8. Glivenko–Cantelli and 0-1 Laws

DASN, NTU https://personal.ntu.edu.sg/wptay/

8.1 Glivenko–Cantelli Theorem

Assume that $F(x)$ is a cumulative distribution function (cdf), i.e.,

- $F(x)$ is non-decreasing.
- $F(x)$ is right continuous.
- $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$.

A r.v. X such that $\mathbb{P}(X \leq x) = F(x)$ is said to have cdf F. From Carathéodory's Extension Theorem (Theorem 3.1), the cdf of X uniquely determines the distribution $\mathbb{P}_X$ of X.

Then even though $F(x)$ may not be a bijective function, we can always define a pseudo-inverse as

$$F^{-1}(\omega) = \inf\{x : F(x) \geq \omega\} \tag{1}$$

for $\omega \in [0, 1]$. We have the following properties:

(i) If $y < F^{-1}(\omega)$, then $F(y) < \omega$. Otherwise, suppose $F(y) \geq \omega$. Then by the definition (1), $F^{-1}(\omega) \leq y$, which is a contradiction.

(ii) If $y > F^{-1}(\omega)$, then $F(y) \geq \omega$ because F is non-decreasing. If ω is a continuity point of F^{-1}, then we have $F(y) > \omega$.

Lemma 8.1. *For any x, we have*

$$\{\omega : \omega \leq F(x)\} = \{\omega : F^{-1}(\omega) \leq x\}.$$

Proof. If $\omega \leq F(x)$, then by definition (1), we have $F^{-1}(\omega) \leq x$. On the other hand, if $F^{-1}(\omega) \leq x$, then since F is non-decreasing, we have $\omega \leq F(y)\ \forall\, y > x$. Since $F(x)$ is right continuous, by letting $y \downarrow x$, we have $\omega \leq F(x)$. □

Lemma 8.2. *$F^{-1}(\omega)$, $\omega \in [0, 1]$, is a random variable on $([0, 1], \mathcal{B}([0, 1]), \lambda)$ with cdf F, where λ is the Lebesgue measure.*

Proof. From Lemma 8.1, we have $\lambda(F^{-1}(\omega) \leq x) = \lambda(\omega \leq F(x)) = F(x)$. □

Suppose that (X_i) are i.i.d. random variables with cdf F, then we define the empirical cdf for $n \geq 1$ as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}.$$

The SLLN implies that for each x, $F_n(x) \to F(x)$ a. s. as $n \to \infty$.

Theorem 8.1 (Glivenko-Cantelli Theorem)**.**

$$\sup_x |F_n(x) - F(x)| \to 0 \text{ a. s. } \textit{as } n \to \infty.$$

Proof. We reduce the proof to that for the probability space $([0,1], \mathcal{B}[0,1], \lambda)$. Let Λ to be the cdf for λ and suppose Y_i are i.i.d. $\sim \lambda$. We define

$$\Lambda_n(F(x)) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq F(x)\}}.$$

If $X_i = F^{-1}(Y_i)$, then from Lemma 8.2, X_i are i.i.d. with cdf F. Therefore, $X_i \leq x$ iff $Y_i \leq F(x)$, and to prove the theorem, it suffices to show that

$$\sup_y |\Lambda_n(y) - \Lambda(y)| \to 0 \text{ a. s.}$$

Fix $\epsilon > 0$, choose m s.t. $\dfrac{1}{m} \leq \dfrac{\epsilon}{2}$. Consider the set

$$E = \left\{ \frac{k}{m} \,:\, k = 0, 1, \ldots, m \right\}.$$

From SLLN, $\Lambda_n(y) \to \Lambda(y)$ for each $y \in E$. Since E is finite, $\exists\, N$ s.t. $\forall\, n \geq N$, $|\Lambda_n(y) - \Lambda(y)| \leq \frac{\epsilon}{2}$ $\forall\, y \in E$. For $x \in [0,1]$, we can always find u, v s.t. $u \leq x < v$, $u, v \in E$, $v - u = \frac{1}{m}$. Since $\Lambda(y) = y$, we have

$$\Lambda_n(x) \geq \Lambda_n(u) \geq u - \frac{\epsilon}{2} \geq x - \frac{1}{m} - \frac{\epsilon}{2} \geq x - \epsilon,$$
$$\Lambda_n(x) \leq \Lambda_n(v) \leq v + \frac{\epsilon}{2} \leq x + \frac{1}{m} + \frac{\epsilon}{2} \leq x + \epsilon,$$

so that

$$|\Lambda_n(x) - \Lambda(x)| \leq \epsilon.$$

The theorem is now proved. $\square$

8.2 Kolmogorov's 0-1 Law

Suppose that $(X_i)_{i\geq 1}$ are independent random variables. Consider the σ-algebra $\sigma(X_n, X_{n+1}, \ldots)$, which is the smallest σ-algebra w.r.t. which all X_m, $m \geq n$, are measurable:

$$\sigma(X_n, X_{n+1}, \ldots) = \sigma\left(\bigcup_{m\geq n} \sigma(X_m)\right).$$

