\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right) = \lim_{n\to\infty} \prod_{k\geq n} (1 - \mathbb{P}(A_k)).$$

Now, $\prod_{k=n}^{\infty}(1 - x_k) = 0$ iff $\sum_{k=n}^{\infty} x_k = \infty$, which is true for $x_k = \mathbb{P}(A_k)$ by hypothesis. Thus the right-hand side tends to 0 as $n \to \infty$, and we are done. □

In particular, note that the conclusion of the first Borel–Cantelli lemma has the form "finitely often a.s." or "eventually a.s." However, the lower bound is a statement of "infinitely often a.s.," which fails to fit the form of the first Borel–Cantelli lemma. This is where Lemma 2 will be applied instead.

Considering ingredients of our previous proofs, we will not need the tail-sum approximation to exploit given finiteness conditions, as Rademachers are already bounded. However, a concentration inequality analogous to Chebyshev's will turn out useful in converting summations of probabilities into more numerical series.

Lemma 3 (Chernoff bound for i.i.d. sum of Rademachers)**.**

For any positive integer a, we have the bound $\mathbb{P}(S_n \geq a) \leq e^{-a^2/(2n)}$.

Proof. Let $t > 0$. As $x \mapsto e^{tx}$ is a monotonic function taking nonnegative values, by Markov's inequality, we obtain the usual Chernoff bound of

$$\mathbb{P}(S_n \geq a) = \mathbb{P}(e^{tS_n} \geq e^{ta}) \leq \frac{\mathbb{E}(e^{tS_n})}{e^{ta}} = \frac{\mathbb{E}(e^{tX_1})^n}{e^{ta}}.$$

Note that $\mathbb{E}(e^{tS_n})$, the moment-generating function of S_n, equals $\prod_{i=1}^n \mathbb{E}(e^{tX_i})$ by independence. In particular, by comparing the coefficients in the Taylor expansions,

$$\mathbb{E}(e^{tX_1}) = \frac{e^t + e^{-t}}{2} \leq e^{t^2/2}.$$

Thus, taking $t = a/n$, the Chernoff bound becomes

$$\mathbb{P}(S_n \geq a) \leq \frac{e^{nt^2/2}}{e^{ta}} = e^{-a^2/(2n)}.$$

□

Let us also consider the *maximal function* or process $M_n \coloneqq \max_{0 \leq k \leq n} S_k$. A result in our toolkit from *2023-01-18* is Kolmogorov's maximal inequality, which bounds the tail probability of M_n in terms of the Chebyshev bound for S_n. Generally, maximal inequalities are valuable because they save an undesirable extra factor of n due to a naïve application of the union bound. To say something about the full history or trajectory, using only the current sample or value — that's quite a powerful idea.

Lemma 4 (Reflection principle)**.**

For any positive integer a, we have the equality $\mathbb{P}(M_n \geq a) = \mathbb{P}(S_n = a) + 2\,\mathbb{P}(S_n > a) \leq 2\,\mathbb{P}(S_n \geq a)$.

Proof. By the law of total probability,

$$\mathbb{P}(M_n \geq a) = \sum_{k=-n}^{a-1} \mathbb{P}(M_n \geq a,\, S_n = k) + \sum_{k=a}^{n} \mathbb{P}(M_n \geq a,\, S_n = k).$$

By reflectional symmetry, the event $\{M_n \geq a, S_n = k\}$ has the same probability as $\{M_n \geq a, S_n = a - (k - a)\}$:

$$= \sum_{k=-n}^{a-1} \mathbb{P}(S_n = 2a - k) + \sum_{k=a}^{n} \mathbb{P}(S_n = k)$$

$$= \mathbb{P}(S_n \geq a + 1) + \mathbb{P}(S_n \geq a).$$

□

There are different variations of Lemma 4 with constant factors of 3, 4, $\frac{4}{3}$, etc., but when you see its application, it becomes clear that we only require the finiteness of the factor. For example, an alternative to Lemma 4 is

Lemma 5 (Etemadi's inequality)**.**

Let $a \geq 0$. Then $\mathbb{P}(\max_{0 \leq k \leq n} |S_k| \geq 3a) \leq 3\,\mathbb{P}(|S_n| \geq a)$.

