Th1: The Law of Large Numbers

Federico Detomaso

1 Meaning

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. According to the law, the average of the results obtained from a large number of independent identical trials should be close to the expected value and tends to become closer to the expected value as more trials are performed.

2 Definition

2.1 Sample mean

Let $\{X_n\}$ be a sequence of random variables.

Let $\overline{X_n}$ be the sample mean of the first n terms of the sequence:

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

A Law of Large Numbers (LLN) states some conditions that are sufficient to guarantee the convergence of $\overline{X_n}$ to a constant, as the sample size n increases.

Typically, all the random variables in the sequence $\{X_n\}$ have the same expected value $E\left[\overline{X_n}\right] = \mu$. In this case, the constant to which the sample mean converges is μ (which is called population mean).

But there are also Laws of Large Numbers in which the terms of the sequence $\{X_n\}$ are not required to have the same expected value. In these cases, which are not treated in this lecture, the constant to which the sample mean converges is an average of the expected values of the individual terms of the sequence $\{X_n\}$.

2.2 Weak Law

The weak law of large numbers (also called Khinchin's law) states that the sample average converges in probability towards the expected value.

$\overline{X_n} \longrightarrow \mu$ when $n \longrightarrow \infty$ (converges in probability)

That is, for any positive number ε,

$\lim_{n\to\infty} Pr(\left|\overline{X_n} - \mu\right| < \varepsilon) = 1$

Interpreting this result, the weak law states that for any nonzero margin specified (ε), no matter how small, with a sufficiently large sample there will be a very high probability that the average of the observations will be close to the expected value; that is, within the margin.

2.3 Strong law

The strong law of large numbers (also called Kolmogorov's law) states that the sample average converges almost surely to the expected value.

$\overline{X_n} \longrightarrow \mu$ when $n \longrightarrow \infty$ (converges almost surely)

That is,

$Pr(\lim_{n\to\infty} \overline{X_n} = \mu) = 1$

What this means is that the probability that, as the number of trials n goes to infinity, the average of the observations converges to the expected value, is equal to one. The modern proof of the strong law is more complex than that of the weak law, and relies on passing to an appropriate subsequence.

The strong law of large numbers can itself be seen as a special case of the pointwise ergodic theorem. This view justifies the intuitive interpretation of the expected value (for Lebesgue integration only) of a random variable when sampled repeatedly as the "long-term average".

3 Proof

3.1 Proof Weak LLN

Given $X_1, X_2, ...$ an infinite sequence of i.i.d. random variables with finite expected value $E(X_1) = E(X_2) = ... = \mu < \infty$, we are interested in the convergence of the sample average

$$\overline{X_n} = \frac{1}{n}(X_1 + ... + X_n)$$

The weak law of large numbers states:

$$\overline{X_n} \longrightarrow \mu \text{ when } n \longrightarrow \infty \text{ (converges in probability)}$$

This proof uses the assumption of finite variance $Var(X_i) = \sigma^2$ (for all i). The independence of the random variables implies no correlation between them, and we have that

$Var(\overline{X_n}) = Var(\frac{1}{n}(X_1 + ... + X_n)) = \frac{1}{n^2}Var(X_1 + ...V_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

The common mean μ of the sequence is the mean of the sample average:

$E\left[\overline{X_n}\right] = \mu$

Using Chebyshev's inequality on $\overline{X_n}$ results in

$Pr(\left|\overline{X_n} - \mu\right| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$

This may be used to obtain the following:

$Pr(\left|\overline{X_n} - \mu\right| < \varepsilon) = 1 - Pr(\left|\overline{X_n} - \mu\right| \geq \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$

As n approaches infinity, the expression approaches 1. And by definition of convergence in probability, we have obtained

$\overline{X_n} \longrightarrow \mu$ when $n \longrightarrow \infty$ (converges in probability)

3.2 Proof Strong LLN

We give a relatively simple proof of the strong law under the assumptions that the X_i are iid, $E[X_i] =: \mu < \infty, Var(X_i) = \sigma^2 < \infty$ and $E[X_i^4] =: \tau < \infty$.

