

LAWS OF LARGE NUMBERS – II

**Contents**

1. The strong law of large numbers
2. The Chernoff bound

**1 THE STRONG LAW OF LARGE NUMBERS**

While the weak law of large numbers establishes convergence of the sample mean, in probability, the strong law establishes almost sure convergence.

Before we proceed, we point out two common methods for proving almost sure convergence.

**Proposition 1.** *Let  $\{X_n\}$  be a sequence of random variables, not necessarily independent.*

(i) *If  $\sum_{n=1}^{\infty} \mathbb{E}[|X_n|^s] < \infty$ , and  $s > 0$ , then  $X_n \xrightarrow{\text{a.s.}} 0$ .*

(ii) *If  $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > \epsilon) < \infty$ , for every  $\epsilon > 0$ , then  $X_n \xrightarrow{\text{a.s.}} 0$ .*

(iii)  *$X_n \xrightarrow{\text{a.s.}} 0$  iff for every  $\epsilon > 0$  we have  $\mathbb{P}[\sup_{m \geq n} |X_m| > \epsilon] \rightarrow 0$  as  $n \rightarrow \infty$ .*

**Proof.** (i) By the monotone convergence theorem, we obtain  $\mathbb{E}[\sum_{n=1}^{\infty} |X_n|^s] < \infty$ , which implies that the random variable  $\sum_{n=1}^{\infty} |X_n|^s$  is finite, with probability 1. Therefore,  $|X_n|^s \xrightarrow{\text{a.s.}} 0$ , which also implies that  $X_n \xrightarrow{\text{a.s.}} 0$ .

(ii) Setting  $\epsilon = 1/k$ , for any positive integer  $k$ , the Borel-Cantelli Lemma shows that the event  $\{|X_n| > 1/k\}$  occurs only a finite number of times, with probability 1. Thus,  $\mathbb{P}(\limsup_{n \rightarrow \infty} |X_n| > 1/k) = 0$ , for every positive integer  $k$ .

Note that the sequence of events  $\{\limsup_{n \rightarrow \infty} |X_n| > 1/k\}$  is monotone and converges to the event  $\{\limsup_{n \rightarrow \infty} |X_n| > 0\}$ . The continuity of probability measures implies that  $\mathbb{P}(\limsup_{n \rightarrow \infty} |X_n| > 0) = 0$ . This establishes that  $X_n \xrightarrow{\text{a.s.}} 0$ .

(iii) This follows since

$$\{\omega : X_n(\omega) \not\rightarrow 0\} = \bigcup_{\epsilon > 0} \bigcap_{n \geq 1} \{\omega : \sup_{m \geq n} |X_m(\omega)| > \epsilon\}$$

□

**Theorem 1:** Let  $X, X_1, X_2, \dots$  be i.i.d. random variables, and assume that  $\mathbb{E}[|X|] < \infty$ . Let  $S_n = X_1 + \dots + X_n$ . Then,  $S_n/n$  converges to a finite constant  $c$  almost surely if and only if  $\mathbb{E}[|X|] < \infty$  and  $c = \mathbb{E}[X]$ .

**Proof of necessity of  $\mathbb{E}[|X|] < \infty$ .** Note that

$$\frac{1}{n} \sum_{k=1}^n a_k \rightarrow 0 \quad \Rightarrow \quad \frac{a_n}{n} \rightarrow 0$$

(just write  $a_n = \sum_{k=1}^n a_k - \sum_{k=1}^{n-1} a_k$ ). Thus, we have

$$\frac{X_n - c}{n} \xrightarrow{\text{a.s.}} 0$$

And by Borel-Cantelli this implies

$$\sum_{n=1}^{\infty} \mathbb{P}[|X - c| > n] < \infty.$$

On the other hand, from  $\mathbb{E}[|Y|] = \int_0^{\infty} \mathbb{P}[|Y| > t] dt$  we derive

$$\mathbb{E}[|X - c|] \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}[|X - c| > n] < \infty,$$

which implies  $\mathbb{E}[|X|] < \infty$ . By what is to be shown, whenever  $\mathbb{E}[|X|] < \infty$  we should have

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{a.s.}} \mathbb{E}[X]$$

which implies  $c = \mathbb{E}[X]$ .