Proof. We can assume that S_n is any i.i.d. sum. We follow the same strategy as in the proof of Ottaviani's inequality: partition the event of interest A^* into $A_k := \{|S_k| \geq 3a, \max_{1\leq j\leq k} |S_j| < 3a\}$. Then

$$\mathbb{P}(A^*) = \sum_{k=1}^{n} \mathbb{P}(A_k) \leq \mathbb{P}(|S_n| \geq a) + \sum_{k=1}^{n} \mathbb{P}(A_k \cap |S_n| < a).$$

Ignore the $\mathbb{P}(|S_n| \geq a)$ term that stays out front. We observe that $A_k \cap \{|S_n| < a\} \subseteq A_k \cap \{|S_n - S_k| > 2a\}$, and note that $|S_n - S_k|$ is $\sigma(X_{k+1}, \ldots, X_n)$-measurable, independent of $A_k \in \sigma(X_1, \ldots, X_k)$. Thus the above is

$$\begin{aligned} \cdots &\leq \sum_{k=1}^{n} (\mathbb{P}(A_k) \cdot \mathbb{P}(|S_n - S_k| > 2x)) \\ &\leq \mathbb{P}(A^*) \cdot \max_{1\leq k\leq n} \mathbb{P}(|S_n - S_k| > 2x). \end{aligned}$$

By contrapositive, $\{|S_n - S_k| > 2x\} \subseteq \{|S_n| > x\} \cup \{|S_k| > x\}$. Adding back in the $\mathbb{P}(|S_n| \geq a)$ term, we have

$$\begin{aligned} \cdots &\leq \mathbb{P}(|S_n| \geq a) + 2\,\mathbb{P}(A^*) \max_{1\leq k\leq n} \mathbb{P}(|S_k| > a) \\ &\leq 3 \max_{1\leq k\leq n} \mathbb{P}(|S_k| \geq a). \end{aligned}$$

Returning to our special case, suppose $0 \ll a < n/(1+2^{1/2})$. We claim that $\mathbb{P}(|S_n| \geq a) = \max_{1\leq k\leq n} \mathbb{P}(|S_k| \geq a)$, which gives the result above. We want

$$\mathbb{P}(|S_{k-1}| \geq a) = \mathbb{P}(|S_k| \geq a+1) + \frac{1}{2}\,\mathbb{P}(|S_k| = a-1) \leq \mathbb{P}(|S_k| \geq a),$$

which holds iff the following inequalities are true:

$$\begin{aligned} \mathbb{P}(|S_k| = a-1) &\leq 2\,\mathbb{P}(|S_k| = a+1) \\ 2 \cdot \frac{k!}{(a-1)!(k-a+1)!} 2^{-k} &\leq 4 \cdot \frac{k!}{(a+1)!(k-a-1)!} 2^{-k} \\ (a+1)a &\leq 2(k-a+1)(k-a). \end{aligned}$$

I may have made a mistake above in terms of only considering $2(n-a+1)(n-a)$, but no matter: if the maximum tail probability is attained by $|S_{k_n}|$, $a \ll k_n \leq n$, this merely gives a weaker upper bound in the main proof. In any case, I only mention this lemma as a possible alternative to Lemma 4, so it's somewhat fine to fail here. □

The lemmas above are enough to prove the upper bound result in Theorem 1. But, as we noted with the second Borel–Cantelli lemma, the lower bound result requires a different approach and thus a different set of lemmas.

Lemma 6 (Local Central Limit Theorem, or local CLT)**.**

$\mathbb{P}(S_n = k) \in \Theta(e^{-k^2/(2n)}/\sqrt{\pi n})$.

