Another proof of the strong law of large numbers

Alex Fu

2023-01-18

In this note, we will continue on our quest to further quantify and make precise the strong law of large numbers from *2023-01-02*. Recall the strategy of our previous proof:

- We reduced to the case of X_n nonnegative without loss of generality.
- Then, we defined the truncated sequence $\bar{X}_i = X_i \cdot \mathbb{1}_{X_i \le i}$. The only property we were given is the finiteness condition $\mathbb{E}(X_1) < \infty$, so we used the Borel–Cantelli lemma and the tail-sum approximation to show that the limit of the sample mean $\bar{S}_n/n$ is unaffected by truncation almost surely. Here is finiteness in action:

$$\sum_{i=1}^{\infty} \mathbb{P}(X_i \neq \bar{X}_i) = \sum_{i=1}^{\infty} \mathbb{P}(X_i > i) \approx \mathbb{E}(X_1) < \infty.$$

- With bounded random variables, we now gain access to higher-order Markov's inequalities, such as Chebyshev's inequality. By also choosing a sufficiently fast subsequence $(\bar{X}_{k_n})_{n\ge1}$, we get even more control over the rate of decay of the Chebyshev bound. Now, invoking the Borel–Cantelli lemma, Chebyshev's inequality, and the tail-sum approximation once more, we convert the question of almost sure convergence into one of finiteness:

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{\bar{S}_{k_n} - \mathbb{E}(\bar{S}_{k_n})}{k_n}\right| > \varepsilon\right) \lesssim \sum_{n=1}^{\infty} \left[\frac{1}{k_n^2} \sum_{i=1}^{k_n} \mathbb{E}(X_{k_n}^2)\right] \approx \sum_{j=1}^{\infty} \left[j\,\mathbb{P}(X_1 > j) \sum_{n:k_n \ge j} \frac{1}{k_n}\right] \approx \mathbb{E}(X_1).$$

- Finally, we simply leverage monotonicity to upgrade to convergence along the full sequence.

Almost sure convergence is rather unique in that it is really *pointwise* convergence in its essence: as such, we can take inspiration from and make use of quantitative results on the convergence of sequences of real numbers for the purposes of almost sure convergence.

For example, while the harmonic series ($p = 1$) is the "boundary" between convergence and divergence for p-series $\sum_{n=1}^{\infty} n^{-p}$, it is not the most precise boundary. A finer result states that

$$\sum_{n=1}^{\infty} \frac{1}{n(\log n)^c}$$

converges for $c > 1$ and diverges for $c \le 1$, and thus $\sum_{n=1}^{\infty} (n \log n)^{-1} = \infty$ is a better boundary.

Now, we know that $S_n/n \to 0$ almost surely, and $S_n/\sqrt{n}$ tends to an $O(1)$ tight random variable. We wish to go further: indeed, we can show that $S_n/n^{0.5+\varepsilon} \to 0$ a.s. for $\varepsilon > 0$. First, we will need a few lemmas.

Lemma 1 (Kolmogorov's maximal inequalilty)**.**

Let $X_1, X_2, \ldots$ be independent and zero-mean. Then

$$\mathbb{P}\left(\sup_{1 \leq k \leq n} |S_k| > x\right) \leq \frac{\mathbb{E}(S_n^2)}{x^2}.$$

This is a stronger result than Chebyshev's inequality, which only bounds the probability of deviation for the "endpoint" S_n/n, not the maximum of the "process" S_k/k over all $1 \leq k \leq n$. As a side note, Lemma 2 is a special case of the *optional stopping theorem* for submartingales.

Proof. Let A^* be the event of interest. A useful strategy also used in the proof of *Ottaviani's inequality*: we observe that if $\sup_{1 \leq k \leq n} |S_k| > x$, then there must be an index k at which $|S_k|$ **first** exceeds x. Thus, let us partition A^* into the events $A_k = \{|S_k| > x \text{ and } |S_j| \leq x \text{ for all } j < k\}$. Now,

$$\begin{aligned}\mathbb{E}(S_n^2) \geq \mathbb{E}(S_n^2 \cdot \mathbb{1}_{A^*}) = \sum_{k=1}^{n} \mathbb{E}(S_n^2 \cdot \mathbb{1}_{A_k})& \\ &= \mathbb{E}(S_k^2 \cdot \mathbb{1}_{A_k}) + 2\,\mathbb{E}(S_k(S_n - S_k) \cdot \mathbb{1}_{A_k}) + \mathbb{E}((S_n - S_k)^2 \cdot \mathbb{1}_{A_k}).\end{aligned}$$

