Law of Large numbers and Borel Cantelli Lemmas

5.1 Law of Large Numbers (LLN)

Theorem 5.1 ($\mathbf{L^2}$ – Law of Large Numbers). *Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}\, X_1 = \mu$ and $\mathbb{E}\, X_1^2 < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbf{L^2}} \mu \text{ as } n \to \infty.$$

Proof Sketch: We have

$$\left\| \frac{1}{n}\sum_{i=1}^{n} X_i - \mu \right\|_2^2 = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) = \frac{1}{n}\operatorname{Var}(X_1) \to 0.$$

■

For the Weak Law of Large Numbers (WLLN) introduced below, no assumption on the $\mathbb{E}\, X^2$ is needed.

Theorem 5.2 (Weak Law of Large Numbers – WLLN). *Let $X_1, X_2, \dots$ be i.i.d. with $\mathbb{E}\, X_1 = \mu$ is finite. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mu \text{ as } n \to \infty.$$

The convergence also holds in L^1.

Proof. Note that, it is enough to prove L^1-convergence, *i.e.*, $\left\| \frac{1}{n}\sum_{i=1}^{n} X_i - \mu \right\|_1 \to 0$ as L^1 convergence implies convergence in probability.

Fix $x > 0$. Define, $Y_i := X_i \mathbb{1}_{|X_i| < x}$ for $i \geqslant 1$. We can write $X_i = X_i \mathbb{1}_{|X_i| \geqslant x} + Y_i$. Note that, by DCT we have, $\mathbb{E}\left(|X_1| \mathbb{1}_{|X_1| > x}\right) \to 0$ as $x \uparrow \infty$. The main idea behind the proof is to truncate the r.v.s, so that the bounded part converges using L^2 law of large numbers and the unbounded part has small expectation.

We have, for any $x > 0$

$$\begin{aligned}
\left\| \frac{1}{n}\sum_{i=1}^{n} X_i - \mu \right\|_1 &\leqslant \mathbb{E}\left| \frac{1}{n}\sum_{i=1}^{n} X_i \mathbb{1}_{|X_i| \geqslant x} - \mathbb{E}(X_1 \mathbb{1}_{|X_1| \geqslant x}) \right| + \mathbb{E}\left| \frac{1}{n}\sum_{i=1}^{n} Y_i - \mathbb{E}\, Y_1 \right| \\
&\leqslant 2\,\mathbb{E}(|X_1| \mathbb{1}_{|X_1| \geqslant x}) + \sqrt{\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)} \\
&= 2\,\mathbb{E}(|X_1| \mathbb{1}_{|X_1| \geqslant x}) + \sqrt{n^{-1}\operatorname{Var}(Y_1)} \leqslant 2\,\mathbb{E}(|X_1| \mathbb{1}_{|X_1| \geqslant x}) + n^{-1/2}\left\| X_1 \mathbb{1}_{|X_1| < x} \right\|_2 .
\end{aligned}$$

Taking $x = n^{1/4}$ and letting $n \to \infty$, we get

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right\|_1 = 0.$$

Note that, to prove convergence in Probability it was enough to prove that $\mathbb{P}(\max_{1\leqslant i\leqslant n} |X_i| > x) \to 0$ for any $x > 0$ satisfying $\mathrm{Var}(X_1 \mathbb{1}_{|X_1|>x}) \ll n$. ■

Exercise 5.3. *Show that $\mathcal{X}$, the set of all random variables, is metrizable under convergence in probability, that is, $\exists$ a metric $\boldsymbol{d}$ on $\mathcal{X}$ such that $X_n \xrightarrow{\mathbb{P}} X$ if and only if $\boldsymbol{d}(X_n, X) \to 0$ and $\mathcal{X}$ is complete under the metric.*

Hint: *There are many choices:*

$$\begin{aligned}
\boldsymbol{d}(X,Y) &= \mathbb{E}\left(|X-Y| \wedge 1\right)\\
\boldsymbol{d}(X,Y) &= \mathbb{E}\left(|X-Y|/(1+|X-Y|)\right)\\
\boldsymbol{d}(X,Y) &= \inf\{\mathbb{P}(|X-Y| > \varepsilon) + \varepsilon \mid \varepsilon > 0\}.
\end{aligned}$$

We can strengthen the conclusion of WLLN without any extra assumption.