Proof. See the proof of the De Moivre–Laplace Central Limit Theorem in *2023-01-03*. □

In probability, lower bounds are harder to come by. You may have noticed that most of the common concentration inequalities — Markov's, Chebyshev's, Chernoff, Hoeffding, Azuma's, etc. — are all upper bounds. But, we need the divergence of a summation of tail probabilities for the second Borel–Cantelli lemma, which means we need a lower bound. Here, we can leverage the discreteness of the S_n and the precise local asymptotic analysis for $\mathbb{P}(S_n = k)$ in order to approximate $\mathbb{P}(S_n \geq k)$.

Lemma 7 (Lower bound on tail probability of sum)**.**

Let $k > n^{1/2}$. Then $\mathbb{P}(S_n \geq k) \geq e^{-k^2/(2n)}(n^{1/2}/k)$ up to some positive constant multiplicative factor.

Proof. Throughout, let $c > 0$ denote the constant factor (which we do not care about) at each step. First,

$$\mathbb{P}(S_n \geq k) \geq \mathbb{P}\Big(k \leq S_n \leq k + \frac{n}{k}\Big) \geq cn^{-1/2} \sum_{m=k}^{k+(n/k)} e^{-m^2/(2n)}$$

per Lemma 6. For m in this range, we have that

$$\exp\left(-\frac{m^2}{2n}\right) \geq \exp\left(-\frac{(k+\frac{n}{k})^2}{2n}\right) = \exp\left(-\frac{k^2}{2n} - 1 - \frac{n}{2k^2}\right) \geq c\exp\left(-\frac{k^2}{2n}\right),$$

where $k^2 > n$ by hypothesis. Observe that placing an upper bound of $k + \frac{n}{k}$ on m is what allowed us to write this inequality. Now, as there are $\frac{n}{k}$ values of m in the sum,

$$\mathbb{P}(S_n \geq k) \geq \frac{n}{k} \cdot cn^{-1/2}e^{-k^2/(2n)}.$$

□

Verily, with every hardship, there is relief. To show that a limit supremum has a lower bound, it suffices to show that a subsequence is eventually greater than said lower bound. We *don't* need to upgrade to the full sequence and *don't* need to consider maximal inequalities; any "infinitely often" subsequence is all we need.

Finally, here we go.

Proof of upper bound. Let $\varepsilon > 0$. It suffices to show that $\limsup_{n\to\infty} S_n/f(n) \leq 1 + \varepsilon$ a.s.

(More formally, we can take $\varepsilon \to 0^+$ along $(\frac{1}{k})_{k\geq 1}$ such that the corresponding events form a decreasing sequence, then invoke the monotone continuity of probability. We will not worry about this familiar argument.)

i. Let $\alpha > 1$, and consider the exponential subsequence $(M_{\alpha^k})_{k\geq 1}$. We write $\alpha^k = \lceil \alpha^k \rceil$ out of convenience. In order to invoke the first Borel–Cantelli lemma, consider the following probability:

$$\mathbb{P}\left(\max_{0\leq n\leq \alpha^k} S_n \geq (1+\varepsilon)f(\alpha^k)\right) \leq 2\,\mathbb{P}\big(S_{\alpha^k} \geq (1+\varepsilon)f(\alpha^k)\big),$$

which we bounded by the reflection principle.

ii. Now, by the Chernoff bound for S_n, the above is at most

$$\cdots \leq 2\exp\left(-\frac{(1+\varepsilon)^2 f(\alpha^k)^2}{2\alpha^k}\right) = 2\exp\big(-(1+\varepsilon)^2 \log\log \alpha^k\big) = 2(k\log\alpha)^{-(1+\varepsilon)^2}.$$

Note that $-(1^+)\log\log\alpha^k = \log(k\log\alpha)^{-(1^+)}$ produced the summable term $ck^{-(1^+)}$ after exponentiation.

