Intuition for the law of the iterated logarithm

Alex Fu

2023-02-06

This note is an addendum to *2023-01-19*, in which I naïvely stated there may be no satisfying explanation for the term $\sqrt{2n \log \log n}$ in the law of the iterated logarithm. Of course, it turns out there is one, and I just hadn't looked hard enough. The main ideas in this note were relayed to me by Professor Shirshendu Ganguly; certain additions — elaborations, errors, extended metaphors — are my own.

Like putting cling wrap tightly over a mountain, the asymptotic envelope of the trajectory of a random walk is not its approximate contour, but is "supported" on certain points of exceptional maximality. Along the way, there are certain peaks that stick out more sharply, and valleys that jut out in the negative direction, poking and prodding the envelope, stretching its boundary and shape.

This is ultimately another story of simplification and approximation.

It's not easy to tell where a "representative peak" will rise from. A mountain range is a connected land mass, and the geographical features of a random walk are locally constant. Any given point has some chance of being representative, but once we know a point, neighboring points in its sphere of influence are bound to be near the same elevation, pulled up to its height or down to its depth. This is the constraint of bounded increments on the terrain — there are no cliffs or trenches.

$|S_{n+d} - S_n| \leq d.$

Perhaps, if we look away far enough from one representative point, we'll find another. But it's not enough to look a fixed additive distance away: as n increases, the typical magnitude of S_n is in the order of $\sqrt{n}$. Then, S_{n+d} could very well be part of the same regional feature as S_n, a fluctuation of size d minuscule relative to the size of S_n. At an apex of the towering Great Wall, would you call the top of a nearby staircase another peak?

So, let us consider the next best thing: to look a fixed multiplicative distance away. We may consider S_n and S_{cn} for any $c > 1$, but for simplicity, let us take $c = 2$. Our goal is in approximation, after all.

S_n and S_{2n} are approximately uncorrelated.

Well, $\mathrm{cov}(S_n, S_{2n}) = \mathbb{E}(S_n S_{2n}) = \mathbb{E}(S_n^2) + 0 = n$, and $\sqrt{\mathrm{var}(S_n)\,\mathrm{var}(S_{2n})} = \sqrt{n} \cdot \sqrt{2n} = \sqrt{2}n$, so the correlation coefficient is $1/\sqrt{2} \approx 0.7$, which doesn't seem very close to 0. However, the correlation $1/\sqrt{c}$ can be made as close to 0 as we'd like; we will see that the exact choice of $c = 2$ will not matter much.

Now, applying the normal approximation given by the Central Limit Theorem, we can leverage a key fact: uncorrelated and *jointly Gaussian* random variables are independent. That is, $S_n, S_{n/2}, S_{n/4}, \ldots$ are approximately uncorrelated and thus approximately *independent*. Like individual volcanoes rising from a random sea, unlike parts of the same continuous mountain range, these $\log_2 n$ points are independent candidates for the support of the envelope, or an approximation of the maximal value attained — the representative peaks we are looking for.

(S_n and $S_{n/2}$ are certainly not independent, but they are decorrelated enough. While the first few terms $S_1, S_2, S_4, \ldots$ are very strongly dependent, consecutive terms S_m, S_{2m} are eventually far enough apart thanks to the multiplicative scaling. Considering additive distances would have still left us with $n/d \in \Theta(n)$ many representative peaks, unlike the reduced $\log_2 n$ bits of information we have now.)

And, the distance between S_n and $S_{n/2}$ may certainly seem a bit large — what if peaks arise in between the indices $n/2$ and n, which is certainly possible? The point of approximation with representative peaks is to have a reasonably close lower bound on the probability that the elevation ever exceeds some level. Points between $S_{n/2}$ and S_n are essentially in the "sphere of influence" of $S_{n/2}$ or S_n; the real **key** is that $S_n, S_{n/2}, \ldots$ are *freely* varying random variables, like volcanoes independently rising from a random sea.

