

The Strong Law of Large Numbers and the Borel-Cantelli Lemmas

Zaikun Zhang *

October 12, 2024 11:25pm +08:00

For a sequence of random variables $\{X_k\}$, we will say it is nonnegative if $X_k \geq 0$ a.s. for each k , and uniformly bounded if there exists a constant M such that $|X_k| \leq M$ a.s. for each k . In addition, we define $\bar{X}_k = k^{-1} \sum_{\ell=1}^k X_\ell$ for each $k \geq 1$.

1 The strong law of large numbers

1.1 The i.i.d. case

Theorem 1.1. *Let $\{X_k\}$ be a sequence of i.i.d. random variables.*

- (a) $\{\bar{X}_k\}$ converges a.s. to a finite random variable if and only if $\mathbb{E}(|X_1|) < \infty$. When it converges a.s., the limit is $\mathbb{E}(X_1)$.
- (b) If $\{X_k\}$ is nonnegative, then $\{\bar{X}_k\}$ converges to $\mathbb{E}(X_1)$ a.s..

Proof. (a) See [8, Theorem 5.23].

(b) If $\mathbb{E}(X_1) < \infty$, then the convergence holds according to (a). Otherwise, for each integer $n \geq 1$, we have

$$\liminf_{k \rightarrow \infty} \bar{X}_k \geq \liminf_{k \rightarrow \infty} \frac{1}{k} \sum_{\ell=1}^k \min\{X_\ell, n\} = \mathbb{E}(\min\{X_1, n\}).$$

Letting $n \rightarrow \infty$, we have $\liminf_k \bar{X}_k = \infty$ a.s., because $\mathbb{E}(\min\{X_1, n\}) \rightarrow \mathbb{E}(X_1) = \infty$ according to Levi's monotone convergence theorem. Hence $\bar{X}_k \rightarrow \infty = \mathbb{E}(X_1)$ a.s. \square

*Email: zaikunzhang@gmail.com.

Remark 1.1. For the “if” part of Theorem 1.1, see also [5, Theorem 1] and [4, Theorem 2.4.1], which show that the mutual independence assumption can be weakened to pairwise independence.

Item (b) of Theorem 1.1 can be generalized to the following, with a very similar proof.

Theorem 1.2. *Let $\{X_k\}$ be a sequence of i.i.d. random variables. If $\mathbb{E}(X_1^+) < \infty$ or $\mathbb{E}(X_1^-) < \infty$, then $\{\bar{X}_k\}$ converges to $\mathbb{E}(X_1)$ a.s.*

Proof. If $\mathbb{E}(X_1^+) < \infty$ and $\mathbb{E}(X_1^-) < \infty$, then $\mathbb{E}(|X_1|) < \infty$. Thus the convergence holds. If $\mathbb{E}(X_1^+) = \infty$ and $\mathbb{E}(X_1^-) < \infty$, then $\mathbb{E}(X_1) = -\infty$ and the convergence can be established by considering the truncated sequence $\{\min\{X_k, n\}\}$ for each integer $n \geq 1$. \square

1.2 Martingale difference sequences

Theorem 1.3 (Martingale convergence theorem [4, Theorem 4.2.11]). *If $\{X_k\}$ is a submartingale with $\sup_k \mathbb{E}(X_k^+) < \infty$, then $X_k \rightarrow X$ a.s. for some random variable X with $\mathbb{E}(|X|) < \infty$.*

Corollary 1.1 ([4, Theorem 4.2.12]). *If $\{X_k\}$ is a nonnegative supermartingale, then $X_k \rightarrow X$ a.s. for some random variable X with $\mathbb{E}(X) \leq \mathbb{E}(X_1)$.*

By Theorem 1.3, if $\{X_k\}$ is an L^1 -bounded martingale, then $X_k \rightarrow X$ a.s. for some random variable $X \in L^1$. However, the convergence may not hold in L^1 . See [4, Example 4.2.13]. In contrast, the following theorem ensures the convergence in L^p for $p > 1$.