Note that the intersection of these σ-algebras is also a σ-algebra,

$$\mathcal{T} = \bigcap_{n\geq 1} \sigma(X_n, X_{n+1}, \ldots),$$

and we call this the tail σ-algebra. The tail σ-algebra contains all events that are not affected by changes in a finite number of random variables X_i.

Let $S_n = \sum_{i=1}^n X_i$. Then $\{\omega : \lim_{n\to\infty} S_n \text{ exists}\} \in \mathcal{T}$. Furthermore, since for any $m \in \mathbb{N}$,

$$\limsup_{n\to\infty} \frac{S_n}{n}(\omega) = \limsup_{m\leq n\to\infty} \frac{S(m,n)}{n},$$

where $S(m,n) = S_n - S_{m-1}$, we have $\{\limsup_{n\to\infty} \frac{S_n}{n} \leq a\} \in \mathcal{T}$ for any $a \in \mathbb{R}$. (To be specific, note that $S(m,n)$ is measurable w.r.t. $\sigma(X_m, X_{m+1}, \ldots)$ for every $n \geq m$. Therefore, $\limsup_{n\to\infty} S(m,n)/n$ is measurable w.r.t. $\sigma(X_m, X_{m+1}, \ldots)$ (cf. Lemma 4.2). This is true for all $m \geq 1$.)

On the other hand, consider the event $\{\limsup_{n\to\infty} S_n \geq a\}$ where X_i is not 0 a.s. It does not belong to $\mathcal{T}$ since it is in $\sigma(X_1, X_2, \ldots)$ but not in $\sigma(X_2, X_3, \ldots)$.

Lemma 8.3 (Grouping Lemma)**.** *Suppose that a set of σ-algebras $\{\mathcal{A}_t : t \in T\}$ indexed by the countable index set T are independent (i.e., any finite subset of σ-algebras $\mathcal{A}_{i_1}, \mathcal{A}_{i_2}, \ldots, \mathcal{A}_{i_n}$ are independent). Then for any finite partition $T = \bigcup_{i=1}^n T_i$ where $T_i \cap T_j = \emptyset \ \forall\, i \neq j$, the σ-algebras $\mathcal{F}_i = \sigma(\bigcup_{t\in T_i} \mathcal{A}_t)$, $i = 1, \ldots, n$ are independent.*

Proof. Let

$$\mathcal{C}_i = \left\{\bigcap_{t\in F} A_t \ : \ \ F \text{ is finite subset of } T_i,\ A_t \in \mathcal{A}_t\right\}.$$

We note that $\mathcal{C}_i$ is a π-system. For any finite $F_i \subset T_i$, $A_{t_i} \in \mathcal{A}_{t_i}$ where $t_i \in F_i$, we have

$$\mathbb{P}\left(\bigcap_{i=1}^n \bigcap_{t_i\in F_i} A_{t_i}\right) = \prod_{i=1}^n \mathbb{P}\left(\bigcap_{t_i\in F_i} A_{t_i}\right),$$

thus $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n$ are independent. From Lemma 5.2, $\sigma(\mathcal{C}_i)$, $i = 1, \ldots, n$ are independent. For each $t \in T_i$, we have $\mathcal{A}_t \subset \mathcal{C}_i$, hence $\mathcal{F}_i \subset \sigma(\mathcal{C}_i)$, which implies that $\mathcal{F}_i$, $i = 1, \ldots, n$ are independent. $\square$

Theorem 8.2 (Kolmogorov's 0-1 law)**.** *Suppose that $(X_i)_{i\geq 1}$ are independent. Let $\mathcal{T}$ be the tail σ-algebra. If $A \in \mathcal{T}$, then $\mathbb{P}(A) = 0$ or 1.*