Theorem 5.4 (Strong Law of Large Numbers – SLLN). *Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}\, X_1 = \mu$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu.$$

Example 5.5. *Let $X_1, X_2, \ldots \sim$ Uniform$([-1,1])$ be i.i.d. r.v. with mean zero. Note $\mathbb{E}\, X_i^2 = \int_{-1}^{1} x^2 \cdot \frac{1}{2} dx = \frac{1}{3}$. So by the SLLN, $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{a.s.} \frac{1}{3}$. Now $(X_1, \ldots, X_n) \sim$ Uniform$([-1,1]^n)$. So an n-dimensional vector with uniform coordinates approximately lies on a sphere with radius $\sqrt{\frac{n}{3}}$ with high probability.*

We need certain results to obtain a.s. convergence from probability estimates.

5.2 Borel Cantelli Lemmas

Definition 5.6. *Let $A_1, A_2, \ldots$ be an infinite sequence of events. We define*

$$\begin{aligned}
\limsup_n A_n = \cap_{m\geqslant 1} \cup_{n\geqslant m} A_n = \lim_{m\to\infty} \cup_{n\geqslant m} A_n &= \{A_n \text{ infinitely often (i.o.)}\}\\
&= \{\omega \text{ that are in infinitely many } A_n\}.
\end{aligned}$$

We have the following:

- $\limsup_n \mathbb{1}_{A_n} = \mathbb{1}_{\limsup_n A_n}$,
- $\liminf_n A_n = \lim_{m\uparrow\infty} \cap_{n\geqslant m} A_n = \cup_{m\geqslant 1} \cap_{n\geqslant m} A_n = \{A_n \text{ eventually (ev.)}\} = \{\omega$ that are in all but finitely many $A_n\}$,

- $\liminf_n \mathbb{1}_{A_n} = \mathbb{1}_{\liminf_n A_n}$,
- $(\limsup_n A_n)^c = \liminf_n A_n^c$,
- $\liminf_n A_n \subseteq \limsup_n A_n$.

Now we will state and prove the Borel Cantelli Lemmas:

Lemma 5.7 (Borel–Cantelli Lemmas). *Let $A_1, A_2, \ldots$ be a sequence of events.*

(i) *If $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$, then*

$$\mathbb{P}(A_n \text{ i.o.}) = 0.$$

(ii) *If $A_1, A_2, \ldots$ are independent and $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$, then*

$$\mathbb{P}(A_n \text{ i.o.}) = 1.$$

Proof. (i) $\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(\lim_m \cup_{n\geqslant m} A_n) = \lim_m \mathbb{P}(\cup_{n\geqslant m} A_m)$ with $\cup_{n\geqslant m} A_m$ decreasing in m. This implies

$$\lim_m \mathbb{P}(\cup_{n\geqslant m} A_n) \leqslant \lim_m \sum_{n\geqslant m} \mathbb{P}(A_n) = 0.$$

(ii) We have

$$\begin{aligned} \mathbb{P}(A_n \text{ i.o.})^c = \mathbb{P}(\lim_m \cap_{n\geqslant m} A_n^c) &= \lim_m \mathbb{P}(\cap_{n\geqslant m} A_n^c) \\ &= \lim_m \prod_{n\geqslant m} \mathbb{P}(A_n^c) = \lim_m \prod_{n\geqslant m} (1 - \mathbb{P}(A_n)) \leqslant \lim_m e^{-\sum_{n\geqslant m} \mathbb{P}(A_n)} = 0, \end{aligned}$$

since $(1-x) \leqslant e^{-x}$ and $\sum_{n\geqslant m} \mathbb{P}(A_n) = \infty$ for all $m \geqslant 1$. ■

