Law of Large Numbers and Central Limit Theorem

Ma Xuejun

School of Mathematical Sciences

Soochow University

https://xuejunma.github.io

2019.10.07

Outline

Review

- Definition of convergence
- Relationship between 4 types of convergence

$$
\begin{array}{l}
X_n \stackrel{a.s.}{\to} X \\
\qquad\qquad\qquad \Rightarrow X_n \stackrel{p}{\to} X \Rightarrow X_n \stackrel{d}{\to} X \\
X_n \stackrel{L_r}{\to} X
\end{array}
$$

- Stochastic orders: O_p, o_p

Terms

- The Weak/Strong Law of Large Numbers
- The Central Limit Theorem

Law of Large Numbers

- Now that we have methods to describe the limit of a sequence of random variables
- Recall the motivating problem: Sample mean $\bar{X}_n$ converges to EX intuitively.
- Question: what is this convergence? Is it convergence in distribution.? probability? a.s.? L_r?

Weak Law of Large Numbers (WLLN)

WLLN

Let $\{X_n\} = X_1, X_2, \ldots,$ be a sequence of independently and identically distributed (i.i.d.) r.v.'s such that $E|X_1| < \infty$, Then

$$\bar{X}_n \xrightarrow{p} EX_1$$

- The condition $E|X_1| < \infty$ is to assure the existence of EX_1.
- The theorem can be extended to many dependence structures, such as Markov chains.
- The theorem can be extended to cases that Xi's are not identical, but share the same 1st and 2nd moments.
- According to properties for convergence in probability, for any cont. function $g(\cdot)$,

$$g\left(\bar{X}_n\right) = g\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \xrightarrow{p} g(\mu).$$

Sketch of Proof of WLLN

Note that when the limit is a constant, convergence in probability is equivalent with convergence in distribution. To prove convergence in distribution, we only need to show

$$\phi_{X_n}(t) \to \phi_{EX_1} = e^{itEX_1}$$

We will use the following result without any proof. For a r.v. X with finite first moment, we have

$$\phi_X(X) = 1 + itEX + o(t)$$

Proof:

$$\begin{aligned}\phi_{\bar{X}_n}(t) &= E\Big[\exp(it\bar{X}_n)\Big] = E\Big[\exp\Big(it\frac{1}{n}\sum_{i=1} nX_i\Big)\Big] \\ &= \prod_{i=1}^{n} E[\exp(itX_i/n)] \\ &= \Big(E[\exp(itX_1/n)]\Big)^n = \phi_{X_1}^n(t/n)\end{aligned}$$

Let $n \to \infty$, then

$$\phi_{\bar{X}_n}(t) = \Big(1 + itEX_1/n + o(1/n)\Big)^n \to e^{itEX_1}$$

Therefore, the convergence is proved.

Strong Law of Large Numbers (WLLN)

SLLN

Let $\{X_n\} = X_1, X_2, \dots,$ be a sequence of independently and identically distributed (i.i.d.) r.v.'s such that $E|X_1| < \infty$, Then

$$\bar{X}_n \overset{a.s.}{\to} EX_1$$

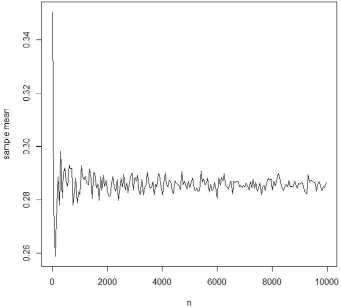
- The conditions can be relaxed. Identical distributions are not required, but there are still constraints on the second moment
- Stronger requirements than WLLN to assure better results
- The proof is beyond the scope of this course
- LLN: when n gets larger, the distribution of sample mean is more concentrated around EX_1.

Example of LLN: Calculate Expectation

Recall: $EX = \inf_{\infty}^{\infty} x f(x) dx$, where $f(x)$ is pdf of X.

- Generate n samples with pdf $f(x)$, and calculate the $\bar{X}_n$. When n is very large, $EX \approx \bar{X}_n$
- Example: Beta distribution with parameters $a = 2$ and $b = 2$,

$$EX = \int_0^1 \frac{\Gamma(7)}{\Gamma(2)\Gamma(5)} x^{2-1}(1-x)^{5-1} dx$$ Hard to Calculate!



Example of LLN: Calculate Expectation

$Beta(2, 5)$

```
rm(list=ls())
n.vec <- seq(1, 10^4, 50)
n.len <- length(n.vec)
mean.full <- NULL
for(i in 1:n.len){
  mean.full[i] <- mean(rbeta(n=n.vec[i], shape1=2,
                              shape2=5))
}
plot(n.vec, mean.full, type="l",xlab = "n",
     ylab = "sample mean")
abline(h=2/7,lwd=1,col="blue")
```

- What's more, according to the continuous mapping theorem, $g(\bar{X}_n) \to g(E(X))$, e.g., $[E(X)]^2 \approx \bar{X}_n^2$

Example of LLN: Calculate Expectation

- LLN can also be used to find $E(g(X))$, where $g(\cdot)$ is a function
- Generate n i.i.d. samples $\{X_i\}_{i=1}^{n}$ with pdf $f(x)$, and let $Y_i = g(X_i)$. Then, $\bar{Y}_n \to E[g(X)]$. When n is very large, $E(g(X)) \approx \bar{Y}_n$
- Example: Beta distribution with parameters $a = 2, b = 5$.

$$Y = X^2 \qquad Z = 2X + 1 \qquad W = e^X$$

Examples of Using LLN: Integration

- Suppose we wish to calculate

$$\int_0^1 g(x)dx$$

where $g(x)$ may be complicated and the integration in not easy to compute.