**Proof of convergence, assuming  $\mathbb{E}[X^2] < \infty$ .** We now consider the case where we assume additionally that  $X$  has a finite second moment  $\mathbb{E}[X^2]$ . We have

$$\mathbb{E} \left[ \left( \frac{S_n}{n} - \mu \right)^2 \right] = \frac{\text{var}(X)}{n}.$$

If we only consider values of  $n$  that are perfect squares, we obtain

$$\sum_{i=1}^{\infty} \mathbb{E} \left[ \left( \frac{S_{i^2}}{i^2} - \mu \right)^2 \right] = \sum_{i=1}^{\infty} \frac{\text{var}(X)}{i^2} < \infty,$$

which implies that  $((S_{i^2}/i^2) - \mathbb{E}[X])^2$  converges to zero, with probability 1. Therefore,  $S_{i^2}/i^2$  converges to  $\mathbb{E}[X]$ , with probability 1.

Suppose that the random variables  $X_i$  are nonnegative. Consider some  $n$  such that  $i^2 \leq n < (i+1)^2$ . We then have  $S_{i^2} \leq S_n \leq S_{(i+1)^2}$ . It follows that

$$\frac{S_{i^2}}{(i+1)^2} \leq \frac{S_n}{n} \leq \frac{S_{(i+1)^2}}{i^2},$$

or

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

As  $n \rightarrow \infty$ , we also have  $i \rightarrow \infty$ . Since  $i/(i+1) \rightarrow 1$ , and since  $S_{i^2} \cdot i^2$  converges to  $\mathbb{E}[X]$ , with probability 1, we see that for almost all sample points,  $S_n/n$  is sandwiched between two sequences that converge to  $\mathbb{E}[X]$ . This proves that  $S_n/n \rightarrow \mathbb{E}[X]$ , with probability 1.

For a general random variable  $X$ , we write it in the form  $X = X^+ - X^-$ , where  $X^+$  and  $X^-$  are nonnegative. The strong law applied to  $X^-$  and  $X^+$  separately, implies the strong law for  $X$  as well.  $\square$

**Proof of convergence (general case).** The proof for the most general case (finite mean, but possibly infinite variance) is conceptually simple: We truncate the distribution of  $X$  and apply previous argument to  $Y = X \cdot 1\{|X| < c\}$ , so that the second moment of the latter is finite. Technically, this involves showing that difference  $Y - X$ , although potentially of infinite variance, cannot contribute much to the limiting value. The method is based upon what is called ‘‘maximal ergodic lemma’’, or ‘‘weak- $L_1$ ’’ estimate of the maximal function, see Lemma 1 below.

Without loss of generality we assume  $\mathbb{E}[X] = 0$ . Then by Proposition 1.(iii) it suffices to show for every  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ \sup_{m \geq n} \frac{|S_m|}{m} > \epsilon \right] = 0. \quad (1)$$

To that end, fix  $\epsilon$ , choose (very large)  $c > 0$  and define

$$Y_n = X_n 1\{|X_n| \leq c\} \quad (2)$$

$$Z_n = X_n 1\{|X_n| > c\} \quad (3)$$

$$T_n = \frac{Y_1 + \cdots + Y_n}{n} \quad (4)$$

$$Z^* = \sup_{n \geq 1} \frac{Z_1 + \cdots + Z_n}{n} \quad (5)$$

Note that  $Y_j$  and  $Z_j$  are iid. By dominated convergence theorem, as  $c \rightarrow \infty$  we have  $\mathbb{E}[Y] \rightarrow 0$  and  $\mathbb{E}[|Z|] \rightarrow 0$ . Therefore for any  $\delta$  it is possible to find  $c > 0$  such that

$$|\mathbb{E}[Y]| \leq \frac{\epsilon}{4} \quad (6)$$

$$\mathbb{E}[|Z|] \leq \delta \frac{\epsilon}{2} \quad (7)$$

By the proof above, we have

$$T_n \xrightarrow{\text{a.s.}} \mathbb{E}[Y] \quad (8)$$

and therefore we have

$$\mathbb{P} \left[ \sup_{m \geq n} \frac{|S_m|}{m} > \epsilon \right] \leq \mathbb{P} \left[ \sup_{m \geq n} \frac{|T_m|}{m} + |Z^*| > \epsilon \right] \quad (9)$$

$$\leq \mathbb{P} \left[ \sup_{m \geq n} \frac{|T_m|}{m} > \frac{\epsilon}{2} \right] + \mathbb{P} \left[ |Z^*| > \frac{\epsilon}{2} \right] \quad (10)$$

