Law of Large Numbers

Jieun Shin
January 11, 2023

Convergence of random variables i

Theorem 1. Convergence in distribution

A sequence $X_1, X_2, \ldots$ of the real-valued random variables, with cumulative distribution functions $F_1, F_2, \ldots,$ is said to converge in distribution, or converge weakly, or converge in law to a random variable X with cumulative distribution function F if

$$\lim_{x \to -\infty} F_n(x) = F(x), \tag{1}$$

for every number $x \in \mathbb{R}$ at which F is continuous.

Theorem 2. Convergence in probability

A sequence $\{X_n\}$ of random variables converge in probabilitytoward the random variable X if for all ϵ

$$\lim_{x \to -\infty} \mathbf{Pr}(|X_n - X| > \epsilon) = 0. \tag{2}$$

Convergence of random variables ii

Theorem 3-1. Almost sure convergence

The sequence $\{X_n\}$ converges almost surely or almost everywhere or with probability 1 or strongly towards X means that

$$\mathbf{Pr}\left(\lim_{x\to\infty} X_n = X\right) = 1, \tag{3}$$

for every number $x \in \mathbb{R}$ at which F is continuous.

Theorem 3-2. Almost sure convergence

This means that the values of X_n approach the value of X, in the sense (see almost surely) that events for which X_n does not converge to X have probability 0. Using the probability space $(\Omega, \mathcal{F}, \mathbf{Pr})$ and the concept of the random variable as a function from Ω to $\mathbb{R}$, this is equivalent to the statement

$$\mathbf{Pr}\left(w \in \Omega : \lim_{x\to\infty} X_n(w) = X(w)\right) = 1, \tag{4}$$

Theorem 4. Convergence in mean

Given a real number $r \geq 1$, we say that the sequence X_n converges in the r-th mean (or in the L_r-norm) towards the random variable X, if the r-th absolute moments $(|X_n|^r)$ and $\mathbb{E}(|X|^r)$ of X_n and X exist, and

$$\lim_{n\to\infty} \mathbb{E}\left(|X_n - X|^r\right) = 0, \tag{5}$$

where the operator $\mathbb{E}$ denotes the expected value. Convergence in r-th mean tells us that the expectation of the r-th power of the difference between X_n and X converges to zero.

Law of Large Numbers i

- In probability theory, the law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times.
- The average of the results obtained from a large number of trials should be close to the expected value and tends to become closer to the expected value as more trials are performed.
- the LLN only applies to the average:

$$\lim_{x \to \infty} \sum_{i=1}^{n} \frac{X_i}{n} = \bar{X}. \tag{6}$$

But, $\frac{1}{n} \sum_{i=1}^{n} X_i - n \cdot \bar{X}$ is not converge toward zero as n increase.

Basic concept

Let $X_1, X_2, \dots$ be an infinite sequence of i.i.d random variables with expected value $E(X_1) = E(X_2) = \cdots = \mu$. Then, the sample average $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ converges to the expected value:

$$\bar{X}_n \to \mu \quad \text{as } n \to \infty. \tag{7}$$

- Does a finite variance is necessary?
  If the variance of the $\bar{X}_n$ is finite, then it can be used to show (6) simply, but it is not necessary. Large or infinite variance will make the convergence slower, but the LLN holds anyway.
- For interpretation of the weak and the strong LLN, see Conbergence of random variables.

Theorem 5. Weak law

The WLLN (a.k.a. Khinchin's law) state that the sample average converges in probability towards the expected value

$$\bar{X}_n \xrightarrow{p} \mu \quad \text{as } n \to \infty. \tag{8}$$

That is, for any positive number ϵ,

$$\lim_{x \to -\infty} \mathbf{Pr}(|X_n - X| < \epsilon) = 1. \tag{9}$$

- If the variance is bounded, then the law can be shown by Chebyshev.

Theorem 6-1. Strong law

The SLLN (a.k.a. Kolmogorov law) state that the sample average converges almost surely to the expected value

$$\bar{X}_n \xrightarrow{a.s.} \mu \quad \text{as } n \to \infty. \tag{10}$$

That is,

$$\mathbf{Pr}(\lim_{x \to \infty} \bar{X}_n = \mu) = 1. \tag{11}$$

- The strong law of large numbers can itself be seen as a special case of the pointwise ergodic theorem.
- Random variables which converge strongly are guaranteed to converge weakly, not vice versa.
- Noting that Kolmogorov's strong law needs the (minimal) assumption that $(X_n)_{n \geq 1}$ are in L_1.