- Relate the integration with expectation. We need a density function. Let $X \sim Unif(0,1)$, then the pdf of X is 1 on $[0,1]$. For function $g(\cdot)$,

$$Eg(X) = \int_0^1 g(X) \times 1dx = \int_0^1 g(x)dx$$

Procedure (apply the method in previous slide for mean):

- Generate n i.i.d samples $X \sim Unif(0,1)$, and calculate $g(X_i)$ correspondingly
- Compute $Eg(X) \approx \overline{g(X_i)} = \frac{1}{n}\sum_{i=1}^n g(X_i)$
- This method is called Monte Carlo method.

Motivation

Suppose that a fair coin is tossed 100 times. What is the probability that the total number of heads is no smaller than 60?

Let X be the total number of heads, then $X \sim Bin(100, 0.5)$. We are interested in $P(X \geq 60)$

- Calculate directly means calculating 40 probs $\{p(X = i)\}_{i=60,61,\ldots}$ and take the summation. COMPLICATED.
- X can be seen as the summation of 100 Bernoulli trials with $p = 0.5$ and limit theorems can be applied.
  - With LLN, we only know $X/100 \xrightarrow{p} 0.5$, CANNOT get $P(X \geq 60)$
  - New Limit Theorem is required to describe the behaviour of X more accurately.

Central Limit Theorem (CLT)

Let $\{X_n\} = X_1, X_2, \ldots,$ be a sequence of independently and identically distributed (i.i.d.) r.v.'s such that $EX_1^2 < \infty$. Let $\sigma^2 = Var(X_1)$ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n}$, then

$$\sqrt{n}[\bar{X}_n - EX_1] = \sqrt{n}\Big[\frac{1}{n}\sum_{i=1}^{n}(X_i - EX_1)\Big] \xrightarrow{d} N(0, \sigma^2)$$

- $EX_1^2 < \infty$ is a regular condition to assure the existence of EX_1 and $Var(X_1)$
- It means that $\bar{X}_n$ can be approximated by a normal distribution, no matter what the distribution for X_i is.
- Here, $n^{-0.5}$ is the convergence rate. Or, say, $\bar{X}_n = O_p(n^{-0.5})$. If we use $n^{0.5+\delta}$ with $\delta > 0$, then there is no meaningful result; if we use $n^{0.5-\delta}$ then it converges to 0.

Comments on CLT

CLT is the most important theorem in statistics

- CLT means that, the sample mean will be approximately normally distributed for large sample sizes, regardless of the distribution of the samples
- Many statistics (say, $\bar{X}_n, \bar{X}_n^2$ have distributions that are approximately normal, even the population distribution is not normal ($\Leftarrow$ The dist. of statistics can be approximated)
- Statistical inference can be derived based on normality, provided the sample size is large.
- In practice, it gives a very rough guideline to approximate $\bar{X}_n$ when n is large (a few hundreds or even more)
- However, the convergence is the weakest convergence, converge in distribution. With the result, for statistics (e.g., $\bar{X}_n$), we can only calculate

$$P(\bar{X}_n \geq a), P(\bar{X}_n \leq a), P(a \leq \bar{X}_n \leq b)$$

Comparison Between LLN and CLT

	LLN	CLT
Results	Focus on $\bar{X}_n$	Focus on $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$
	$\bar{X}_n \xrightarrow{p} \mu$	$\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z$
Convergence	In probability	In distribution
Interpretation	$\bar{X}_n$ converges to μ	The rate $\bar{X}_n$ converges to μ
Usage	Monte Carlo Method	Statistical Inference

-
- Compare to real numbers, LLN means that

$$\frac{2\sqrt{n}+1}{\sqrt{n}} \to 2$$

CLT mean that

$$\sqrt{n}\Big(\frac{2\sqrt{n}+1}{\sqrt{n}} - 2\Big) \to 1.$$

CLT Example

Let $X_i \overset{i.i.d}{\sim} Exp(1), i = 1, 2, \ldots$. We know that $E[X_1] = Var(X_1) = 1$, and so the sample mean converges to 1. How many samples we need so that our error is at most 10%, with probability more than 0.95?

The target is, to figure out n, so that $P(0.9 \leq \bar{X}_n \leq 1.1) \geq 0.95$. For large n, with CLT, we have $\sqrt{n}(\bar{X}_n - 1) \xrightarrow{d} N(0, 1)$. 1). Therefore, we may use standard normal distribution to approximate the probability. Then,

$$\begin{aligned} P(0.9 \leq \bar{X}_n \leq 1.1) &= P(-0.1 \leq \bar{X}_n - 1 \leq 0.1) \\ &= P(-0.1\sqrt{n} \leq \sqrt{n}(\bar{X}_n - 1) \leq 0.1\sqrt{n}) \\ &\approx \Phi(0.1\sqrt{n}) - \Phi(-0.1\sqrt{n}) \\ &= 2\Phi(0.1\sqrt{n}) - 1 \geq 0.95 \end{aligned}$$

Check the normal table, and we can find $n > 384$.