$$\leq \mathbb{P} \left[ \sup_{m \geq n} \frac{|T_m - \mathbb{E}[Y]|}{m} > \frac{\epsilon}{4} \right] + \mathbb{P} \left[ |Z^*| > \frac{\epsilon}{2} \right] \quad (11)$$

where (9) is because  $|S_m - T_m| \leq Z^*$ , (10) follows from the union-bound applied to non-negative  $A, B$ :

$$\mathbb{P}[A + B > 2\epsilon] \leq \mathbb{P}[A > \epsilon] + \mathbb{P}[B > \epsilon]$$

and (11) is because of (6) and  $|T_m| \leq |T_m - \mathbb{E}[Y]| + \mathbb{E}[Y]$ .

Taking limit of (11) as  $n \rightarrow \infty$  the first term disappears due to (8) and Proposition 1.(iii). The Lemma 1 to follow bounds the second term as

$$\mathbb{P} \left[ |Z^*| > \frac{\epsilon}{2} \right] \leq \frac{2\mathbb{E}[|Z|]}{\epsilon} \leq \delta.$$

Alltogether, we have shown

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left[ \sup_{m \geq n} \frac{|S_m|}{m} > \epsilon \right] \leq \delta$$

for every  $\delta > 0$ , which proves (1) and the Theorem.  $\square$

**Lemma 1** (Estimate for the maximum of averages). *Let  $Z_n$  be iid with  $\mathbb{E}[|Z|] < \infty$  then*

$$\mathbb{P} \left[ \sup_{n \geq 1} \frac{|Z_1 + \dots + Z_n|}{n} > a \right] \leq \frac{\mathbb{E}[|Z|]}{a} \quad \forall a > 0$$

**Proof.** The argument for this Lemma has originally been quite involved, until a dramatically simple proof (below) was found by A. Garcia. We note that the result applies to arbitrary stationary process  $\{Z_n, n = 1, \dots\}$ , although we only need an iid version here.

Define

$$S_n = \sum_{k=1}^n Z_k \tag{12}$$

$$L_n = \max\{0, Z_1, \dots, Z_1 + \dots + Z_n\} \tag{13}$$

$$M_n = \max\{0, Z_2, Z_2 + Z_3, \dots, Z_2 + \dots + Z_n\} \tag{14}$$

$$Z^* = \sup_{n \geq 1} \frac{S_n}{n} \tag{15}$$

It is sufficient to show that

$$\mathbb{E}[Z_1 1_{\{Z^* > 0\}}] \geq 0. \tag{16}$$

Indeed, applying (16) to  $\tilde{Z}_1 = Z_1 - a$  and noticing that  $\tilde{Z}^* = Z^* - a$  we obtain

$$\mathbb{E}[Z_1 1_{\{Z^* > a\}}] \geq a \mathbb{P}[Z^* > a],$$

from which Lemma follows by upper-bounding the left-hand side with  $\mathbb{E}[|Z_1|]$ .

In order to show (16) we first notice that  $\{L_n > 0\} \nearrow \{Z^* > 0\}$ . Next we notice that

$$Z_1 + M_n = \max\{S_1, \dots, S_n\}$$

and furthermore

$$Z_1 + M_n = L_n \quad \text{on } \{L_n > 0\}$$

Thus, we have

$$Z_1 1_{\{L_n > 0\}} = L_n - M_n 1_{\{L_n > 0\}}$$

where we do not need indicator in the first term since  $L_n = 0$  on  $\{L_n > 0\}^c$ . Taking expectation we get

$$\mathbb{E}[Z_1 1_{\{L_n > 0\}}] = \mathbb{E}[L_n] - \mathbb{E}[M_n 1_{\{L_n > 0\}}] \quad (17)$$

$$\geq \mathbb{E}[L_n] - \mathbb{E}[M_n] \quad (18)$$

$$= \mathbb{E}[L_n] - \mathbb{E}[L_{n-1}] = \mathbb{E}[L_n - L_{n-1}] \geq 0, \quad (19)$$

where we used  $M_n \geq 0$ , the fact that  $M_n$  has the same distribution as  $L_{n-1}$ , and  $L_n \geq L_{n-1}$ , respectively. Taking limit as  $n \rightarrow \infty$  in (19) we obtain (16).  $\square$

## 2 THE CHERNOFF BOUND

Let again  $X, X_1, \dots$  be i.i.d., and  $S_n = X_1 + \dots + X_n$ . Let us assume, for simplicity, that  $\mathbb{E}[X] = 0$ . According to the weak law of large numbers, we know that  $\mathbb{P}(S_n \geq na) \rightarrow 0$ , for every  $a > 0$ . We are interested in a more detailed estimate of  $\mathbb{P}(S_n \geq na)$ , involving the rate at which this probability converges to zero. It turns out that if the moment generating function of  $X$  is finite on some interval  $[0, c]$  (where  $c > 0$ ), then  $\mathbb{P}(S_n \geq na)$  decays exponentially with  $n$ , and much is known about the precise rate of exponential decay.