Proof. We first show that if $A \in \sigma(X_1, X_2, \ldots)$ and $B \in \mathcal{T}$, then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. For any $k \geq 1$, since $\mathcal{T} \subset \sigma(X_{k+1}, X_{k+2}, \ldots)$, by Lemma 8.3, $\sigma(X_1, \ldots, X_k)$ and $\mathcal{T}$ are independent. It then follows that $\bigcup_{k\geq 1} \sigma(X_1, \ldots, X_k)$ and $\mathcal{T}$ are independent. Since they are both π-systems, by Lemma 5.2, we have

$$\sigma(X_1, X_2, \ldots) = \sigma\left(\bigcup_{k\geq 1} \sigma(X_1, \ldots, X_k)\right) \perp\!\!\!\perp \mathcal{T}.$$

Since $A \in \mathcal{T} \subset \sigma(X_1, X_2, \ldots)$, we obtain that A is independent of itself and

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap A)\\ &= \mathbb{P}(A)\mathbb{P}(A)\\ &= \mathbb{P}(A)^2,\end{aligned}$$

and the theorem follows. □

Example 8.1. *The event $\{S_n = \sum_{i=1}^n X_i$ converges $\} \in \mathcal{T}$. From the Kolmogorov's 0-1 law, S_n converges with probability 0 or 1. Intuitively, this means that if S_n converges in probability, it should also converge a.s. We will show this result rigorously in a future session.*

8.3 Hewitt-Savage 0-1 Law

We now suppose that $(X_i)_{i\geq 1}$ are i.i.d. A finite permutation $\pi : \{1, 2, \ldots\} \mapsto \{1, 2, \ldots\}$ is such that $\pi(i) \neq i$ for finitely many i. If B is such that $(X_i)_{i\geq 1} \in B \implies (X_{\pi(i)})_{i\geq 1} \in B$ for all finite permutations π, then we say B is invariant w.r.t. finite permutations. Let

$$\mathscr{E} = \{\{(X_i)_{i\geq 1} \in B\} : B \text{ is invariant}\}.$$

Then $\mathscr{E}$ is called the exchangeable σ-algebra. (Exercise: check that $\mathscr{E}$ is a σ-algebra.)

$\mathscr{E}$ consists of events that are not affected when we permute a finite number of random variables X_i. Clearly, $\mathcal{T} \subset \mathscr{E}$. On the other hand, $\{\limsup_{n\to\infty} S_n \geq a\} \notin \mathcal{T}$ but $\{\limsup_{n\to\infty} S_n \geq a\} \in \mathscr{E}$ so $\mathscr{E}$ is strictly larger than $\mathcal{T}$.

Theorem 8.3 (Hewitt-Savage 0-1 Law)**.** *Assume $(X_i)_{i\geq 1}$ are i.i.d. If $A \in \mathscr{E}$, then $\mathbb{P}(A) = 0$ or 1.*

To prove this, we first introduce a lemma. Let $A \triangle B = (A\backslash B) \cup (B\backslash A)$. We have

$$|\mathbb{P}(A) - \mathbb{P}(B)| = \left|\int \mathbf{1}_A \,\mathrm{d}\mathbb{P} - \int \mathbf{1}_B \,\mathrm{d}\mathbb{P}\right| \leq \int |\mathbf{1}_A(\omega) - \mathbf{1}_B(\omega)| \,\mathrm{d}\mathbb{P} = \mathbb{P}(A \triangle B).$$

Lemma 8.4 (Approximation Lemma)**.** *If $\mathcal{A}$ is a algebra and $B \in \sigma(\mathcal{A})$, then $\exists\ B_n \in \mathcal{A}$ such that $\mathbb{P}(B \triangle B_n) \to 0$ as $n \to 0$.*

Proof. Let $\mathcal{D} = \{B \in \sigma(\mathcal{A}) : \mathbb{P}(B \triangle B_n) \to 0 \text{ as } n \to 0\}$. Since that this is a λ-system and $\mathcal{A} \subset \mathcal{D}$, from the $\pi - \lambda$ theorem, we have $\sigma(\mathcal{A}) \subset \mathcal{D}$. □

Proof of Hewitt-Savage 0-1 Law. Let $X = (X_i)_{i \geq 1}$. We have $A = \{X \in B\}$ for a B that is invariant w.r.t. finite permutations. The σ-algebra $\sigma(X_1, X_2, \ldots)$ is generated by the algebra

$$\mathcal{A} = \Big\{\{(X_i)_{i \in F} \in C\} \,:\, F \text{ is finite, } C \in \mathcal{B}(\mathbb{R}^{|F|})\Big\}.$$

From Lemma 8.4, for any $\epsilon > 0$, $\exists\, A_n \in \sigma(X_1, \ldots, X_n)$ such that $\mathbb{P}(A \triangle A_n) \leq \epsilon$ for all n sufficiently large since $\mathscr{E} \subset \sigma(X_1, X_2, \ldots)$. We can write $A_n = \{(X_1, \ldots, X_n) \in B_n\}$ for some $B_n \in \mathcal{B}(\mathbb{R}^n)$. Now define $A'_n = \{(X_{n+1}, \ldots, X_{2n}) \in B_n\} \in \sigma(X_{n+1}, \ldots, X_{2n})$. From independence, we have $\mathbb{P}(A_n \cap A'_n) = \mathbb{P}(A_n)\mathbb{P}(A'_n)$.