Remark 5.8. *The assumption of independence in part ii) of Borel-Cantelli Lemma can be replaced by **pairwise independence**. See [Durrett] for the proof.*

Lemma 5.9. *The following are equivalent:*

(i) *$X_n \to X$ a.s.*

(ii) *For every $\varepsilon > 0$ we have $\mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 0$.*

(iii) *There exists $\varepsilon_n \downarrow 0$ such that $\mathbb{P}(|X_n - X| > \varepsilon_n \text{ i.o.}) = 0$.*

(iv) *For $M_n := \max_{k\geqslant n} |X_k - X|$, we have $M_n \to 0$ in probability.*

Proof. i) $\implies$ ii): $\mathbb{P}(\lim_{n\to\infty} |X_n - X| = 0) = 1 = \mathbb{P}(\forall\varepsilon > 0, |X_n - X| \leqslant \varepsilon \text{ ev.})$. This implies $\forall\varepsilon > 0, \mathbb{P}(|X_n - X| \leqslant \varepsilon \text{ ev.}) = 1$. We have the result by taking complement. ■

Exercise 5.10. *Prove the equivalence in Lemma **??**.*

In v) of Lemma **??**, we cannot replace $\exists$ with $\forall$, illustrated by the following example.

Example 5.11. *Let $X_n = U_n/n$, where $U_1, U_2, \ldots$ are i.i.d. Uniform$(-1,1)$ rvs. Then we have*

$$\mathbb{P}(|X_n| > 1/2n \text{ i.o.}) = 1$$

and $X_n \xrightarrow{a.s.} 0$. For the proof, let $A_n = \{|X_n| > 1/2n\}, n \geqslant 1$. Then $\mathbb{P}(A_n) = \frac{1}{2} \implies \sum \mathbb{P}(A_n) = \infty$ and A_n's are independent. Therefore, by second Borel-Cantelli Lemma, $\mathbb{P}(A_n \text{ i.o.}) = 1$ or $\mathbb{P}(|X_n| > 1/2n \text{ i.o.}) = 1$.

5.3 Random Walk on $\mathbb{Z}^d$

In this section we consider random walks in d-dimensions, $\mathbb{Z}^d$. Let $X_1, X_2, \ldots$ be i.i.d. $\mathbb{Z}^d$-valued random vector such that

$$\mathbb{P}(X_1 = \boldsymbol{e}_i) = \mathbb{P}(X_1 = -\boldsymbol{e}_i) = \frac{1}{2d}, \text{ for } i = 1, 2, \ldots, d$$

where $\boldsymbol{e}_i$ is the unit vector in the i-th coordinate direction. The simple symmetric random walk (SSRW) S_n is defined as

$$S_0 = 0,$$
$$S_n = X_1 + X_2 + \cdots + X_n, \ n \geqslant 1.$$

Theorem 5.12. *For the simple symmetric random walk on $\mathbb{Z}^d$ we have,*

$$\mathbb{P}(S_n = 0 \ i.o.) = \begin{cases} 1 \ \text{if } d = 1, 2 \\ 0 \ \text{if } d \geqslant 3. \end{cases}$$

Example 5.13 (Random Walk on $\mathbb{Z}$)**.** *Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{P}(X_i = -1) = \frac{1}{2} = \mathbb{P}(X_i = 1)$ for all $i \in \mathbb{N}$. Define*

$$S_n = X_1 + X_2 + \cdots + X_n$$

for $n \in \mathbb{N}$. It can be shown that $\mathbb{P}(S_n = 0 \ i.o.) = 1$. Notice that

$$\mathbb{P}(S_{2n} = 0) = \binom{2n}{n}\left(\frac{1}{2}\right)^n$$

for any n. By Stirling's approximation : $n! \sim (2\pi)^{\frac{1}{2}} e^{-n} n^{n+\frac{1}{2}}$,

$$\mathbb{P}(S_{2n} = 0) = \frac{2n!}{(n!)^2}\frac{1}{2^n} \sim \frac{(2\pi)^{\frac{1}{2}} e^{-n} 2n^{2n+\frac{1}{2}}}{2\pi e^{-2n} n^{2n+1} 2^{2n}} = \frac{1}{\sqrt{\pi n}}.$$

