

Lecture Notes for ECE 534, Random Processes, Fall 2026

Laws of Large Numbers

These notes treat the laws of large numbers for sums of independent, identically distributed (iid) random variables. We first prove the weak and the strong law under a finite-variance assumption, using nothing beyond Markov's inequality and the Borel–Cantelli lemma: the strong law is obtained along the subsequence of perfect squares, extended to all n for nonnegative variables by a sandwich argument, and then to general variables by splitting into positive and negative parts. We then show that the same second-moment computation gives a law of large numbers in mean square for variables that are merely uncorrelated, not necessarily independent. We then outline Kolmogorov's proof of the strong law under the minimal assumption of a finite mean, work through three examples (Gaussian samples, the empirical CDF, and the Cauchy counterexample), and close with a statement of the central limit theorem.

Contents

1	The weak law of large numbers	2
2	The strong law of large numbers under finite variance	2
2.1	Step 1: convergence along the squares	3
2.2	Step 2: nonnegative variables and the sandwich bound	3
2.3	Step 3: general variables	4
3	The mean square law for uncorrelated variables	4
4	Kolmogorov's strong law under a finite mean	6
4.1	Three ingredients	6
4.2	Outline of the proof of Theorem 4.1(a)	7
4.3	Proof of Theorem 4.1(b)	8
5	Examples	9
5.1	IID standard Gaussian random variables	9
5.2	The empirical CDF converges to the true CDF	10
5.3	A counterexample: the Cauchy distribution	11
6	The central limit theorem	12

Notation and conventions

Throughout, $X_1, X_2, \dots$ are random variables on a common probability space, $S_n = X_1 + \dots + X_n$, and S_n/n is the sample mean of the first n observations. When the X_i have a common finite mean we write $\mu = \mathbb{E}[X_1]$, and when they have a common finite variance we write $\sigma^2 = \text{Var}(X_1)$. Convergence almost surely, in probability, in mean square and in distribution is denoted by $\xrightarrow{\text{a.s.}}$, $\xrightarrow{\text{p}}$, $\xrightarrow{\text{m.s.}}$, and $\xrightarrow{\text{d}}$ respectively.

$\xrightarrow{p.}$, $\xrightarrow{m.s.}$ and $\xrightarrow{d.}$. A *weak* law of large numbers asserts $S_n/n \xrightarrow{p.} \mu$, and a *strong* law asserts $S_n/n \xrightarrow{a.s.} \mu$.

Markov's and Chebyshev's inequalities and the two Borel–Cantelli lemmas are used freely. The first Borel–Cantelli lemma will always enter through the following standard consequence:

$$\sum_{n=1}^{\infty} \mathbb{P}(|Y_n - Y| > \varepsilon) < \infty \text{ for every } \varepsilon > 0 \implies Y_n \xrightarrow{a.s.} Y. \quad (1)$$

1 The weak law of large numbers

Theorem 1.1 (Weak law, finite variance). *Let $X_1, X_2, \dots$ be iid with mean μ and variance $\sigma^2 < \infty$. Then for every $\varepsilon > 0$ and every $n \geq 1$,*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}. \quad (2)$$

Consequently $S_n/n \xrightarrow{p.} \mu$, and also $S_n/n \xrightarrow{m.s.} \mu$.

Proof. By linearity, $\mathbb{E}[S_n/n] = \mu$. Independent random variables are uncorrelated, so the variance of the sum is the sum of the variances:

$$\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = \text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

The right-hand side tends to zero, which is mean square convergence. Applying Markov's inequality to the nonnegative random variable $(S_n/n - \mu)^2$ with $c = \varepsilon^2$ gives (2). $\square$

Remark 1.2 (What the proof used). Only that the X_i are pairwise uncorrelated with a common mean and a common (or merely uniformly bounded) variance. Neither full independence nor identical distributions were needed. Section 3 takes this up.

Remark 1.3 (Sample size). To guarantee $\mathbb{P}(|S_n/n - \mu| \geq \varepsilon) \leq \delta$, the bound (2) asks for $n \geq \sigma^2/(\delta\varepsilon^2)$ samples. For $\sigma = 1$, $\varepsilon = 0.1$ and $\delta = 0.05$ this is $n \geq 2000$. The bound holds for *every* distribution with variance σ^2 , and is correspondingly conservative; see Example 5.1 and Remark 6.4.

Remark 1.4 (Why this does not give the strong law). The bound $\sigma^2/(n\varepsilon^2)$ tends to zero but $\sum_n 1/n = \infty$, so the Borel–Cantelli criterion (1) cannot be applied directly. The next section gets around this by first passing to a subsequence along which the bound *is* summable.

2 The strong law of large numbers under finite variance

Theorem 2.1 (Strong law, finite variance). *Let $X_1, X_2, \dots$ be iid with mean μ and variance $\sigma^2 < \infty$. Then*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu.$$

The proof has three steps.

Step 1. Convergence along the perfect squares $n = k^2$, by Chebyshev and Borel–Cantelli.

Step 2. Extension to all n when the variables are nonnegative, by monotonicity of S_n and a sandwich bound.

Step 3. Extension to general variables via $X = X^+ - X^-$.