Theorem 1.4 (Martingale L^p convergence theorem [4, Theorem 4.4.6]). *If $p > 1$ and $\{X_k\}$ is a martingale with $\sup_{k \geq 1} \|X_k\|_p < \infty$, then there exists a random variable $X \in L^p$ such that $X_k \rightarrow X$ a.s. and in L^p .*

Lemma 1.1 (Kronecker’s Lemma). *Let $\{a_k\}$ be a sequence of nonnegative numbers and $\{b_k\}$ be an increasing sequence with $b_k \rightarrow \infty$.*

(a) *If $\sum_{k=1}^{\infty} b_k^{-1} a_k < \infty$, then $b_k^{-1} \sum_{\ell=1}^k a_{\ell} \rightarrow 0$.*

(b) *If $\sum_{k=1}^{\infty} a_k < \infty$, then $b_k^{-1} \sum_{\ell=1}^k b_{\ell} a_{\ell} \rightarrow 0$.*

Remark 1.2. In item (a) of Lemma 1.1, we trivially have $b_k^{-1} a_k \rightarrow 0$, and Kronecker’s lemma strengthens it to $b_k^{-1} \sum_{\ell=1}^k a_{\ell} \rightarrow 0$. In item (b), we trivially have $b_k^{-1} \sum_{\ell=1}^k a_{\ell} \rightarrow 0$, and Kronecker’s lemma strengthens it to $b_k^{-1} \sum_{\ell=1}^k b_{\ell} a_{\ell} \rightarrow 0$.

Definition 1.1. Let $\{X_k\}$ be a sequence of random variables and $\{\mathcal{F}_k\}$ be a filtration. Then $\{X_k\}$ is called a martingale difference sequence adapted to $\{\mathcal{F}_k\}$ if there exists a martingale $\{S_k\}$ adapted to $\{\mathcal{F}_k\}$ with $S_0 = 0$ such that $X_k = S_k - S_{k-1}$ for each $k \geq 1$, or, equivalently, if $\mathbb{E}(|X_k|) < \infty$ and $\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) = 0$ a.s. for each $k \geq 1$.

Theorem 1.5 (Chow [3]). *Let $\{X_k\}$ be a martingale difference sequence. If there exists a constant $p \geq 1$ such that*

$$\sum_{k=1}^{\infty} k^{-p-1} \mathbb{E}(|X_k|^{2p}) < \infty,$$

then $\bar{X}_k \rightarrow 0$ a.s.

Remark 1.3. Chow's theorem is a consequence of [2, Theorem 2] and [1, Theorem 9]. Chow [3] mentions these references without specifying the theorems. See also

<https://math.stackexchange.com/questions/4965300/>.

Remark 1.4. The special case of Theorem 1.5 with $p = 1$ can be obtained by the martingale L^p convergence theorem and the Kronecker Lemma. Define $Y_k = \sum_{\ell=1}^k \ell^{-1} X_\ell$ for each $k \geq 1$. Then $\{Y_k\}$ is a martingale. In addition, noting that

$$\mathbb{E}(X_i X_j) = \mathbb{E}(X_j \mathbb{E}(X_i \mid \mathcal{F}_{i-1})) = 0, \quad i > j \geq 1,$$

we have

$$\mathbb{E}(Y_k^2) = \sum_{\ell=1}^k \ell^{-2} \mathbb{E}(X_\ell^2) \leq \sum_{\ell=1}^{\infty} \ell^{-2} \mathbb{E}(X_\ell^2) < \infty.$$

Hence $\{Y_k\}$ is an L^2 -bounded martingale, and hence converges a.s. Invoking the Kronecker lemma, we have $\bar{X}_k \rightarrow 0$ a.s.