iii. This upper bound is summable in k. Thus, by the Borel–Cantelli lemma,

$$\left\{\max_{0\leq n\leq\alpha^k} S_n \leq (1+\varepsilon)f(\alpha^k) \text{ eventually in } k\right\} \text{ a.s.}$$

iv. To upgrade to the full sequence, consider $\alpha^{k-1} \le n \le \alpha^k$. Then

$$\frac{S_n}{f(n)} = \frac{S_n}{f(\alpha^k)} \cdot \frac{f(\alpha^k)}{\alpha^k} \cdot \frac{\alpha^k}{n} \cdot \frac{n}{f(n)} \le (1+\varepsilon)\alpha \text{ eventually (in } n) \text{ a.s.}$$

Note that $S_n/f(\alpha^k) \le (1+\varepsilon)$ by step iii above; $\alpha^k/n \le \alpha^k/\alpha^{k-1} = \alpha$; and $g(n) = f(n)/n$ is an eventually decreasing function, so $g(\alpha^k) \cdot 1/g(n) \le 1$. In other words, we have just shown that

$$\limsup_{n\to\infty} \frac{S_n}{f(n)} \le (1+\varepsilon)\alpha \text{ a.s.}$$

Taking $\alpha \to 1^+$, we are done.

□

Proof of lower bound. Let us show that $\limsup_{n\to\infty} S_n/f(n) \ge 1-\varepsilon$ a.s. for $0 < \varepsilon < 1$.

i. Let $\alpha > 1$, and consider the subsequence $(S_{\alpha^k})_{k\ge 1}$. By the symmetric version of the upper bound,

$$\liminf_{k\to\infty} \frac{S_{\alpha^k}}{f(\alpha^k)} \ge \liminf_{n\to\infty} \frac{S_n}{f(n)} \ge -(1+\varepsilon) \text{ a.s.}$$

To prove that $S_{\alpha^k}/f(\alpha^k)$ *exceeds* 1 infinitely often, let us consider the differences $S_{\alpha^k} - S_{\alpha^{k-1}}$.

ii. Define the events $A_k \coloneqq \{S_{\alpha^k} - S_{\alpha^{k-1}} \ge (1-\varepsilon)f(\alpha^k - \alpha^{k-1})\}$. For convenience, we write $n = \alpha^k - \alpha^{k-1}$. Note that each $S_{\alpha^k} - S_{\alpha^{k-1}}$ is $\sigma(X_{\alpha^{k-1}+1}, \dots, X_{\alpha^k})$-measurable, so $(A_k)_{k\ge 1}$ is mutually independent. Moreover, $S_{\alpha^k} - S_{\alpha^{k-1}}$ is equal in distribution to S_n, i.e. $\mathbb{P}(A_k) = \mathbb{P}(S_n \ge (1-\varepsilon)f(n))$.

By the second Borel–Cantelli lemma, to show that $\mathbb{P}(A_k \text{ i.o.}) = 1$, it suffices to show that $\sum_{k=1}^\infty \mathbb{P}(A_k) = \infty$.

iii. By the lower bound on the tail probability of S_n, up to multiplicative constants,

$$\mathbb{P}(A_k) \ge \frac{n^{1/2}}{(1-\varepsilon)f(n)} \exp\left(-\frac{(1-\varepsilon)^2 f(n)^2}{2n}\right) \sim \frac{c}{\sqrt{\log\log n}} (\log n)^{-(1-\varepsilon)^2}.$$