2.1 Step 1: convergence along the squares

Lemma 2.2. *Under the assumptions of Theorem 2.1,*

$$\frac{S_{k^2}}{k^2} \xrightarrow{a.s.} \mu \quad \text{as } k \rightarrow \infty.$$

Proof. Fix $\varepsilon > 0$. By (2) with $n = k^2$,

$$\sum_{k=1}^{\infty} \mathbb{P}\left(\left|\frac{S_{k^2}}{k^2} - \mu\right| > \varepsilon\right) \leq \sum_{k=1}^{\infty} \frac{\sigma^2}{k^2 \varepsilon^2} = \frac{\pi^2 \sigma^2}{6 \varepsilon^2} < \infty.$$

The Borel–Cantelli criterion (1), applied to the sequence $Y_k = S_{k^2}/k^2$, gives the claim. $\square$

2.2 Step 2: nonnegative variables and the sandwich bound

Lemma 2.2 controls the sample mean only at the times $1, 4, 9, 16, \dots$. Between consecutive squares the sample mean could, a priori, wander. For nonnegative summands it cannot wander far, because the partial sums are then nondecreasing in n .

Lemma 2.3. *In addition to the assumptions of Theorem 2.1, suppose $X_i \geq 0$ for all i . Then $S_n/n \xrightarrow{a.s.} \mu$.*

Proof. For $n \geq 1$ let $k = k(n) = \lfloor \sqrt{n} \rfloor$, so that

$$k^2 \leq n < (k+1)^2.$$

Since the X_i are nonnegative, S_n is nondecreasing in n , whence $S_{k^2} \leq S_n \leq S_{(k+1)^2}$. Dividing by n and using $k^2 \leq n < (k+1)^2$ once more (together with $S_{k^2} \geq 0$ and $S_{(k+1)^2} \geq 0$),

$$\frac{S_{k^2}}{(k+1)^2} \leq \frac{S_n}{n} \leq \frac{S_{(k+1)^2}}{k^2}.$$

Rewriting each outer term as a ratio of squares times a sample mean taken *at a square* gives the sandwich bound

$$\left(\frac{k}{k+1}\right)^2 \frac{S_{k^2}}{k^2} \leq \frac{S_n}{n} \leq \left(\frac{k+1}{k}\right)^2 \frac{S_{(k+1)^2}}{(k+1)^2}, \quad k = \lfloor \sqrt{n} \rfloor. \quad (3)$$

Let Ω_0 be the event, of probability one by Lemma 2.2, on which $S_{k^2}/k^2 \rightarrow \mu$. Fix $\omega \in \Omega_0$ and let $n \rightarrow \infty$. Then $k(n) \rightarrow \infty$, the two prefactors in (3) tend to 1, and both S_{k^2}/k^2 and $S_{(k+1)^2}/(k+1)^2$ tend to μ . Both ends of (3) therefore converge to μ , and so does S_n/n . $\square$

Remark 2.4 (Why squares). The subsequence $n_k = k^2$ meets two competing requirements. It is *sparse* enough that the Chebyshev bounds are summable, $\sum_k 1/n_k < \infty$, and *dense* enough that consecutive terms are asymptotically indistinguishable, $n_{k+1}/n_k \rightarrow 1$, so that the prefactors in (3) tend to 1. The subsequence $n_k = k$ fails the first requirement and $n_k = 2^k$ fails the second. (With $n_k = 2^k$ the sandwich would only give $\frac{1}{2}\mu \leq \liminf S_n/n \leq \limsup S_n/n \leq 2\mu$.)

2.3 Step 3: general variables

For a real number x write $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$, so that

$$x = x^+ - x^-, \quad |x| = x^+ + x^-, \quad x^+ \geq 0, \quad x^- \geq 0.$$

Proof of Theorem 2.1. The random variables $X_1^+, X_2^+, \dots$ are iid, being a fixed function of the iid sequence (X_i) ; they are nonnegative; and they have finite variance because $\mathbb{E}[(X_1^+)^2] \leq \mathbb{E}[X_1^2] < \infty$. Lemma 2.3 therefore applies to them, and likewise to $X_1^-, X_2^-, \dots$:

$$\frac{1}{n} \sum_{i=1}^n X_i^+ \xrightarrow{\text{a.s.}} \mathbb{E}[X_1^+], \quad \frac{1}{n} \sum_{i=1}^n X_i^- \xrightarrow{\text{a.s.}} \mathbb{E}[X_1^-].$$

On the intersection of the two probability-one events on which these limits hold,

$$\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i^+ - \frac{1}{n} \sum_{i=1}^n X_i^- \longrightarrow \mathbb{E}[X_1^+] - \mathbb{E}[X_1^-] = \mathbb{E}[X_1] = \mu. \quad \square$$

Remark 2.5 (What the proof used). Step 1 used only Chebyshev's inequality, hence only that the X_i are pairwise uncorrelated with common mean and variance. Step 3 needs the same for (X_i^+) and for (X_i^-) , which is guaranteed if the X_i are *pairwise independent* and identically distributed. So Theorem 2.1 holds verbatim under pairwise independence.