So $\sum_{n\geqslant 1} \mathbb{P}(S_n = 0) = \infty$. However

$$\mathbb{P}(S_{2n} = 0, S_{2n+2} = 0) = \mathbb{P}(S_{2n} = 0) \times \frac{2}{2^2} \neq \mathbb{P}(S_{2n} = 0)\,\mathbb{P}(S_{2n+2} = 0).$$

So Borel-Cantelli Lemma cannot be applied to show $\mathbb{P}(S_n = 0 \ i.o.) = 1$.

Example 5.14 (Another Random Walk on $\mathbb{Z}^d$)**.** *Let $\tilde{X}_i = (x_1, x_2, \ldots, x_d)$ with probability $\frac{1}{2^d}$ where each $x_i \in \{1, -1\}$. Then, for each n, define*

$$S_n = \tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_n.$$

Analogous to random walk on $\mathbb{Z}$, we are interested in calculating $\mathbb{P}(S_{2n} = 0)$. Write

$$S_n = (S_n^1, \ldots, S_n^d)$$

where $(S_n^1)_{n\geqslant 1}, \ldots, (S_n^d)_{n\geqslant 1}$ are i.i.d. 1-dimensional random walk. Observe that

$$\mathbb{P}(S_{2n} = 0) = \mathbb{P}(S_{2n}^1 = 0, \ldots, S_{2n}^d = 0) \sim \frac{1}{(n\pi)^{\frac{d}{2}}}.$$

Then $\sum_{n\geqslant 1} \mathbb{P}(S_{2n} = 0) < \infty$ if and only if $d \geqslant 3$. In particular, for $d \geqslant 3$, $\mathbb{P}(S_{2n} = 0 \ i.o.) = 0$.

5.4 Proof of Strong Law of Large Number

Theorem 5.15 (Strong Law of Large number). *Let $X_1, X_2, \dots$ be an i.i.d. sequence of random variables with $\mathbb{E}\, X_1 = \mu$ finite, then*

$$n^{-1} \sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu.$$

Proof Sketch: WLOG, assume $X_i \geqslant 0$. (For general random variables X_i, we can write $X_i = X_i^+ - X_i^-$ and apply the case when each $X_i \geqslant 0$ separately to $(X_i^+)_{i\geqslant 1}$ and $(X_i^-)_{i\geqslant 1}$.) We will use the following steps:

1. Truncation (inhomogeneous)
2. Error analysis for truncation
3. Subsequence argument.

Let

$$Y_k := X_k \mathbb{1}_{X_k \leqslant k}.$$

Notice that if we set $T_n = \sum_{i=1}^n Y_i$, then it suffices to show that $\frac{T_n}{n} \xrightarrow{a.s.} \mu$ because of the following lemma.

Lemma 5.16. $\mathbb{P}(X_k \neq Y_k \text{ } i.o.) = 0$.

Proof. We will apply Borel-Cantelli Lemma (i) to prove this result. So it is enough to show that $\sum_k \mathbb{P}(X_k \neq Y_k) < \infty$. Notice that $\mathbb{P}(X_k \neq Y_k) = \mathbb{P}(X > k)$. So

$$\sum_k \mathbb{P}(X_k \neq Y_k) = \sum_{k \geqslant 1} \mathbb{P}(X > k) = \mathbb{E}(\sum_{k\geqslant 1} \mathbb{1}_{X>k}) \leqslant \mathbb{E}\, X < \infty$$

where the last inequality follows by $0 \leqslant \sum_{k\geqslant 1} \mathbb{1}_{X>k} \leqslant X$. ■

In particular, $n^{-1}\sum_{i=1}^n X_i - n^{-1}\sum_{i=1}^n Y_i \xrightarrow{a.s.} 0$ because $\left|\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i\right| \leqslant \sum_{i=1}^\infty |X_i - Y_i|$, a finite random variable.