Note that A_k is $\sigma(X_1, \ldots, X_k)$-measurable, as is S_k, and $\sigma(X_1, \ldots, X_k)$ is independent of $\sigma(X_{k+1}, \ldots, X_n)$. Thus the cross term factorizes as $\mathbb{E}(S_k \mathbb{1}_{A_k}) \cdot \mathbb{E}(S_n - S_k)$ and vanishes by zero-meanness. As $(S_n - S_k)^2 \geq 0$, and $S_k^2 > x^2$ on the event A_k by its definition,

$$\begin{aligned}&\geq \mathbb{E}(S_k^2 \cdot \mathbb{1}_{A_k}) \\ &\geq x^2\,\mathbb{P}(A_k).\end{aligned}$$

□

Lemma 2 (Summability of variances implies almost-sure summability of sequence)**.**

Let $X_1, X_2, \ldots$ be independent and zero-mean. If $\sum_{i=1}^{\infty} \mathsf{var}(X_i) < \infty$, then $\sum_{i=1}^{\infty} X_i < \infty$ almost surely.

Proof. When we want to prove almost sure convergence without necessarily knowing what the limit explicitly is, a common strategy is to leverage the completeness of the reals, i.e. show that the sequence is almost surely Cauchy. (The same strategy works for L^p limits, as L^p space is complete.) Let $\varepsilon > 0$. We want N such that

$$|S_n - S_N| < \varepsilon \text{ almost surely for all } n \geq N.$$

Writing the Cauchyness condition in terms of only one variable allows us to invoke the Chebyshev-like Kolmogorov's maximal inequality and bound the probability of the complement:

$$\mathbb{P}\left(\sup_{N \leq m \leq n} |S_n - S_m| \geq \varepsilon\right) \leq \sum_{i=N}^{\infty} \frac{\mathsf{var}(X_i)}{\varepsilon^2}.$$

The event on the left-hand side increases as $n \to \infty$, but the upper bound, which does not depend on n, still holds. Moreover, as $N \to \infty$, the right-hand side tends to 0 as $1/\varepsilon^2$ times the tail of a convergent series, while the left-hand side tends to the probability that S_n is not a.s. Cauchy. Thus S_n converges almost surely, and we are done. □

Alternatively, one can be a bit more rigorous in using the monotone continuity of probability measures and formulating the event of being almost sure Cauchy in terms of countable unions and intersections, but this is besides the point.

Lemma 3 (Kronecker's lemma)**.**

Let $(a_n)_{n\geq 1}$ be a sequence of real numbers, and suppose $b_n \uparrow \infty$. If $\sum_i (a_i/b_i) < \infty$, then $(\sum_{k=1}^n a_k)/b_n \to 0$.

This is a statement about real numbers, and its proof requires no probability whatsoever.

Proof. Let $\gamma_k \coloneqq \sum_{i=1}^k (a_i/b_i)$, and let $\gamma \coloneqq \lim_{k\to\infty} \gamma_k < \infty$. We observe that the difference of consecutive values $\gamma_i - \gamma_{i-1}$ isolates the term a_i/b_i. Let us write

$$\sum_{k=1}^n a_k = \sum_{k=1}^n (\gamma_k - \gamma_{k-1}) b_k.$$

Using *summation by parts*, the discrete analogue of integration by parts, or *Abel's lemma*, the above is equal to

$$b_n \gamma_n - \sum_{k=1}^{n-1} (b_{k+1} - b_k)\gamma_k,$$

where $\gamma_0 = 0$. Alternatively, simply expand the sum and group by γ_k instead of b_k. Now,

$$\frac{1}{b_n} \sum_{k=1}^n a_k = \gamma_n - \sum_{k=1}^{n-1} \frac{(b_{k+1} - b_k)}{b_n} \gamma_k.$$

For $\varepsilon > 0$, take any N such that $k \geq N$ implies $|\gamma_k - \gamma| < \varepsilon$. Then for any $n \geq N$, the above is sandwiched by

$$\gamma_n - \sum_{k=N}^{n-1} \frac{b_{k+1} - b_k}{b_n} (\gamma + \varepsilon) \leq \cdots \leq \gamma_n - \sum_{k=N}^{n-1} \frac{b_{k+1} - b_k}{b_n} (\gamma - \varepsilon).$$

Note that the first finitely many terms in the sum do not matter in the limit: $\frac{1}{b_n} \sum_{k=1}^{N-1} (b_{k+1} - b_k)\gamma_k \to 0$ because $b_n \uparrow \infty$ and N is fixed. From here, by telescoping cancellation, the bounds become

$$\gamma_n - \frac{b_n - b_N}{b_n} (\gamma + \varepsilon) \leq \cdots \leq \gamma_n - \frac{b_n - b_N}{b_n} (\gamma - \varepsilon).$$

Finally, passing to the limit as $n \to \infty$, the above simplifies to

$$-\varepsilon \leq \lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^n a_k \leq \varepsilon.$$

As $\varepsilon > 0$ is arbitrary, we have shown that $\lim_{n\to\infty} (\sum_{k=1}^n a_k)/b_n = 0$, as desired. $\square$

And, finally, the main event.