Remark 2.6 (A second route past the squares). One can avoid the positive/negative decomposition by bounding the fluctuation of S_n between consecutive squares directly. Assume $\mu = 0$ without loss of generality and set $D_k = \max_{k^2 \leq n < (k+1)^2} |S_n - S_{k^2}|$. Bounding the maximum of nonnegative numbers by their sum,

$$\mathbb{E}[D_k^2] \leq \sum_{n=k^2+1}^{(k+1)^2-1} \mathbb{E}[(S_n - S_{k^2})^2] = \sigma^2 \sum_{i=1}^{2k} i = \sigma^2 k(2k+1) \leq 3\sigma^2 k^2,$$

so $\mathbb{P}(D_k > \varepsilon k^2) \leq 3\sigma^2/(\varepsilon^2 k^2)$ by Markov's inequality. This is summable in k , hence $D_k/k^2 \xrightarrow{\text{a.s.}} 0$ by (1), and for $k^2 \leq n < (k+1)^2$,

$$\left| \frac{S_n}{n} \right| \leq \frac{|S_{k^2}| + D_k}{k^2} \xrightarrow{\text{a.s.}} 0.$$

The cruder union-type bound used here is replaced, in Kolmogorov's proof below, by a sharp *maximal inequality*.

3 The mean square law for uncorrelated variables

Independence entered the proofs of Theorem 1.1 and Lemma 2.2 at exactly one point: the variance of a sum is the sum of the variances. That identity needs only that the summands be uncorrelated. This section records what the same computation gives when independence and identical distributions are both dropped. Mean square convergence is the natural mode here, since it is a statement about second moments only.

Theorem 3.1 (Law of large numbers in mean square). *Let $X_1, X_2, \dots$ be random variables with finite variances such that*

- (i) $\mathbb{E}[X_i] = \mu$ for all i ;
- (ii) $\text{Cov}(X_i, X_j) = 0$ for all $i \neq j$;
- (iii) $\text{Var}(X_i) \leq c$ for all i , for some constant $c < \infty$.

Then for every $n \geq 1$,

$$\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] \leq \frac{c}{n}, \quad (4)$$

so $S_n/n \xrightarrow{\text{m.s.}} \mu$, and consequently $S_n/n \xrightarrow{p} \mu$. The X_i need be neither independent nor identically distributed.

Proof sketch. By (i), $\mathbb{E}[S_n/n] = \mu$, so the mean square error of S_n/n is its variance. Expanding the variance of the sum into its n^2 covariances,

$$\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = \frac{1}{n^2} \text{Var}(S_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \leq \frac{c}{n}.$$

The third equality is (ii): the $n^2 - n$ off-diagonal terms vanish. Only the n diagonal terms survive, each at most c by (iii), against a normalization of n^2 . For convergence in probability, apply Markov's inequality to $(S_n/n - \mu)^2$: $\mathbb{P}(|S_n/n - \mu| \geq \varepsilon) \leq c/(n\varepsilon^2)$. $\square$

Remark 3.2 (Geometric picture). Regard random variables with finite second moments as vectors with inner product $\langle X, Y \rangle = \mathbb{E}[XY]$ and norm $\|X\| = \sqrt{\mathbb{E}[X^2]}$. The centered variables $X_i - \mu$ are then *orthogonal* by (ii), and the proof is the Pythagorean theorem:

$$\|S_n - n\mu\|^2 = \sum_{i=1}^n \|X_i - \mu\|^2 \leq nc.$$

A sum of n orthogonal vectors of bounded length has length of order $\sqrt{n}$, not n , and dividing by n sends it to zero. At the other extreme, if $X_i = X_1$ for all i the vectors are parallel, $\|S_n - n\mu\| = n\|X_1 - \mu\|$, and $S_n/n = X_1$ does not converge to μ .

Remark 3.3 (What can be relaxed). The proof rests on the exact identity

$$\mathbb{E}\left[\left(\frac{S_n - \mathbb{E}[S_n]}{n}\right)^2\right] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j),$$

so each hypothesis can be weakened as far as this double sum allows.

- (a) *Unbounded variances.* For uncorrelated variables, $(S_n - \mathbb{E}[S_n])/n \xrightarrow{\text{m.s.}} 0$ if and only if $n^{-2} \sum_{i=1}^n \text{Var}(X_i) \rightarrow 0$. This allows, for example, $\text{Var}(X_i) \leq c i^\alpha$ with $\alpha < 1$.
- (b) *Unequal means.* If $\mathbb{E}[X_i] = \mu_i$, the conclusion becomes $\frac{1}{n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{\text{m.s.}} 0$, and then $S_n/n \xrightarrow{\text{m.s.}} \mu$ if and only if $\frac{1}{n} \sum_{i=1}^n \mu_i \rightarrow \mu$.
- (c) *Correlated variables.* Suppose (X_i) is wide-sense stationary with mean μ and autocovariance function $C_X(k) = \text{Cov}(X_{i+k}, X_i)$. Collecting the terms of the double sum along diagonals,

$$\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = \frac{1}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) C_X(k) \leq \frac{1}{n} \sum_{k=-(n-1)}^{n-1} |C_X(k)|.$$

The right-hand side tends to zero whenever $C_X(k) \rightarrow 0$ as $|k| \rightarrow \infty$, being a Cesàro average of a sequence that tends to zero. So it is enough that the variables be *asymptotically uncorrelated*.

Example 3.4 (Uncorrelated but strongly dependent). Let Θ be uniformly distributed on $[0, 2\pi]$ and $X_i = \sqrt{2} \cos(i\Theta)$ for $i \geq 1$. Then $\mathbb{E}[X_i] = 0$, and by the product-to-sum formula,

$$\mathbb{E}[X_i X_j] = \mathbb{E}[\cos((i-j)\Theta)] + \mathbb{E}[\cos((i+j)\Theta)] = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

So (i)–(iii) hold with $\mu = 0$ and $c = 1$, with equality in (iii), and the computation in the proof of Theorem 3.1 gives $\mathbb{E}[(S_n/n)^2] = 1/n \rightarrow 0$. Yet the X_i are as far from independent as possible: all of them are deterministic functions of the single random variable Θ .