As $\log n \sim k$ by definition of n, we see that the above is not summable, i.e. $\sum_{k=1}^\infty \mathbb{P}(A_k) = \infty$. We remark that this exponential lower bound is really quite similar in form to the Chernoff upper bound. In both cases, we consider S_n for $n \simeq \alpha^k$, and our bound is $(e^{f(n)^2/(2n)})^{-c} = (e^{\log\log n})^{-c} = (k\log\alpha)^{-c}$. For $c = 1^+$, this term is summable in k, but not for $c = 1^-$. This is one possible reason for the appearance of $f(n)^2/(2n) = \log\log n$: it produces the exact "boundary" k^{-1} under exponentiation, and we then land on different sides of convergence or divergence based on our allowance of $k^{-(1+\varepsilon)^2}$ or $k^{-(1-\varepsilon)^2}$.

iv. We have shown that $S_{\alpha^k} \ge S_{\alpha^{k-1}} + (1-\varepsilon)f(\alpha^k - \alpha^{k-1})$ infinitely often in k a.s. Now, combined with step i,

$$\frac{S_{\alpha^k}}{f(\alpha^k)} \ge \frac{-(1+\varepsilon)f(\alpha^{k-1})}{f(\alpha^k)} + (1-\varepsilon)\frac{f(\alpha^k - \alpha^{k-1})}{f(\alpha^k)} \to -\frac{(1+\varepsilon)}{\alpha^{1/2}} + (1-\varepsilon)\left(1 - \frac{1}{\alpha}\right)^{1/2}.$$

(If we want, we may insert an extra $-\delta$ in the limit to ensure it is an eventual lower bound, then let $\delta \to 0^+$.) As the infinite subsequence $(S_{\alpha^k}/f(\alpha^k))_{k\ge 1}$ is eventually above this lower bound, for the full sequence,

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{f(n)} \ge -\frac{(1+\varepsilon)}{\alpha^{1/2}} + (1-\varepsilon)\left(1 - \frac{1}{\alpha}\right)^{1/2}\right) = 1.$$

Taking $\varepsilon \to 0^+$ and $\alpha \uparrow \infty$, we are done.

□

Interestingly, $f(n) = \sqrt{2n \log\log n}$ is also a "boundary" between almost sure and in probability convergence for the random walk S_n: we have just shown that $S_n/f(n)$ cannot converge to 0 a.s., and yet $S_n/f(n) \to 0$ in probability. (By the CLT and Slutsky's theorem, $n^{1/2}/f(n) \cdot S_n/n^{1/2} \to 0$ in distribution, and convergence in distribution to a *constant* is equivalent to convergence in probability to that constant.)

The proof of the LIL gives a perhaps unsatisfying answer to our question about the origin of the $\log\log n$ term: the iterated logarithm appears out of a rather technical derivation, amidst a combination of exponential bounds, limits, approximations, etc. $\log\log n$ could have been simply conjectured, or possibly found via simulation, although given that its first proof was in the 1920s, this is slightly unlikely.

Or, $f(n) = \sqrt{2n \log\log n}$ could have been derived from some clever and careful technical observations. After all, the first proofs would have been attempted in search of the goalpost, not building a path towards a known destination. We are perhaps lucky that the upper and lower bounds on S_n do coincide at a single "point" $f(n)$; not many bounds will meet this precisely, or even at all. Perhaps it is a bit unintuitive, but it is still a beautiful result.

So, the next time you go on a random walk, be almost sure to look for the hidden roots and iterated logs lining your path. [To the best of my knowledge, this final, beautifully lame line is my own.]

This note made use of the following resources:

1. `Lecture notes` by Ron Peled and scribe Aya Vituri.
2. `Chapter 1 of notes` by Joseph G. Conlon.
3. `Reflection principle for simple random walk` by Math StackExchange user Basj.
4. `Maximal inequality for a sequence of partial sums` by Math StackExchange user saz.
5. `Khinchin's inequality and Etemadi's inequality` by Jordan Bell.
6. `Section 1.9` of *Probability and Measure*, 3rd edition by Patrick Billingsley.
7. `Chapter 8` of *Probability: A Graduate Course* by Allan Gut.
8. *Large-Scale Simulation and Proof for Khinchin's Law of the Iterated Logarithm* by Zhehang Du.

■