Theorem 1.6 ([6, Theorem 2.18]). *If $1 \leq p \leq 2$ and $\{X_k\}$ is a martingale difference sequence adapted to a filtration $\{\mathcal{F}_k\}$, then*

$$\{\bar{X}_k \rightarrow 0\} \supset \left\{ \sum_{k=1}^{\infty} k^{-p} \mathbb{E}(|X_k|^p \mid \mathcal{F}_{k-1}) < \infty \right\}.$$

In particular, if $\sum_{k=1}^{\infty} k^{-p} \mathbb{E}(|X_k|^p) < \infty$, then $\bar{X}_k \rightarrow 0$ a.s.

1.3 The independent case with moment conditions

Lemma 1.2. *Let X be a random and $p \geq 1$ be a constant.*

$$(a) \mathbb{E}(|X|) \leq \mathbb{E}(|X|^p)^{\frac{1}{p}}.$$

$$(b) \text{ If } \mathbb{E}(|X|) < \infty, \text{ then } \mathbb{E}(|X - \mathbb{E}(X)|^p) \leq 2^p \mathbb{E}(|X|^p).$$

Proof. (a) Jensen's inequality (or Hölder's inequality).

$$(b) \text{ By the triangle inequality, we have } \|X - \mathbb{E}(X)\|_p \leq \|X\|_p + |\mathbb{E}(X)| \leq 2\|X\|_p. \quad \square$$

Theorem 1.7. *Let $\{X_k\}$ be a sequence of independent random variables. If either*

$$\sum_{k=1}^{\infty} k^{-p-1} \mathbb{E}(|X_k|^{2p}) < \infty \tag{1.1}$$

for a constant $p \geq 1$, or

$$\sum_{k=1}^{\infty} k^{-p} \mathbb{E}(|X_k|^p) < \infty \tag{1.2}$$

for a constant $p \in [1, 2]$, then $\bar{X}_k - \mathbb{E}(\bar{X}_k) \rightarrow 0$ a.s.

Proof. According to Lemma 1.2, we have $\mathbb{E}(|X_k|) < \infty$ for each $k \geq 1$ in both cases. Define $Y_k = X_k - \mathbb{E}(X_k)$ for each $k \geq 1$. Then $\{Y_k\}$ is a martingale difference sequence. In addition,

$$\mathbb{E}(|Y_k|^p) \leq 2^p \mathbb{E}(|X_k|^p)$$

for all $k \geq 1$ and $p \geq 1$. Hence (1.3) implies

$$\sum_{k=1}^{\infty} k^{-p-1} \mathbb{E}(|Y_k|^{2p}) < \infty, \tag{1.3}$$

and (1.4) implies

$$\sum_{k=1}^{\infty} k^{-p} \mathbb{E}(|Y_k|^p) < \infty. \tag{1.4}$$

In both cases, $\bar{Y}_k \rightarrow 0$ a.s. by Theorems 1.5 and 1.6. Thus $\bar{X}_k - \mathbb{E}(\bar{X}_k) \rightarrow 0$ a.s. \square

2 The Borel-Cantelli Lemmas

Theorem 2.1 (The Borel-Cantelli Lemmas). *Let $\{E_k\}$ be a sequence of events.*

$$(a) \text{ If } \sum_{k=1}^{\infty} \mathbb{P}(E_k) < \infty, \text{ then } \mathbb{P}(E_k \text{ i.o.}) = 0.$$

(b) If $\{E_k\}$ is mutually independent and $\sum_{k=1}^{\infty} \mathbb{P}(E_k) = \infty$, then $\mathbb{P}(E_k \text{ i.o.}) = 1$.

Remark 2.1. The Borel-Cantelli Lemmas lead to a 0-1 law: For a sequence of mutually independent events $\{E_k\}$, the probability $\mathbb{P}(\{E_k \text{ i.o.}\})$ is either 0 or 1, corresponding to $\sum_{k=1}^{\infty} \mathbb{P}(E_k) < \infty$ or $\sum_{k=1}^{\infty} \mathbb{P}(E_k) = \infty$, respectively.