Remark 3.5 (Almost sure convergence also holds). Mean square convergence does not in general imply almost sure convergence. Under (i)–(iii), however, $S_n/n \xrightarrow{\text{a.s.}} \mu$ as well, by the method of Section 2. Step 1 used only Chebyshev’s bound, which (4) supplies with c in place of σ^2 , so $S_{k^2}/k^2 \xrightarrow{\text{a.s.}} \mu$. To pass from the squares to all n , use Remark 2.6: its only input is $\mathbb{E}[(S_n - S_{k^2})^2] \leq c(n - k^2)$ for the centered sums, which holds for uncorrelated variables. The route through positive and negative parts (Step 3) is *not* available here, because X_i^+ and X_j^+ need not be uncorrelated when X_i and X_j merely are.

4 Kolmogorov’s strong law under a finite mean

The finite-variance assumption in Theorem 2.1 is a convenience of the method, not a feature of the phenomenon. For iid sequences the strong law holds *if and only if* the mean is finite.

Theorem 4.1 (Kolmogorov’s strong law). *Let $X_1, X_2, \dots$ be iid.*

- (a) *If $\mathbb{E}|X_1| < \infty$ and $\mu = \mathbb{E}[X_1]$, then $S_n/n \xrightarrow{\text{a.s.}} \mu$.*
- (b) *If $\mathbb{E}|X_1| = \infty$, then $\limsup_{n \rightarrow \infty} |S_n|/n = \infty$ almost surely. In particular, S_n/n does not converge to a finite limit.*

We outline the proof of (a), giving the main ideas of each ingredient, and then prove (b) in full since it is short.

4.1 Three ingredients

Without a second moment, Chebyshev’s inequality is unavailable for the X_i themselves. Kolmogorov’s proof restores second moments by *truncation*, and replaces the subsequence and sandwich device of Section 2 by an inequality that controls all partial sums at once.

Lemma 4.2 (Kolmogorov’s maximal inequality). *Let $Z_1, \dots, Z_n$ be independent with mean zero and finite variances, and let $T_k = Z_1 + \dots + Z_k$. Then for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |T_k| \geq \varepsilon\right) \leq \frac{\text{Var}(T_n)}{\varepsilon^2}.$$

This is Chebyshev’s bound for $|T_n|$ alone, yet it controls the whole path $T_1, \dots, T_n$ at no extra cost.

Proof sketch. Let $A_k = \{|T_k| \geq \varepsilon, |T_j| < \varepsilon \text{ for } j < k\}$ be the event that the path first leaves $(-\varepsilon, \varepsilon)$ at time k . The A_k are disjoint with union $\{\max_{k \leq n} |T_k| \geq \varepsilon\}$. Write $T_n = T_k + (T_n - T_k)$ and expand the square:

$$\mathbb{E}[T_n^2 \mathbf{1}_{A_k}] = \mathbb{E}[T_k^2 \mathbf{1}_{A_k}] + 2\mathbb{E}[T_k \mathbf{1}_{A_k} (T_n - T_k)] + \mathbb{E}[(T_n - T_k)^2 \mathbf{1}_{A_k}] \geq \mathbb{E}[T_k^2 \mathbf{1}_{A_k}] \geq \varepsilon^2 \mathbb{P}(A_k).$$

The cross term vanishes because $T_k \mathbf{1}_{A_k}$ is a function of $Z_1, \dots, Z_k$, hence independent of $T_n - T_k$, which has mean zero. Summing over k gives $\text{Var}(T_n) = \mathbb{E}[T_n^2] \geq \varepsilon^2 \mathbb{P}(\max_{k \leq n} |T_k| \geq \varepsilon)$. $\square$

Lemma 4.3 (Kolmogorov's convergence criterion). *Let $Z_1, Z_2, \dots$ be independent with mean zero and $\sum_n \text{Var}(Z_n) < \infty$. Then the series $\sum_n Z_n$ converges almost surely.*

Proof sketch. Apply Lemma 4.2 to $Z_{m+1}, \dots, Z_n$ and let $n \rightarrow \infty$:

$$\mathbb{P}\left(\sup_{k>m} |T_k - T_m| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{j>m} \text{Var}(Z_j) \xrightarrow{m \rightarrow \infty} 0.$$

Hence $M_m = \sup_{j,k>m} |T_j - T_k| \leq 2 \sup_{k>m} |T_k - T_m|$ tends to zero in probability. Since M_m is nonincreasing in m it also has an almost sure limit, which must then be 0. So (T_k) is almost surely a Cauchy sequence. $\square$

Lemma 4.4 (Kronecker's lemma). *Let $a_1, a_2, \dots$ be real numbers. If $\sum_n a_n/n$ converges, then $\frac{1}{n} \sum_{i=1}^n a_i \rightarrow 0$.*

Proof sketch. Let $b_n = \sum_{i=1}^n a_i/i$, $b_0 = 0$, and $b = \lim b_n$. Since $a_i = i(b_i - b_{i-1})$, summation by parts gives

$$\frac{1}{n} \sum_{i=1}^n a_i = b_n - \frac{1}{n} \sum_{i=1}^{n-1} b_i \rightarrow b - b = 0,$$

because the Cesàro averages of a convergent sequence converge to the same limit. $\square$

We also need the standard comparison between a mean and a sum of tail probabilities.