### 2.1 Upper bound

Let  $M(s) = \mathbb{E}[e^{sX}]$ , and assume that  $M(s) < \infty$ , for  $s \in [0, c]$ , where  $c > 0$ . Recall that  $M_{S_n}(s) = \mathbb{E}[e^{s(X_1 + \dots + X_n)}] = (M(s))^n$ . For any  $s > 0$ , the Markov inequality yields

$$\mathbb{P}(S_n \geq na) = \mathbb{P}(e^{sS_n} \geq e^{nsa}) \leq e^{-nsa} \mathbb{E}[e^{sS_n}] = e^{-nsa} (M(s))^n.$$

Every nonnegative value of  $s$ , gives us a particular bound on  $\mathbb{P}(S_n \geq a)$ . To obtain the tightest possible bound, we minimize over  $s$ , and obtain the following result.

**Theorem 2. (Chernoff upper bound)** Suppose that  $\mathbb{E}[e^{sX}] < \infty$  for some  $s > 0$ , and that  $a > 0$ . Then,

$$\mathbb{P}(S_n \geq na) \leq e^{-n\phi(a)},$$

where

$$\phi(a) = \sup_{s \geq 0} (sa - \log M(s)).$$

For  $s = 0$ , we have

$$sa - \log M(s) = 0 - \log 1 = 0,$$

where we have used the generic property  $M(0) = 1$ . Furthermore,

$$\frac{d}{ds}(sa - \log M(s)) \Big|_{s=0} = a - \frac{1}{M(s)} \cdot \frac{d}{ds}M(s) \Big|_{s=0} = a - 1 \cdot \mathbb{E}[X] > 0.$$

Since the function  $sa - \log M(s)$  is zero and has a positive derivative at  $s = 0$ , it must be positive when  $s$  is positive and small. It follows that the supremum  $\phi(a)$  of the function  $sa - \log M(s)$  over all  $s \geq 0$  is also positive. In particular, for any fixed  $a > 0$ , the probability  $\mathbb{P}(S_n \geq na)$  decays at least exponentially fast with  $n$ .

**Example:** For a standard normal random variable  $X$ , we have  $M(s) = e^{s^2/2}$ . Therefore,  $sa - \log M(s) = sa - s^2/2$ . To maximize this expression over all  $s \geq 0$ , we form the derivative, which is  $a - s$ , and set it to zero, resulting in  $s = a$ . Thus,  $\phi(a) = a^2/2$ , which leads to the bound

$$\mathbb{P}(X \geq na) \leq e^{-a^2n/2}.$$

## 2.2 Lower bound

Remarkably, it turns out that the estimate  $\phi(a)$  of the decay rate is tight, under minimal assumptions. To keep the argument simple, we introduce some simplifying assumptions.

### Assumption 1.

- (i)  $M(s) = \mathbb{E}[e^{sX}] < \infty$ , for all  $s \in \mathbb{R}$ .
- (ii) The random variable  $X$  is continuous, with PDF  $f_X$ .
- (iii) The random variable  $X$  does not admit finite upper and lower bounds. (Formally,  $0 < F_X(x) < 1$ , for all  $x \in \mathbb{R}$ .)

We then have the following lower bound.

**Theorem 2. (Chernoff lower bound)** Under Assumption 1, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -\phi(a), \quad (20)$$

for every  $a > 0$ .

We note two consequences of our assumptions, whose proof is left as an exercise:

- (a)  $\lim_{s \rightarrow \infty} \frac{\log M(s)}{s} = \infty$ ;
- (b)  $M(s)$  is differentiable at every  $s$ .