Lemma 2.1. For a sequence $\{x_k\} \subset [0, 1)$, $\sum_{k=1}^{\infty} x_k = \infty$ if and only if $\prod_{k=1}^{\infty} (1 - x_k) = 0$.

Proof. If $\sum_{k=1}^{\infty} x_k = \infty$, then $\prod_{k=1}^{\infty} (1 - x_k) \leq \exp(-\sum_{k=1}^{\infty} x_k) = 0$. If $\sum_{k=1}^{\infty} x_k < \infty$, then there exists an integer K such that, for each $k \geq K$, we have $x_k \leq 1/2$ and hence $1 - x_k \geq \exp(-2x_k)$. Thus $\prod_{k=K}^{\infty} (1 - x_k) \geq \exp(-2\sum_{k=K}^{\infty} x_k) > 0$, implying that $\prod_{k=1}^{\infty} (1 - x_k) > 0$. \square

Theorem 2.2. Let $\{X_k\}$ be a sequence of nonnegative random variables.

- (a) If $\sum_{k=1}^{\infty} \mathbb{E}(X_k) < \infty$, then $\sum_{k=1}^{\infty} X_k < \infty$ a.s.
- (b) If $\sum_{k=1}^{\infty} \mathbb{E}(X_k) = \infty$, then $\sum_{k=1}^{\infty} X_k = \infty$ a.s. provided that the random variables $\{X_k\}$ are i.i.d. or mutually independent and uniformly bounded.

Proof. (a) By the Fubini-Tonelli theorem, we have

$$\mathbb{E}\left(\sum_{k=1}^{\infty} X_k\right) = \sum_{k=1}^{\infty} \mathbb{E}(X_k) < \infty.$$

Hence $\sum_{k=1}^{\infty} X_k < \infty$ a.s.

(b) Suppose that $\{X_k\}$ is i.i.d.. By the strong law of large numbers (Theorem 1.1), we have $k^{-1} \sum_{\ell=1}^k X_{\ell} \rightarrow \mathbb{E}(X_1)$ a.s., which is positive (possibly infinite) since $\sum_{k=1}^{\infty} \mathbb{E}(X_k) = \infty$. Hence $\sum_{k=1}^{\infty} X_k = \infty$ a.s.

Now suppose that $\{X_k\}$ is a sequence of mutually independent and uniformly bounded random variables. Without loss of generality, we assume that $X_k < 1$ a.s. for each $k \geq 1$. Then $\mathbb{E}(X_k) < 1$ for each $k \geq 1$. According to Lemma 2.1, $\sum_{k=1}^{\infty} \mathbb{E}(X_k) = \infty$ implies that $\prod_{k=1}^{\infty} [1 - \mathbb{E}(X_k)] = 0$. Due to the independence of $\{X_k\}$, we have

$$\mathbb{E}\left[\prod_{k=1}^{\infty} (1 - X_k)\right] = \prod_{k=1}^{\infty} \mathbb{E}(1 - X_k) = \prod_{k=1}^{\infty} [1 - \mathbb{E}(X_k)] = 0.$$

Thus $\prod_{k=1}^{\infty} (1 - X_k) = 0$ a.s. since this is a nonnegative random variable, implying that $\sum_{k=1}^{\infty} X_k = \infty$ a.s. according to Lemma 2.1. \square

Remark 2.2. The Borel-Cantelli Lemmas are special cases of Theorem 2.3 with $X_k = \mathbb{1}(E_k)$. Note that $\{E_k \text{ i.o.}\} = \{\sum_{k=1}^{\infty} X_k = \infty\}$.

The following theorem generalizes item (b) of Theorem 2.2 in the uniformly bounded case.