Let $k = \log n$, and let $T_k, T_{k-1}, T_{k-2}, \ldots \coloneqq S_n, S_{n/2}, S_{n/4}, \ldots$ be independent. Then

$$\mathbb{P}\left(\max_{1 \le m \le n} S_m \ge t\right) \approx \mathbb{P}\left(\max_{1 \le i \le k} T_i \ge t\right).$$

We can also look at a real mountain range, or rather the clouds above it, imagining $S_n, S_{n/2}, S_{n/4}, \ldots$ as a small flock of clouds grazing on the mountaintops, distant enough that each is free to float as it pleases. We have used the technique of taking an exponential subsequence in two proofs of the SLLN and a proof of the LIL, where it gave a fast subsequence; here, it takes on the role of independent proxies for the supremum.

Now, we can further approximate the T_i as being *identically distributed* as $\mathcal{N}(0, n) = \sqrt{n} \cdot \mathcal{N}(0, 1)$. The variances of $S_n, S_{n/2}, S_{n/4}, \ldots$ are all in the order of n. (Again, the first few terms $S_1, S_2, S_4, \ldots$ are not quite $\mathcal{N}(0, n)$, but eventually the variance is comparable to n by the multiplicative scaling.) Normalizing $Z_i = T_i/\sqrt{n}$, we can treat the "clouds" as $k = \log n$ i.i.d. standard normal random variables.

For a given threshold t, the probability that one of $Z_1, \ldots, Z_k$ i.i.d. $\mathcal{N}(0, 1)$ exceeds t is

$$\mathbb{P}\left(\max_{1 \le i \le k} Z_i \ge t\right) = 1 - (1 - \mathbb{P}(Z_1 \ge t))^k = 1 - (1 - e^{-t^2/2})^k$$

by De Morgan's laws and independence. We want $\mathbb{P}(\max_{1 \le i \le k} Z_i \ge t) \in \Theta(1)$ to be some large probability constant in k. If it tends to 1 quickly as $k \to \infty$, for example when $t = \sqrt{k}$, then t is not a proper envelope: the trajectory will exceed it too easily. If it tends to 0, then the opposite is true: t is not tight enough.

We will use the approximation $1 - (1 - \frac{1}{k})^k \approx 1 - e^{-1} \approx 0.63$. Of course, any other $\frac{c}{k} \in \Theta(\frac{1}{k})$ would be reasonable as well; we take $\frac{1}{k}$ for simplicity. This means that we want

$$e^{-t^2/2} \approx \frac{1}{k} \implies t \approx \sqrt{2 \log k}.$$

Recall that $\log k = \log \log n$ by definition. Lastly, this threshold is scaled by $\sqrt{n}$ to revert the normalization performed on $S_n, S_{n/2}, \ldots$, which gives us the term $\sqrt{2n \log \log n}$. In other words,

$\mathbb{P}\big(\max_{1\le m\le n} S_m \ge \sqrt{2n\log\log n}\big) \approx 1 - e^{-1}$.

Or, more straightforwardly, we treat the k clouds $S_n, S_{n/2}, \ldots$ as i.i.d. $\mathcal{N}(0,n)$ random variables to find the threshold

$$\mathbb{P}\left(\max_{1\le i\le k} T_i \ge t\right) = 1 - (1 - e^{-t^2/(2n)})^k \approx 1 - e^{-1} \implies t \approx \sqrt{2n\log k} = \sqrt{2n\log\log n}.$$

The remaining probability of e^{-1} perhaps accounts for the lands in between the representative peaks. This probability is an approximate lower bound, after all. In any case, we have a "tight" asymptotic estimate of $\sup_{1\le m\le n} S_m$: it lies above and below $\sqrt{2n\log\log n}$ *both* with high probability, which indicates that $\limsup S_n \sim \sqrt{2n\log\log n}$ perhaps almost surely. Indeed, previous arguments show that this threshold is the right envelope to consider.

Finally, we also have a more direct explanation for the term $\sqrt{2n\log\log n}$. By the CLT, the tail probability of S_n is asymptotically $e^{-t^2/(2n)}$, whose inverse is $\sqrt{2n\log u^{-1}}$. A factor of $\log n$ appears in u^{-1} from the approximation of $\sup_{1\le m\le n} S_n$ using the semi-independent exponential subsequence $S_n, S_{n/2}, S_{n/4}, \ldots$, which has $\log n$ terms.

So, the next time you try to cling wrap a mountain range, look to the clouds for guidance.

■