A proof of the weak law of large numbers

Alex Fu

2023-01-20

Theorem 1 (Weak law of large numbers, or WLLN)**.**

Let $X_1, X_2, \ldots$ be i.i.d. zero-mean random variables with $\mathbb{E}(|X_1|) < \infty$. Then $S_n/n \to 0$ in probability.

This note continues an unintentional mini-series on limit theorems in probability: *2023-01-02*, *01-03*, *01-18*, *01-19*. This note is largely adapted from Professor Shirshendu Ganguly's fall 2022 offering of Math C218A / Stat C205A.

The weak law of large numbers may seem like an unremarkable consequence of the stronger law and the fact that almost sure convergence implies convergence in probability. However, there are a few reasons why this weaker law is deserving of its own name, even beyond its historical precedence. The weaker conclusion holds in more situations than the almost sure conclusion, as we saw in *2023-01-19*, or with the *Cauchy distribution*, which lacks a first moment.

Note that the weak law of large numbers is usually presented before the strong law because its proof is easier. Many of our remarks and arguments in *2023-01-02* still apply: For instance, we assume the X_i are zero-mean without loss of generality; a LLN is a statement about cancellation, which the bound $\mathbb{E}(|S_n|) \le n\,\mathbb{E}(|X_1|)$ fails to capture; and Chebyshev's inequality applies to bounded random variables.

We start with some formal checks of integrability.

Lemma 1 (Expectation of S_n^2)**.**

Let $X_1, X_2, \ldots$ be i.i.d. zero-mean random variables with $\mathbb{E}(X_1^2) < \infty$. Then $\mathbb{E}(S_n^2) = n\,\mathbb{E}(X_1^2)$.

Proof. Expanding out S_n^2, by the linearity of expectation,

$$\mathbb{E}(S_n^2) = \sum_{i=1}^{n} \mathbb{E}(X_i^2) + \sum_{i \neq j} \mathbb{E}(X_i X_j) = n\,\mathbb{E}(X_1^2) + \binom{n}{2}\,\mathbb{E}(X_1 X_2)$$

Note that $\mathbb{E}(X_1^2) < \infty$ by hypothesis, and $\mathbb{E}(|X_i X_j|) \le \frac{1}{2}\,\mathbb{E}(X_i^2 + X_j^2) < \infty$ by monotonicity. Moreover, $\mathbb{E}(X_i X_j)$ factorizes as $\mathbb{E}(X_i)\,\mathbb{E}(X_j) = 0$ by independence (or Fubini's theorem), and we are done. $\square$

Now, Lemma 2 is a simple application of Chebyshev's inequality.

Lemma 2 (WLLN with second moment assumption)**.**

Let $X_1, X_2, \ldots$ be i.i.d. zero-mean random variables with $\mathbb{E}(X_1^2) < \infty$. Then $S_n/n \to 0$ in probability.

Proof. Let $\varepsilon > 0$. Then by Chebyshev's inequality,

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{\mathbb{E}(S_n^2)}{n^2\varepsilon^2} = \frac{\mathbb{E}(X_1^2)}{n\varepsilon^2} \to 0.$$

□

To remove the second moment assumption, we introduce a more general setting: *triangular arrays*.

$$\begin{array}{lll} X_{1,1} & & \\ X_{2,1} & X_{2,2} & \\ X_{3,1} & X_{3,2} & X_{3,3} \\ \vdots & & \end{array}$$

where $(X_{n,i})_{i=1}^n$ is independent along each row, and we don't care about how the rows relate. A simple example is given by the partial sequences $(X_i)_{i=1}^n$ of an i.i.d. sequence $X_1, X_2, \ldots$, but triangular arrays provide a more general method to deal with a "sequence of sequences," where the $X_{n,i}$ depend on n as well as i, similar in spirit to a Cantor diagonalization argument. We will encounter triangular arrays again in the *Lindeberg–Feller Central Limit Theorem*.

Lemma 3 (Weak law for triangular arrays)**.**

Let $(X_{n,i})_{i=1}^n$ be a triangular array, let $(b_n)_{n=1}^\infty$ be a sequence of real numbers such that $b_n \uparrow \infty$, and define the truncated random variables $\bar{X}_{n,i} \coloneqq X_{n,i} \cdot \mathbb{1}_{|X_{n,i}| \leq b_n}$.

1. $\sum_{i=1}^n \mathbb{P}(|X_{n,i}| > b_n) \to 0$.
2. $\sum_{i=1}^n \mathbb{E}(\bar{X}_{n,i}^2)/b_n^2 \to 0$.