Theorem 2.3. *Let $\{X_k\}$ be a sequence of uniformly bounded random variables adapted to a filtration $\{\mathcal{F}_k\}$. Define*

$$A = \left\{ \sum_{k=1}^{\infty} X_k = \infty \right\} \quad \text{and} \quad B = \left\{ \sum_{k=1}^{\infty} \mathbb{E}(X_k \mid \mathcal{F}_{k-1}) = \infty \right\}.$$

(a) *If $\{\mathbb{E}(X_k \mid \mathcal{F}_{k-1})\}$ is nonnegative, then $\mathbb{P}(A \cap B^c) = 0$.*

(b) *If $\{X_k\}$ is nonnegative, then $\mathbb{P}((A \cap B^c) \cup (B \cap A^c)) = 0$.*

Proof. This proof is essentially the same as that of [4, Theorem 4.3.4]. Define

$$Y_k = \sum_{\ell=1}^k X_{\ell} - \sum_{\ell=1}^k \mathbb{E}(X_{\ell} \mid \mathcal{F}_{\ell-1}).$$

Then $\{Y_k\}$ is a martingale with respect to the filtration $\{\mathcal{F}_k\}$, and it has bounded increments if $\{X_k\}$ is uniformly bounded. Hence $\mathbb{P}(C \cup D) = 1$ with

$$C = \left\{ \lim_{k \rightarrow \infty} Y_k \text{ exists and is finite} \right\},$$

$$D = \left\{ \liminf_{k \rightarrow \infty} Y_k = -\infty \text{ and } \limsup_{k \rightarrow \infty} Y_k = \infty \right\}.$$

It is clear that

$$A \cap C = B \cap C. \tag{2.1}$$

Suppose that $\{\mathbb{E}(X_k \mid \mathcal{F}_{k-1})\}$ is nonnegative. Then

$$B^c = \left\{ \sum_{k=1}^{\infty} \mathbb{E}(X_k \mid \mathcal{F}_{k-1}) < \infty \right\}.$$

Therefore, we can check that

$$A \cap B^c \cap D = \emptyset. \tag{2.2}$$

Combining (2.1) and (2.2), we have

$$(A \cap B^c) \cap (C \cup D) = \emptyset,$$

which implies that $\mathbb{P}(A \cap B^c) = 0$ since $\mathbb{P}(C \cap D) = 1$.

If $\{X_k\}$ is nonnegative, then so is $\{\mathbb{E}(X_k \mid \mathcal{F}_{k-1})\}$, and hence $\mathbb{P}(A \cap B^c) = 0$. In addition, we have

$$A^c = \left\{ \sum_{k=1}^{\infty} X_k < \infty \right\},$$

which implies $\mathbb{P}(B \cap A^c) = 0$ in a way similar to the above. Thus $\mathbb{P}((A \cap B^c) \cup (B \cap A^c)) = 0$. \square

Remark 2.3. $\mathbb{P}((A \cap B^c) \cup (B \cap A^c)) = 0$ means that the symmetric difference between the events A and B is a null event, or the two events are equal except for a null event.

Remark 2.4. [4, Theorem 4.3.4] is a special case of Theorem 2.3, where $X_k = \mathbb{1}(E_k)$ with an event $E_k \in \mathcal{F}_k$ for each $k \geq 1$. However, the conclusion of [4, Theorem 4.3.4] is

$$\{E_k \text{ i.o.}\} = \left\{ \sum_{k=1}^{\infty} \mathbb{P}(E_k \mid \mathcal{F}_{k-1}) = \infty \right\}, \quad (2.3)$$

which seems not completely rigorous. The correct conclusion should be that the events on the left and right sides of (2.3) are equal except for a null event. In addition, in the second line of the proof of [4, Theorem 4.3.4], the A and B seem to be reversed. Finally, the proof refers to Doob's decomposition theorem ([4, Theorem 4.3.2]), which seems not necessary.