Lemma 4.5. *If $W \geq 0$, then $\sum_{n=1}^{\infty} \mathbb{P}(W > n) \leq \mathbb{E}[W] \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}(W > n)$.*

Proof. Use $\mathbb{E}[W] = \int_0^{\infty} \mathbb{P}(W > t) dt$ and the monotonicity of $t \mapsto \mathbb{P}(W > t)$: $\mathbb{P}(W > n) \leq \int_{n-1}^n \mathbb{P}(W > t) dt$ and $\int_n^{n+1} \mathbb{P}(W > t) dt \leq \mathbb{P}(W > n)$. Sum the first over $n \geq 1$ and the second over $n \geq 0$, and bound $\mathbb{P}(W > 0) \leq 1$. $\square$

4.2 Outline of the proof of Theorem 4.1(a)

Step 1. Truncate. Define $Y_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}$. The Y_n are independent and bounded, but no longer identically distributed. By Lemma 4.5,

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n) \leq \mathbb{E}|X_1| < \infty.$$

By the first Borel–Cantelli lemma, almost surely $X_n = Y_n$ for all but finitely many n , and therefore

$$\frac{1}{n} \sum_{i=1}^n (X_i - Y_i) \xrightarrow{\text{a.s.}} 0.$$

It thus suffices to prove the strong law for the truncated sequence.

Step 2. The truncated means are right on average. Since $|X_1 \mathbf{1}_{\{|X_1| \leq n\}}| \leq |X_1|$ and $\mathbb{E}|X_1| < \infty$, dominated convergence gives $\mathbb{E}[Y_n] = \mathbb{E}[X_1 \mathbf{1}_{\{|X_1| \leq n\}}] \rightarrow \mu$, and hence also

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \rightarrow \mu.$$

Step 3. The truncated variances are summable after scaling. This is the key estimate:

$$\sum_{n=1}^{\infty} \frac{\text{Var}(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}[X_1^2 \mathbf{1}_{\{|X_1| \leq n\}}]}{n^2} = \mathbb{E}\left[X_1^2 \sum_{n \geq \max\{1, |X_1|\}} \frac{1}{n^2}\right] \leq 2\mathbb{E}|X_1| < \infty. \quad (5)$$

The first inequality is $\text{Var}(Y_n) \leq \mathbb{E}[Y_n^2]$, and the interchange of sum and expectation is justified by nonnegativity. For the last inequality, note that for integers $m \geq 1$ we have $\sum_{n \geq m} n^{-2} \leq m^{-2} + \int_m^{\infty} t^{-2} dt \leq 2/m$, and hence, for every real $x \geq 0$,

$$x^2 \sum_{n \geq \max\{1, x\}} \frac{1}{n^2} \leq \frac{2x^2}{\max\{1, x\}} \leq 2x.$$

Note how a *first*-moment assumption produces a *second*-moment bound. The truncation level n is large enough that truncation is eventually invisible (Step 1), and small enough that the truncated variances are under control.

Step 4. Apply the convergence criterion and Kronecker. The variables $Z_n = (Y_n - \mathbb{E}[Y_n])/n$ are independent with mean zero, and $\sum_n \text{Var}(Z_n) < \infty$ by (5). By Lemma 4.3, the series $\sum_n (Y_n - \mathbb{E}[Y_n])/n$ converges almost surely. Applying Lemma 4.4 for each ω in the event of convergence,

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \xrightarrow{\text{a.s.}} 0.$$

Step 5. Combine. Almost surely,

$$\frac{S_n}{n} = \underbrace{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)}_{\rightarrow 0 \text{ (Step 1)}} + \underbrace{\frac{1}{n} \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])}_{\rightarrow 0 \text{ (Step 4)}} + \underbrace{\frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i]}_{\rightarrow \mu \text{ (Step 2)}} \rightarrow \mu.$$

4.3 Proof of Theorem 4.1(b)

Fix $c > 0$. By Lemma 4.5 applied to $W = |X_1|/c$,

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > cn) = \sum_{n=1}^{\infty} \mathbb{P}(|X_1|/c > n) \geq \frac{\mathbb{E}|X_1|}{c} - 1 = \infty.$$

The events $\{|X_n| > cn\}$ are independent, so by the second Borel–Cantelli lemma, almost surely $|X_n| > cn$ for infinitely many n . Since $X_n = S_n - S_{n-1}$,

$$\frac{|X_n|}{n} \leq \frac{|S_n|}{n} + \frac{|S_{n-1}|}{n-1} \quad (n \geq 2),$$

and taking $\limsup$ on both sides gives $c \leq 2 \limsup_n |S_n|/n$ almost surely. Letting c run through the positive integers shows $\limsup_n |S_n|/n = \infty$ almost surely. $\square$

Remark 4.6 (Comparison with Section 2; Etemadi’s proof). The two proofs share a skeleton: get a summable second-moment bound, then apply Borel–Cantelli. They differ in how they pass from “sparse” control to all n . Section 2 uses a subsequence and monotonicity; Kolmogorov uses the maximal inequality, which requires full independence. Etemadi [4] showed that the elementary method of Section 2 is already enough to prove Theorem 4.1(a): truncate as in Step 1 above, reduce to nonnegative variables, apply Chebyshev along the geometric subsequence $n_k = \lfloor \alpha^k \rfloor$ with $\alpha > 1$, sandwich as in (3) to get $\mu/\alpha \leq \liminf S_n/n \leq \limsup S_n/n \leq \alpha\mu$, and let $\alpha \downarrow 1$. That argument needs only *pairwise* independence.