If conditions 1 and 2 above hold, then

$$\frac{1}{b_n}\sum_{i=1}^n (X_{n,i} - \mathbb{E}(\bar{X}_{n,i})) \to 0 \text{ in probability.}$$

Proof. We write $S_n = \sum_{i=1}^n X_{n,i}$ and $\bar{S}_n = \sum_{i=1}^n \bar{X}_{n,i}$.

i. By Chebyshev's inequality for the truncated sum, which has finite second moment by boundedness,

$$\mathbb{P}\left(\frac{|\bar{S}_n - \mathbb{E}(\bar{S}_n)|}{b_n} > \varepsilon\right) \leq \frac{\mathrm{var}(\bar{S}_n)}{b_n^2\varepsilon^2} \leq \frac{1}{\varepsilon^2}\sum_{i=1}^n \frac{\mathbb{E}(\bar{X}_{n,i}^2)}{b_n^2}.$$

By condition 2, this tends to 0, which shows that $(\bar{S}_n - \mathbb{E}(\bar{S}_n))/b_n \to 0$ in probability.

ii. Now, we use a common union bound argument to show that the truncated sum does not differ much from the original sum. By condition 1,

$$\mathbb{P}(S_n \neq \bar{S}_n) \leq \sum_{i=1}^n \mathbb{P}(X_{n,i} \neq \bar{X}_{n,i}) = \sum_{i=1}^n \mathbb{P}(|X_{n,i}| > b_n) \to 0.$$

iii. Finally, combining $(\bar{S}_n - \mathbb{E}(\bar{S}_n))/b_n \to 0$ and $(S_n - \bar{S}_n) \to 0$, we see that $(S_n - \mathbb{E}(\bar{S}_n))/b_n \to 0$ in probability as well. A triangle inequality and union bound argument proves that $X_n + Y_n \to X + Y$ for $X_n \to X$, $Y_n \to Y$ in probability in general:

$$\begin{aligned} \mathbb{P}(|X_n + Y_n - (X+Y)| > \varepsilon) &\leq \mathbb{P}(|X_n - X| + |Y_n - Y| > \varepsilon) \\ &\leq \mathbb{P}\Big(|X_n - X| > \frac{\varepsilon}{2}\Big) + \mathbb{P}\Big(|Y_n - Y| > \frac{\varepsilon}{2}\Big) \\ &\to 0. \end{aligned}$$

□

Proof of Theorem 1. The crux of the argument is the weak law for triangular arrays.

i. Set $X_{n,i} = X_i$ and $b_n = n$. By the i.i.d. condition,

$$\sum_{i=1}^{n} \mathbb{P}(|X_{n,i}| > b_n) = n\,\mathbb{P}(|X_1| > n).$$

This is bounded by $\mathbb{E}(|X_1|)$ by Markov's inequality, but we want this to tend to 0 for condition 1 to hold. By dominated convergence, where $\mathbb{E}(|X_1|) < \infty$, we have that

$$n\,\mathbb{P}(|X_1| > n) = \mathbb{E}(n \cdot \mathbb{1}_{|X_1|>n}) \leq \mathbb{E}(|X_1| \cdot \mathbb{1}_{|X_1|>n}) \to 0.$$

ii. Now, by the i.i.d. condition and the tail-sum approximation, condition 2 becomes

$$\frac{1}{n}\,\mathbb{E}(\bar{X}_1^2) \approx \frac{1}{n}\sum_{i=1}^{n} 2i\,\mathbb{P}(|X_1| \geq i) \to 0,$$

where $\bar{X}_1 = X_1 \cdot \mathbb{1}_{|X_1|\leq n}$ implicitly depends on n. By the previous step, $i\,\mathbb{P}(|X_1| \geq i) \to 0$ implies that the running average $\frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(|X_i| \geq i)$ tends to the same limit of 0.

iii. Invoking the weak law for triangular arrays, $(S_n/n) - \mathbb{E}(\bar{X}_1) \to 0$ in probability. Now, it suffices to show that $\mathbb{E}(X_1) - \mathbb{E}(\bar{X}_1) \to 0$ in probability as $n \to \infty$. We write

$$\mathbb{E}(X_1) - \mathbb{E}(\bar{X}_1) = \mathbb{E}(X_1 \cdot \mathbb{1}_{|X_1|>n}),$$

which tends to 0 by dominated convergence. A convergent sequence of real numbers converges in probability, so we are done.

□

■