Theorem 6-2. Strong law in L_2

Suppose $\{X_n\}_{n\geq 1}$ are i.i.d random variables defined on the probability same space. Let $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = Var(X_i) < \infty$. Then,

$$\bar{X}_n \xrightarrow{a.s.} \mu \quad \text{as } n \to \infty. \tag{12}$$

If we add the assumption that $Var(X_i) < \infty$, then **Theorem 5** converge in L_2

proof of (11)

We assume $\mu = 0$ without loss of generality. Let $Y_n = \bar{X}_n$. Then $\mathbb{E}(Y_n) = 0$ and $\mathbb{E}(Y_n^2) = \sigma^2/n$. So,

$$\sum_{n=1}^{\infty} \mathbb{E}(Y_n^2) = \sum_{n=1}^{\infty} \frac{\sigma^2}{n} < \infty.$$

(Theorem 9-2 in book)

If $\sum_{n=1}^{\infty} \mathbb{E}\{X_n\} < \infty$, then $\sum_{n=1}^{\infty} \{X_n\}$ convergence a.s.

Here, since $\sum_{n=1}^{\infty} 1/n = \infty$, we alternatively use $\sum_{n=1}^{\infty} 1/n^2$. By Thm 9.2, since $\sum_{n=1}^{\infty} \mathbb{E}(Y_n^2) < \infty$ a.s., $Y_n^2 \to 0$ a.s.

Define $p(n) = \lfloor \sqrt{n} \rfloor, n = 1, 2, \ldots$. Then $p(n)^2 \leq n < (p(n)+1)^2$. Note that

$$\frac{1}{n} \sum_{j=p(n)^2+1}^{n} X_j = \frac{1}{n} \sum_{j=n}^{n} - \frac{1}{n} \sum_{j=n}^{p(n)^2} X_j$$

$$= Y_n - \frac{p(n)^2}{n} Y_{p(n)^2},$$

Law of Large Numbers vii

and that

$$
\begin{aligned}
\mathbb{E}\left[\left(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2}\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{j=p(n)^2+1}^{n} X_j\right)^2\right] \\
&= \frac{1}{n^2} Var\left[\sum_{j=p(n)^2+1}^{n} X_j\right] \\
&= \frac{n - p(n)^2}{n^2}\sigma^2 \\
&\leq \frac{2p(n)+1}{n^2}\sigma^2 && (\because n \leq (p(n)+1)^2) \\
&\leq \frac{3\sqrt{n}}{n^2}\sigma^2 && (\because p(n)^2 \leq n) \\
&= \frac{3}{n^{3/2}}\sigma^2.
\end{aligned}
$$

Thus,

$$\sum_{n=1}^{\infty} \mathbb{E}\left[\left(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2}\right)^2\right] \leq \sum_{n=1}^{\infty} \frac{3}{n^{3/2}} \sigma^2 < \infty.$$

By Thm 9.2,

$$\begin{aligned} &\left(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2}\right)^2 < \infty \quad \text{a.s.} \\ &\Rightarrow Y_n - \frac{p(n)^2}{n} Y_{p(n)^2} \to 0 \quad \text{a.s.} \\ &\Rightarrow Y_n \to 0 \quad \text{a.s.,} \end{aligned}$$

the third line hold by $Y_{p(n)^2} \to 0$ a.s. and $0 < \frac{p(n)^2}{n} < 1$. ■

What does SLLN mean?

Let $Y_1, \ldots, Y_n$ be a Bernoulli distributed random variable with probability $p = 0.5$. We toss a coin n times and get event $w = (Y_1, Y_2, \ldots, Y_n) = (1, 1, \ldots, 1)$, which all results are 1. Then,

$$\lim_{x \to \infty} \bar{X}_n = \lim_{x \to \infty} \frac{1 + 1 + \cdots + 1}{n} = \lim_{x \to \infty} \frac{n}{n} = 1.$$

If we expected $\bar{X}_n = 0.5$ nearby, then the results are not the expected value. What if we collect a set of events that satisfy the limit? So consider:

$$\mathcal{E} = \left\{ w \in \Omega : \lim_{x \to \infty} \bar{X}_n = \mu \right\}.$$

SLLN means that the probability of event $\mathcal{E}$ occurring is 1. Because of the concept of a limit, we call it almost sure 1 rather than clearly 1.

Example 1. Bernoulli distribution

Let X_i be a Bernoulli distributed random variable with probability p. Then,

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

Example 2. Monte Carlo Approximation

Suppose f is a L_1 measurable function defined on $[0, 1]$ and $U_i \sim^{i.i.d}$ Unif$(0, 1)$. Then,

$$\mathbb{E}[f(U_i)] = \int_0^1 f(x)dx.$$

Thus,

$$\lim_{x \to -\infty} \frac{1}{n} \sum_{i=1}^{n} f(U_i) \xrightarrow{a.s.} \int_0^1 f(x)dx.$$

■

Q. 20.1 (A Weak Law of Large Numbers). Let (X_i) be a sequence of random variables such that $\sup_i \mathbb{E}\left\{X_i^2\right\} = c < \infty$ and $E\{X_j X_k\} = 0$ if $j \neq k$. Let $S_n = \sum_{i=1}^n X_i$.

a) Show that $\mathbf{Pr}\left(\left|\frac{1}{n}S_n\right| \geq \varepsilon\right) \leq \frac{c}{n\varepsilon^2}$ for $\varepsilon > 0$.
b) $\lim_{n\to\infty} \frac{1}{n}S_n = 0$ in L^2 and in probability.