Remark 4.7 (The weak law under a finite mean). Since almost sure convergence implies convergence in probability, Theorem 4.1(a) shows that the weak law also holds for iid sequences assuming only $\mathbb{E}|X_1| < \infty$ (Khinchin’s weak law).

5 Examples

5.1 IID standard Gaussian random variables

Example 5.1. Let $X_1, X_2, \dots$ be iid $N(0, 1)$. Here $\mu = 0$ and $\sigma^2 = 1$, so Theorems 1.1 and 2.1 apply and $S_n/n \rightarrow 0$ in probability, in mean square, and almost surely. Because sums of independent Gaussians are Gaussian, everything can also be computed exactly:

$$\frac{S_n}{n} \sim N\left(0, \frac{1}{n}\right), \quad \mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) = 2Q(\varepsilon\sqrt{n}),$$

where $Q(x) = 1 - \Phi(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ is the standard normal tail function.

A much better bound from the same inequality. Markov’s inequality can be applied to any nonnegative function of S_n . Choosing an exponential instead of a square (the *Chernoff bound*) gives, for $\theta > 0$,

$$\mathbb{P}\left(\frac{S_n}{n} \geq \varepsilon\right) = \mathbb{P}(e^{\theta S_n} \geq e^{\theta n\varepsilon}) \leq e^{-\theta n\varepsilon} \mathbb{E}[e^{\theta S_n}] = \exp\left(-\theta n\varepsilon + \frac{n\theta^2}{2}\right),$$

using $\mathbb{E}[e^{\theta X_1}] = e^{\theta^2/2}$. The exponent is minimized at $\theta = \varepsilon$. By symmetry the same bound holds for the lower tail, so

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq 2e^{-n\varepsilon^2/2}. \tag{6}$$

A direct proof of the strong law. The bound (6) decays geometrically in n and is therefore summable over *all* n . The criterion (1) gives $S_n/n \xrightarrow{\text{a.s.}} 0$ immediately, with no need for subsequences or sandwich bounds. The work in Section 2 was needed precisely because, with only a second moment, Chebyshev’s $1/n$ decay is the best available.

Numerical comparison. Table 1 compares the three quantities for $\varepsilon = 0.1$. Chebyshev’s bound is valid for every finite-variance distribution, and for that reason cannot see the Gaussian tail.

Scaling. Since $S_n/\sqrt{n} \sim N(0, 1)$ exactly for every n , the fluctuations of the sample mean are of order $1/\sqrt{n}$, and the central limit theorem of Section 6 holds here with equality rather than in the limit.

n	Chebyshev	$1/(n\varepsilon^2)$	Chernoff	$2e^{-n\varepsilon^2/2}$	Exact	$2Q(\varepsilon\sqrt{n})$
100	1		1.21 (trivial)		0.317	
1,000	0.1		1.35×10^{-2}		1.57×10^{-3}	
10,000	0.01		3.86×10^{-22}		1.52×10^{-23}	

Table 1: Bounds on $\mathbb{P}(|S_n/n| \geq \varepsilon)$ for iid $N(0, 1)$ samples, $\varepsilon = 0.1$.

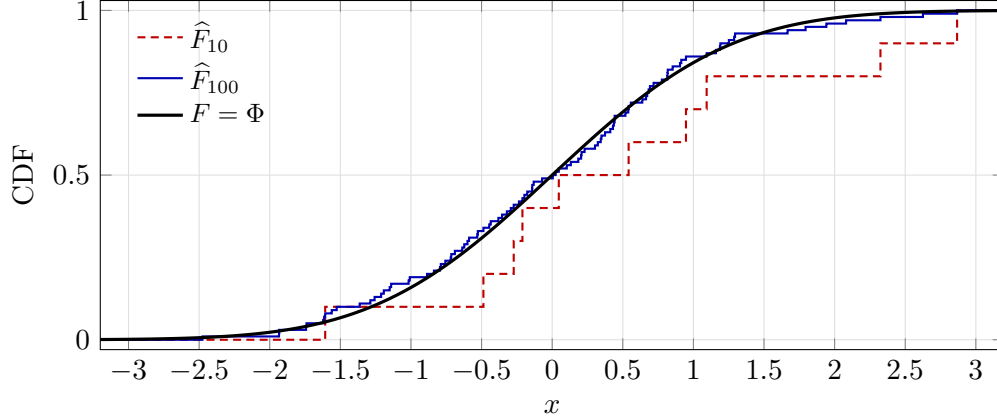


Figure 1: Empirical CDFs of the first 10 and the first 100 terms of one simulated iid $N(0, 1)$ sequence, together with the true CDF Φ .

5.2 The empirical CDF converges to the true CDF

Example 5.2. Let $X_1, X_2, \dots$ be iid with CDF $F(x) = \mathbb{P}(X_1 \leq x)$; no moment assumption is made. The *empirical CDF* of the first n samples is

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}, \quad x \in \mathbb{R},$$

the fraction of the first n observations that are at most x . It is a random, piecewise constant CDF with a jump of size $1/n$ at each observation (Figure 1).