The first property guarantees that for any  $a > 0$  we have  $\lim_{s \rightarrow \infty} (\log M(s) - sa) = \infty$ . Since  $M(s) > 0$  for all  $s$ , and since  $M(s)$  is differentiable, it follows that  $\log M(s)$  is also differentiable and that there exists some  $s^* \geq 0$  at which  $\log M(s) - sa$  is minimized over all  $s \geq 0$ . Taking derivatives, we see that such a  $s^*$  satisfies  $a = M'(s^*)/M(s^*)$ , where  $M'$  stands for the derivative of  $M$ . In particular,

$$\phi(a) = s^* a - \log M(s^*). \quad (21)$$

Let us introduce a new PDF

$$f_Y(x) = \frac{e^{s^* x}}{M(s^*)} f_X(x).$$

This is a legitimate PDF because

$$\int f_Y(x) dx = \frac{1}{M(s^*)} \int e^{s^* x} f_X(x) dx = \frac{1}{M(s^*)} \cdot M(s^*) = 1.$$

The moment generating function associated with the new PDF is

$$M_Y(s) = \frac{1}{M(s^*)} \int e^{sx} e^{s^* x} f_X(x) dx = \frac{M(s + s^*)}{M(s^*)}.$$

Thus,

$$\mathbb{E}[Y] = \frac{1}{M(s^*)} \cdot \frac{d}{ds} M(s + s^*) \Big|_{s=0} = \frac{M'(s^*)}{M(s^*)} = a,$$

where the last equality follows from our definition of  $s^*$ . The distribution of  $Y$  is called a “tilted” version of the distribution of  $X$ .

Let  $Y_1, \dots, Y_n$  be i.i.d. random variables with PDF  $f_Y$ . Because of the close relation between  $f_X$  and  $f_Y$ , approximate probabilities of events involving  $Y_1, \dots, Y_n$  can be used to obtain approximate probabilities of events involving  $X_1, \dots, X_n$ .

We keep assuming that  $a > 0$ , and fix some  $\delta > 0$ . Let

$$B = \left\{ (x_1, \dots, x_n) \mid a - \delta \leq \frac{1}{n} \sum_{i=1}^n x_i \leq a + \delta \right\} \subset \mathbb{R}^n.$$



Let  $S_n = X_1 + \dots + X_n$  and  $T_n = Y_1 + \dots + Y_n$ . We have

$$\begin{aligned}
\mathbb{P}(S_n \geq n(a - \delta)) &\geq \mathbb{P}(n(a - \delta) \leq S_n \leq n(a + \delta)) \\
&= \int_{(x_1, \dots, x_n) \in B} f_X(x_1) \cdots f_X(x_n) dx_1 \cdots dx_n \\
&= \int_{(x_1, \dots, x_n) \in B} (M(s^*))^n e^{-s^* x_1} f_Y(x_1) \cdots e^{-s^* x_n} f_Y(x_n) dx_1 \cdots dx_n \\
&\geq (M(s^*))^n e^{-ns^*(a+\delta)} \int_{(x_1, \dots, x_n) \in B} f_Y(x_1) \cdots f_Y(x_n) dx_1 \cdots dx_n \\
&= (M(s^*))^n e^{-ns^*(a+\delta)} \mathbb{P}(T_n \in B). \tag{22}
\end{aligned}$$

The second inequality above was obtained because for every  $(x_1, \dots, x_n) \in B$ , we have  $x_1 + \dots + x_n \leq n(a + \delta)$ , so that  $e^{-s^* x_1} \cdots e^{-s^* x_n} \geq e^{-ns^*(a+\delta)}$ .

By the weak law of large numbers, we have

$$\mathbb{P}(T_n \in B) = \mathbb{P}\left(\frac{Y_1 + \dots + Y_n}{n} \in [na - n\delta, na + n\delta]\right) \rightarrow 1,$$

as  $n \rightarrow \infty$ . Taking logarithms, dividing by  $n$ , and then taking the limit of the two sides of Eq. (22), and finally using Eq. (21), we obtain

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na - n\delta) \geq \log M(s^*) - s^* a - s^* \delta = -\phi(a) - s^* \delta.$$

This inequality is true for every  $a > 0$  and  $\delta > 0$ . By replacing  $a$  with  $a + \delta$ , we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) \geq -\phi(a + \delta) - s^* \delta.$$

The proof of the lower bound in Eq. (20) is completed by verifying that the function  $\phi$  is continuous (the proof is omitted and is left as an exercise) and letting  $\delta \downarrow 0$ .

MIT OpenCourseWare  
<https://ocw.mit.edu>

6.436J / 15.085J Fundamentals of Probability  
Fall 2018

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>