(Note: The usual i.i.d. assumptions have been considerably weakened here.)

Solution for 20.1

a)

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}S_n\right)^2\right] &= \frac{1}{n^2}\mathbb{E}\left[(X_1 + X_2 + \cdots + X_i)^2\right] \quad \because E(X_j X_k) = 0 \ (j \neq k) \\
&= \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^n X_i^2\right) \\
&= \frac{1}{n^2} \cdot \sum_{i=1}^n \mathbb{E}\left(X_i^2\right) \\
&= \frac{c}{n}
\end{aligned}
$$

이다. 그러므로 chebyshev‘s inequality에 의해서 주어진 $\epsilon > 0$에 대하여

$$P\left(\left|\frac{1}{n}S_n\right| \geq \epsilon\right) \leq \frac{\mathbb{E}\left[\frac{1}{n^2}S_n^2\right]}{\epsilon^2} = \frac{c}{n\varepsilon^2}$$

가 성립한다.

b) (a)로 부터 $\mathbb{E}\left[\left(\frac{1}{n}S_n\right)^2 - 0\right] = \frac{1}{n}c$임을 알고있다. 따라서

$$\mathbb{E}\left[\left(\frac{1}{n}S_n\right)^2 - 0\right] = \frac{c}{n} \to 0 \quad \text{as } n \to \infty$$

이므로 $\frac{1}{n}S_n \to 0$ in L^2 이다. 또한

$$\mathbf{Pr}\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) \leq \frac{c}{n\epsilon^2} \to 0 \quad \text{as } n \to \infty$$

이므로 $\frac{1}{n}S_n \xrightarrow{p} 0$도 성립한다. ■

Q. 20.2

Let $(Y_i)_{i \geq 1}$ be a sequence of independent Binomial random variables, all defined on the same probability space, and with law $B(p, 1)$. Let $X_n = \sum_{i=1}^{n} Y_i$. Show that X_i is $B(p, i)$ and that $\frac{X_i}{i}$ converges a.s. to p.

Solution for 20.2

$(Y_i)_{i \geq 1} \overset{\text{iid}}{\sim} B(1, p)$의 mgf는

$$M_Y(t) = q + pe^t, \quad (q = 1 - p)$$

이므로 $X_n = \sum_{i=1}^{n} Y_i$의 mgf는

$$M_X(t) = (q + pe^t)^n$$

이는 $B(n, p)$의 mgf 이므로 $X_n \sim B(n, p)$임을 알 수 있다. S.L.L.N에 의해

$$\frac{1}{n} X_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \to E(Y_1) = p \quad \text{a.s.}$$

이 성립한다. ■

Q. 20.4 Let $(X_i)_{i\geq 1}$ be i.i.d. with X_i in L^1 and $E\{X_i\} = \mu$. Let $(Y_i)_{i\geq 1}$ be also i.i.d. with Y_i in L^1 and $E\{Y_i\} = \nu \neq 0$. Show that

$$\lim_{n\to\infty} \frac{1}{\sum_{i=1}^{n} Y_i} \sum_{i=1}^{n} X_i = \frac{\mu}{\nu} \quad \text{a.s.}$$

Solution for 20.4

S.L.L.N의 정의에 의해

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to u \quad \text{a.s} \quad \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \to \nu$$

이다. 즉 집합 A, B

$$A = \left\{w;\, \lim_{n\to\infty} \bar{X}_n(w) = u\right\}, B = \left\{w;\, \lim_{n\to\infty} \bar{Y}_n(w) = \nu\right\}$$

에 대하여 $P(A) = P(B) = 1$을 만족한다. 이때 두 집합 A, B에 대하여

$$\begin{aligned} A \cap B &= \left\{ w;\ \lim_{n\to\infty} \bar{X}_n(w) = u,\ \lim_{n\to\infty} \bar{Y}_n(w) = V \right\} \\ &\subset \left\{ w;\ \lim_{n\to\infty} \frac{\bar{X}_n(w)}{\bar{Y}_n(w)} = \frac{u}{v} \right\} = C \end{aligned}$$

이므로 $P(A \cap B) \le p(c)$이고,
$1 \ge P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \Rightarrow P(A \cap B) \ge 1$ 이므로 $P(c) = 1$ 아다. 따라서,

$$\lim_{n\to\infty} \frac{1}{\sum_{i=1}^{n} Y_i} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \frac{\bar{X}_n}{\bar{Y}_n} = \frac{u}{\nu} \quad \text{a.s.}$$

■