Theorem 1 (Strong law of large numbers)**.**

Let $X_1, X_2, \ldots$ be independent and identically distributed zero-mean random variables, with $\mathbb{E}(|X_1|) < \infty$. Then $S_n/n \to 0$ almost surely.

Proof. Even with a new strategy, do not forget our old friends, the Borel–Cantelli lemma and the tail-sum approximation. The theme of quantification still runs through us.

i. By Lemma 3, letting $a_n = X_n(\omega)$ and $b_n = n \uparrow \infty$, it suffices to show that $\sum_{k=1}^{\infty} X_k/k < \infty$ almost surely.

ii. By Lemma 2, it suffices to show that $\sum_{k=1}^{\infty} \mathsf{var}(X_k)/k^2 < \infty$. However, the X_k do not necessarily have finite second moment. As such, to regain higher-order moments, we turn to the usual method of *truncation*.

iii. Let $\bar{X}_k = X_k \cdot \mathbb{1}_{|X_k| \le k}$, and let $Y_k = \bar{X}_k - \mathbb{E}(\bar{X}_k)$. By the same argument as before,

$$\sum_{k=1}^{\infty} \frac{\mathsf{var}(\bar{X}_k)}{k^2} < \infty \xrightarrow{\text{Lemma 2}} \sum_{k=1}^{\infty} \frac{Y_k}{k} < \infty \text{ a.s.} \xrightarrow{\text{Lemma 3}} \sum_{k=1}^{n} \frac{\bar{X}_k}{n} \to 0 \text{ a.s.}$$

Now, a familiar argument proves that truncation has no effect in the limit: $\bar{S}_n/n = S_n/n$ as $n \to \infty$ almost surely because $X_k \neq \bar{X}_k$ only finitely often. Simply combine Borel–Cantelli and

$$\sum_{k=1}^{\infty} \mathbb{P}(X_k \neq \bar{X}_k) = \sum_{k=1}^{\infty} \mathbb{P}(X_k > k) \approx \mathbb{E}(X_1) < \infty.$$

We also needed the observation that $\mathbb{E}(\bar{X}_k) \to 0$ by dominated convergence, taking $\bar{X}_k \overset{\mathsf{d}}{=} X_1 \mathbb{1}_{|X_1| \le k} \le X_1$. This shows that the running average $\sum_{k=1}^{n} \mathbb{E}(\bar{X}_k)/n \to 0$, which we combine with the almost sure convergence of $\sum_{k=1}^{n} Y_k/n \to 0$ to get the conclusion of $\bar{S}_n/n \to 0$ above.

iv. It suffices to show that $\sum_{k=1}^{\infty} \mathsf{var}(\bar{X}_k)/k^2 < \infty$. Using Tonelli's theorem to exchange the summations, this is upper bounded by

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}(|X_k|^2 \cdot \mathbb{1}_{|X_k| \le k})}{k^2} \approx \sum_{k=1}^{\infty} \left[\frac{1}{k^2} \sum_{j=1}^{k} j\, \mathbb{P}(|X_1| > j) \right] \approx \sum_{j=1}^{\infty} \left[j\, \mathbb{P}(|X_1| > j) \sum_{k: k > j} \frac{1}{k^2} \right].$$

Note that $\sum_{k:k>j} 1/k^2 \approx 1/j$ by Riemann approximation, so this bound is

$$\approx \sum_{j=1}^{\infty} \mathbb{P}(|X_1| > j) \approx \mathbb{E}(|X_1|) < \infty.$$

□

As we see, stronger quantitative control is given by the choice of $b_n \uparrow \infty$. How far can we push this technique? Per our remarks in the beginning, $\sum_{i=1}^{\infty} i^{-1}(\log i)^{-(1+\varepsilon)} < \infty$ for $\varepsilon > 0$. If $\mathbb{E}(X_1^2) < \infty$, then letting $b_n = k^{1/2}(\log k)^{1/2+\varepsilon}$, we see that the finiteness of

$$\sum_{k=1}^{\infty} \frac{\mathsf{var}(X_k)}{(k^{1/2}(\log k)^{1/2+\varepsilon})^2} = \mathbb{E}(X_1^2) \left(\sum_{k=2}^{\infty} \frac{1}{k(\log k)^{1+2\varepsilon}} \right) < \infty$$

shows that $S_n/b_n \to 0$ almost surely by the same argument as above. $S_n/\sqrt{n(\log n)^c} \to 0$ for $c > 1$ is the limit of what we can reach for now. A further destination, for another day, is the *law of the iterated logarithm*.

This note was largely adapted from Professor Shirshendu Ganguly's fall 2022 offering of Math C218A / Stat C205A at UC Berkeley.

■