Now recall that $T_n = \sum_{i=1}^n Y_i$. Take $Z_n = \frac{T_n - \mathbb{E}\, T_n}{n}$. By Chebyshev's inequality we have

$$\mathbb{P}(|Z_n| > \varepsilon) \leqslant \frac{\mathrm{Var}(T_n)}{n^2\varepsilon^2} = \frac{\sum_{i=1}^n \mathrm{Var}(Y_i)}{n^2\varepsilon^2} \leqslant \frac{\sum_{i=1}^n \mathbb{E}\, Y_i^2}{n^2\varepsilon^2} = \frac{1}{n^2\varepsilon^2}\sum_{i=1}^n \mathbb{E}\, X_1^2 \mathbb{1}_{X_1 \leqslant i} \leqslant \frac{1}{n\varepsilon^2}\mathbb{E}\, X_1^2 \mathbb{1}_{X_1\leqslant n}.$$

In general, for any subsequences (Z_{n_k}), we have

$$\sum_{k\geqslant 1} \mathbb{P}(|Z_{n_k}| > \varepsilon) \leqslant \sum_{k=1}^\infty \frac{1}{n_k \varepsilon^2}\mathbb{E}\, X_1^2 \mathbb{1}_{X_1 \leqslant n_k} \leqslant \frac{1}{\varepsilon^2}\mathbb{E}\left(X_1^2 \cdot \sum_{k=1}^\infty \frac{\mathbb{1}_{n_k \geqslant X_1}}{n_k}\right).$$

We want a sequence n_k such that

$$\sum_{k=1}^\infty \frac{\mathbb{1}_{n_k \geqslant x}}{n_k} \approx \frac{1}{x}$$

Geometric series satisfies this criteria. So we take $n_k = \lceil \alpha^k \rceil$ for some $\alpha > 1$. Then, for all $x > 0$ we have

$$\sum_{k:n_k \geqslant x} \frac{1}{n_k} \leqslant \frac{c(\alpha)}{x}$$

for some finite constant $c(\alpha)$ depending only on α. So

$$\mathbb{E}\left(X_1^2 \cdot \sum_{k=1}^{\infty} \frac{\mathbb{1}_{n_k \geqslant X_1}}{n_k} \right) \leqslant c(\alpha)\, \mathbb{E}\, X_1 < \infty.$$

In particular, $Z_{n_k} \xrightarrow{a.s.} 0$ where $n_k = \lceil \alpha^k \rceil$, or $\frac{T_{n_k}}{n_k} - \frac{\mathbb{E}\, T_{n_k}}{n_k} \to 0$ as $k \uparrow \infty$. Now,

$$\frac{\mathbb{E}\, T_n}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\, X_1 \mathbb{1}_{X_1 \leqslant i} \to \mathbb{E}\, X_1 \text{ as } n \to \infty \text{ since } \mathbb{E}\, X_1 \mathbb{1}_{X_1 \leqslant i} \to \mathbb{E}\, X_1 \text{ as } i \to \infty$$

or

$$\frac{T_{n_k}}{n_k} \xrightarrow{a.s.} \mu.$$

Notice that $T_{n_k} \leqslant T_n \leqslant T_{n_{k+1}}$ whenever $n_k < n \leqslant n_{k+1}$, or

$$\frac{T_{n_k}}{n_{k+1}} \leqslant \frac{T_n}{n} \leqslant \frac{T_{n_{k+1}}}{n_k}.$$

So

$$\frac{\mu}{\alpha} \leqslant \liminf_k \frac{T_{n_k}}{n_{k+1}} \leqslant \liminf_n \frac{T_n}{n} \leqslant \limsup_n \frac{T_n}{n} \leqslant \limsup_k \frac{T_{n_{k+1}}}{n_k} \leqslant \mu\alpha.$$

Since $\alpha > 1$ is arbitrary,

$$\limsup_n \frac{T_n}{n} = \liminf_n \frac{T_n}{n} = \mu \text{ or } \frac{T_n}{n} \xrightarrow{a.s.} \mu.$$

■