Pointwise convergence. Fix x . The indicators $\mathbf{1}_{\{X_i \leq x\}}$, $i \geq 1$, are iid Bernoulli random variables with mean $F(x)$ and variance $F(x)(1 - F(x)) \leq \frac{1}{4}$. The laws of large numbers therefore apply to them whatever the distribution of the X_i may be:

$$\mathbb{P}(|\hat{F}_n(x) - F(x)| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2} \quad (\text{Theorem 1.1}), \quad \hat{F}_n(x) \xrightarrow{\text{a.s.}} F(x) \quad (\text{Theorem 2.1}).$$

The same reasoning applied to the indicators $\mathbf{1}_{\{X_i < x\}}$ gives $\hat{F}_n(x-) \xrightarrow{\text{a.s.}} F(x-)$, where $G(x-)$ denotes the left limit of G at x , so that $F(x-) = \mathbb{P}(X_1 < x)$.

Pointwise convergence is a statement about each x separately. Monotonicity of CDFs upgrades it to *uniform* convergence, by another sandwich argument.

Theorem 5.3 (Glivenko–Cantelli). *Let $X_1, X_2, \dots$ be iid with CDF F . Then*

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0.$$

Proof. Fix an integer $m \geq 1$ and place a grid at the quantiles of F : let $x_0 = -\infty$, $x_m = +\infty$, and

$$x_j = \inf\{x : F(x) \geq j/m\}, \quad j = 1, \dots, m-1.$$

Then $x_0 \leq x_1 \leq \dots \leq x_m$, and by right-continuity of F and the definition of the infimum, $F(x_{j-}) \leq j/m \leq F(x_j)$. Consequently

$$F(x_j-) - F(x_{j-1}) \leq \frac{1}{m}, \quad j = 1, \dots, m, \quad (7)$$

with the conventions $F(-\infty) = \hat{F}_n(-\infty) = 0$ and $F(+\infty-) = \hat{F}_n(+\infty-) = 1$.

Every real x lies in exactly one interval $[x_{j-1}, x_j)$. For such x , monotonicity of $\hat{F}_n$ and F together with (7) gives

$$\begin{aligned} \hat{F}_n(x) - F(x) &\leq \hat{F}_n(x_j-) - F(x_{j-1}) \leq [\hat{F}_n(x_j-) - F(x_j-)] + \frac{1}{m}, \\ \hat{F}_n(x) - F(x) &\geq \hat{F}_n(x_{j-1}) - F(x_j-) \geq [\hat{F}_n(x_{j-1}) - F(x_{j-1})] - \frac{1}{m}. \end{aligned}$$

Therefore

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \leq \Delta_{n,m} + \frac{1}{m}, \quad \Delta_{n,m} = \max_{0 \leq j \leq m} \max\{|\hat{F}_n(x_j) - F(x_j)|, |\hat{F}_n(x_j-) - F(x_j-)|\}.$$

$\Delta_{n,m}$ is a maximum of finitely many terms, each of which tends to zero almost surely by Example 5.2. So there is an event Ω_m of probability one on which $\Delta_{n,m} \rightarrow 0$ as $n \rightarrow \infty$, and on Ω_m we get $\limsup_n \sup_x |\hat{F}_n(x) - F(x)| \leq 1/m$. On the probability-one event $\bigcap_{m \geq 1} \Omega_m$ this holds for every m , which proves the theorem. $\square$

Remark 5.4. The Glivenko–Cantelli theorem is the reason histograms and empirical quantiles are meaningful: with enough iid data the entire distribution, not just its mean, can be recovered. A quantitative version is the Dvoretzky–Kiefer–Wolfowitz inequality, $\mathbb{P}(\sup_x |\hat{F}_n(x) - F(x)| > \varepsilon) \leq 2e^{-2n\varepsilon^2}$, valid for every F , every n , and every $\varepsilon > 0$.

5.3 A counterexample: the Cauchy distribution

Example 5.5. Let $X_1, X_2, \dots$ be iid standard Cauchy, with density and CDF

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x, \quad x \in \mathbb{R}.$$

The density is symmetric about 0, which is the median and the obvious “center” of the distribution. Nevertheless the mean does not exist:

$$\mathbb{E}|X_1| = \frac{2}{\pi} \int_0^\infty \frac{x}{1+x^2} dx = \frac{1}{\pi} \lim_{a \rightarrow \infty} \ln(1+a^2) = \infty.$$

Equivalently, the tails are too heavy: $\mathbb{P}(|X_1| > x) = \frac{2}{\pi} \arctan(1/x) \sim \frac{2}{\pi x}$ as $x \rightarrow \infty$, which is not integrable.

The sample mean has the same distribution as a single sample. The characteristic function of the standard Cauchy distribution is $\varphi_X(u) = \mathbb{E}[e^{juX}] = e^{-|u|}$. By independence,