The following proposition shows that $\mathbb{P}(B \cap A^c)$ is not necessarily zero in item (a) of Theorem 2.3.

Proposition 2.1. *Let $\{X_k\}$ be a sequence of mutually independent random variables taking either 1 or -1 with $\mathbb{E}(X_k) = Ck^{-\alpha}$ for each $k \geq 1$, where C and α are positive constants. Then, almost surely,*

$$\liminf_{k \rightarrow \infty} \sum_{\ell=1}^k X_{\ell} = \begin{cases} +\infty & \text{if } \alpha < \frac{1}{2}, \\ -\infty & \text{if } \alpha \geq \frac{1}{2}. \end{cases}$$

Proof. According to Kolmogorov's iterated law of logarithm [9] (see also [10, Theorems 7.1–7.3]), we have

$$\liminf_{k \rightarrow \infty} \frac{\sum_{\ell=1}^k [X_{\ell} - \mathbb{E}(X_{\ell})]}{\sqrt{2B_k \log \log B_k}} = -1, \quad (2.4)$$

where $B_k = \sum_{\ell=1}^k \text{Var}(X_{\ell})$. The law is applicable because $B_k \rightarrow \infty$ and

$$\|X_k - \mathbb{E}(X_k)\|_{\infty} = o(\sqrt{B_k / \log \log B_k}).$$

Since $B_k = \sum_{\ell=1}^k (1 - 1/\ell)$, it is clear that $(B_k \log \log B_k)/(k \log \log k) \rightarrow 1$. Hence (2.4) implies

$$\liminf_{k \rightarrow \infty} \frac{\sum_{\ell=1}^k [X_\ell - \mathbb{E}(X_\ell)]}{\sqrt{2k \log \log k}} = -1. \quad (2.5)$$

Meanwhile, since $\mathbb{E}(X_k) = Ck^{-\alpha}$, we have

$$\lim_{k \rightarrow \infty} \frac{\sum_{\ell=1}^k \mathbb{E}(X_\ell)}{\sqrt{2k \log \log k}} = \begin{cases} +\infty & \text{if } \alpha < \frac{1}{2}, \\ 0 & \text{if } \alpha \geq \frac{1}{2}. \end{cases} \quad (2.6)$$

Combining (2.5) and (2.6), we have

$$\liminf_{k \rightarrow \infty} \frac{\sum_{\ell=1}^k X_\ell}{\sqrt{2k \log \log k}} = \begin{cases} +\infty & \text{if } \alpha < \frac{1}{2}, \\ -1 & \text{if } \alpha \geq \frac{1}{2}, \end{cases}$$

which implies the desired conclusion. \square

Remark 2.5. For more discussions on Proposition 2.1, including a perspective from the Kakutani's Dichotomy theorem [7], see

<https://math.stackexchange.com/questions/4968996/>.

References

- [1] D. L. Burkholder. Martingale transforms. *Ann. Math. Statist.*, 37:1494–1504, 1966.
- [2] Y. S. Chow. A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc.*, 11:107–111, 1960.
- [3] Y. S. Chow. On a strong law of large numbers for martingales. *Ann. Math. Statist.*, 38, 1967.
- [4] R. Durrett. *Probability: Theory and Examples*. Camb. Ser. Stat. Probab. Math. Cambridge University Press, Cambridge, fifth edition, 2019.
- [5] N. Etemadi. An elementary proof of the strong law of large numbers. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 55:119–122, 1981.
- [6] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980.

- [7] S. Kakutani. On equivalence of infinite product measures. *Ann. of Math.*, 49(1):214–224, 1948.
- [8] O. Kallenberg. *Foundations of Modern Probability*. Springer Nature Switzerland AG, Gewerbestrasse, third edition, 2021.
- [9] A. Kolmogoroff. Über das Gesetz des iterierten Logarithmus. *Math. Ann.*, 101:126–135, 1929.
- [10] V. V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Clarendon Press, Oxford, 1995.