Let $\pi(X) = (X_{n+1}, \ldots, X_{2n}, X_1, \ldots, X_n, X_{2n+1}, \ldots)$. We have

$$\begin{aligned}
\mathbb{P}(A'_n \triangle A) &= \mathbb{P}(\{(X_{n+1}, \ldots, X_{2n}) \in B_n\} \triangle \{X \in B\}) \\
&= \mathbb{P}(\{(X_{n+1}, \ldots, X_{2n}) \in B_n\} \triangle \{\pi(X) \in \pi(B)\}) \\
&= \mathbb{P}(\{(X_{n+1}, \ldots, X_{2n}) \in B_n\} \triangle \{\pi(X) \in B\}) \ \because \pi(B) = \{\pi(x) \,:\, x \in B\} = B \\
&= \mathbb{P}(\{(X_1, \ldots, X_n) \in B_n\} \triangle \{X \in B\}) \ \because X_i \text{ are i.i.d.} \\
&= \mathbb{P}(A_n \triangle A) \leq \epsilon.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{P}((A'_n \cap A_n) \triangle A) &\leq \mathbb{P}((A_n \triangle A) \cup (A'_n \triangle A)) \\
&\leq \mathbb{P}(A_n \triangle A) + \mathbb{P}(A'_n \triangle A) \\
&\leq 2\epsilon,
\end{aligned}$$

and since $|\mathbb{P}(A) - \mathbb{P}(A_n \cap A'_n)| \leq \mathbb{P}((A'_n \cap A_n) \triangle A)$, we have $|\mathbb{P}(A) - \mathbb{P}(A_n \cap A'_n)| \to 0$ as $n \to \infty$. Similarly, we have $|\mathbb{P}(A) - \mathbb{P}(A_n)| \to 0$ and $|\mathbb{P}(A) - \mathbb{P}(A'_n)| \to 0$ so that $\mathbb{P}(A_n \cap A'_n) = \mathbb{P}(A_n)\mathbb{P}(A'_n) \to \mathbb{P}(A)^2$. Therefore, we have $\mathbb{P}(A) = \mathbb{P}(A)^2$. □

8.4 Discussions

Suppose $(X_i)_{i \geq 1}$ are i.i.d., $\mathbb{E}X_1 = 0$, and $\mathbb{E}X_1^2 = 1$. Take $0 < b_n \to \infty$, then $\frac{S_n}{b_n} \to 0$ a.s. if $\sum_{n \geq 1} \frac{1}{b_n^2} < \infty$. For example, if we choose $b_n = \sqrt{n \log n}$, then $\frac{S_n}{\sqrt{n \log n}} \to 0$ a.s., i.e., S_n grows slower than $\sqrt{n \log n}$.

On the other hand, consider $X_i \sim N(0, 1)$, then $\frac{S_n}{\sqrt{n}} \sim N(0, 1)$. For any $r \in \mathbb{R}$, we have

$$\mathbb{P}\left(\limsup_{n \to \infty} \frac{S_n}{\sqrt{n}} \geq r\right) = \lim_{m \to \infty} \mathbb{P}\left(\bigcup_{n \geq m} \left\{\frac{S_n}{\sqrt{n}} \geq r\right\}\right) \geq \lim_{m \to \infty} \mathbb{P}\left(\frac{S_m}{\sqrt{m}} \geq r\right) > 0.$$

Since $\{\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} \geq r\}$ is a tail event, by Kolmogorov's 0-1 law, we then have

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} \geq r\right) = 1,$$

which implies that $\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = \infty$ a.s., i.e., S_n grows faster than $\sqrt{n}$.

The exact rate of growth of S_n is given by the following theorem.

Theorem 8.4 (Law of Iterated Logarithm)**.** *Suppose* $(X_i)_{i\geq 1}$ *are i.i.d.,* $\mathbb{E}X_1 = 0$*, and* $\mathbb{E}X_1^2 = 1$*. Let* $S_n = \sum_{i=1}^n X_i$*. We have*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1 \ \text{ a. s.}$$

The proof of this theorem is left as a research exercise.