$$\varphi_{S_n/n}(u) = [\varphi_X(u/n)]^n = [e^{-|u|/n}]^n = e^{-|u|},$$

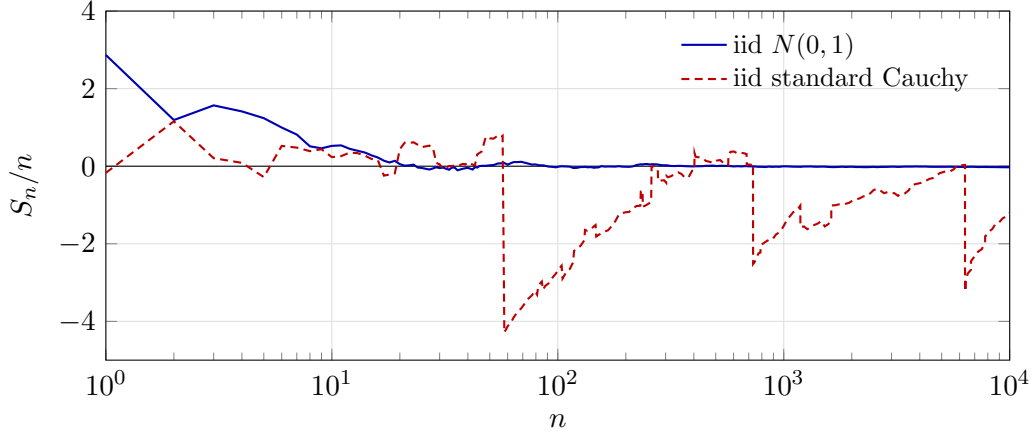


Figure 2: Running sample means S_n/n , $1 \leq n \leq 10^4$, for one simulated sequence of iid $N(0, 1)$ variables and one of iid standard Cauchy variables. The Gaussian path settles at $\mu = 0$; the Cauchy path is repeatedly knocked away by single large observations and never settles.

so S_n/n is again standard Cauchy, *for every* n . Averaging does not concentrate the distribution at all:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) = 1 - \frac{2}{\pi} \arctan \varepsilon \quad \text{for all } n \geq 1,$$

which does not tend to zero. So S_n/n does not converge in probability to 0, nor to any other constant.

No weak law, to any limit. In fact S_n/n does not converge in probability to any random variable. If it did, then $S_{2n}/(2n) - S_n/n \xrightarrow{\mathbb{P}} 0$. But

$$\frac{S_{2n}}{2n} - \frac{S_n}{n} = \frac{1}{2}(V_n - U_n), \quad U_n = \frac{S_n}{n}, \quad V_n = \frac{S_{2n} - S_n}{n},$$

where U_n and V_n are independent and each standard Cauchy. The characteristic function of $\frac{1}{2}(V_n - U_n)$ is $e^{-|u|/2} e^{-|u|/2} = e^{-|u|}$, so this difference is itself standard Cauchy for every n and, for instance, $\mathbb{P}(|S_{2n}/(2n) - S_n/n| \geq 1) = \frac{1}{2}$ for all n .

No strong law. Since $\mathbb{E}|X_1| = \infty$, Theorem 4.1(b) gives $\limsup_n |S_n|/n = \infty$ almost surely. The mechanism is visible in the proof: infinitely often a single new observation X_n is of the same order as n , and by itself moves the running average by an amount of order one (Figure 2).

Moral. The Cauchy example shows that the moment assumption in the laws of large numbers cannot simply be dropped, and Theorem 4.1 shows that a finite mean is exactly the right assumption for the strong law. It also shows that convergence in distribution is much weaker than the other modes: S_n/n converges in distribution trivially (its distribution does not depend on n) while failing to converge in probability. Finally, the center of a Cauchy distribution *can* be estimated consistently, just not by the sample mean: since F is continuous and strictly increasing, Theorem 5.3 implies that the sample median converges almost surely to the true median 0.

6 The central limit theorem

The laws of large numbers say that $S_n/n - \mu \rightarrow 0$. The central limit theorem identifies the rate, $1/\sqrt{n}$, and the shape of the fluctuations at that scale.

Theorem 6.1 (Central limit theorem). *Let $X_1, X_2, \dots$ be iid with mean μ and variance $0 < \sigma^2 < \infty$. Then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mu \right) \xrightarrow{d} N(0, 1).$$

That is, for every $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \right) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Remark 6.2 (Three scalings of the same sum). For iid variables with mean zero and finite variance $\sigma^2 > 0$:

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0 \quad (\text{law of large numbers}), \quad \frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2) \quad (\text{central limit theorem}),$$

while S_n itself does not converge. The limit in the CLT is in distribution only: $S_n/\sqrt{n}$ does not converge in probability or almost surely.

Remark 6.3 (Universality and its limits). The limit depends on the distribution of the X_i only through μ and σ^2 . The finite-variance assumption cannot be dropped: for iid standard Cauchy variables (Example 5.5), S_n/n is standard Cauchy for every n , so no normalization of S_n yields a Gaussian limit.

Remark 6.4 (Sample size revisited). The CLT suggests the approximation $\mathbb{P}(|S_n/n - \mu| \geq \varepsilon) \approx 2Q(\varepsilon\sqrt{n}/\sigma)$. For $\sigma = 1$, $\varepsilon = 0.1$ and a target of 0.05 this gives $n \approx (1.96/0.1)^2 \approx 384$, compared with the guarantee $n \geq 2000$ obtained from Chebyshev's inequality in Remark 1.3. The Chebyshev figure is a bound valid for every n and every finite-variance distribution; the CLT figure is an approximation whose accuracy depends on the distribution. If in addition $\rho = \mathbb{E}|X_1 - \mu|^3 < \infty$, the Berry–Esseen theorem quantifies the error:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \right) - \Phi(x) \right| \leq \frac{C\rho}{\sigma^3\sqrt{n}}$$

for a universal constant C (one may take $C = 1/2$).

References

- [1] B. Hajek, *Random Processes for Engineers*, Cambridge University Press, 2015.
- [2] R. Durrett, *Probability: Theory and Examples*, 5th ed., Cambridge University Press, 2019.
- [3] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley, 1995.
- [4] N. Etemadi, “An elementary proof of the strong law of large numbers,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 55, pp. 119–122, 1981.