Let us first note that without loss of generality we can assume that $\mu = 0$ by centering. In this case, the strong law says that

$Pr(\lim_{n\to\infty} \overline{X_n} = 0) = 1$

or

$Pr(\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = 0) = 1$

It is equivalent to show that

$Pr(\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} \neq 0) = 1$

Note that

$\lim_{n\to\infty} \frac{S_n(\omega)}{n} \neq 0 \iff \exists \epsilon > 0, \left|\frac{S_n(\omega)}{n}\right| \geq \epsilon$ infinitely often,

and thus to prove the strong law we need to show that for every $\epsilon > 0$, we have

$Pr(\omega : |S_n(\omega)| \geq n\epsilon$ infinitely often $) = 0$.

Define the events $A_n = \omega : |S_n| \geq n\epsilon$, and if we can show that

$\sum_{n=1}^{\infty} Pr(A_n) < \infty$,

then the Borel-Cantelli Lemma implies the result. So let us estimate $Pr(A_n)$.

We compute

$E[S_n^4] = E[(\sum_{i=1}^{n} X_i)^4] = E[\sum_{1\leq i,j,k,l\leq n} X_i X_j X_k X_l]$.

We first claim that every term of the form $X_i^3 X_j, X_i^2 X_j X_k, X_i X_j X_k X_l$ where all subscripts are distinct, must have zero expectation. This is because $E[X_i^3 X_j] = E[X_i^3]E[X_j]$ by independence, and the last term is zero — and similarly for the other terms. Therefore the only terms in the sum with nonzero expectation are $E[X_i^4]$ and $E[X_i^2 X_j^2]$. Since the X_i are identically distributed, all of these are the same, and moreover $E[X_i^2 X_j^2] = (E[X_i^2])^2$. There are n terms of the form $E[X_i^4]$ and $3n(n-1)$ terms of the form $(E[X_i^2])^2$, and so

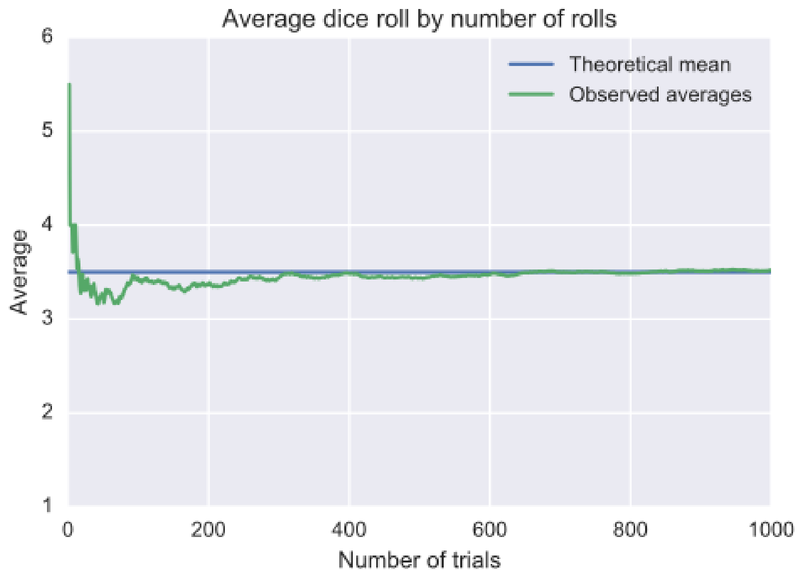
$E[S_n^4] = n\tau + 3n(n-1)\sigma^4$.

Note that the right-hand side is a quadratic polynomial in n, and as such there exists a $C > 0$ such that $E[S_n^4] \leq Cn^2$ for n sufficiently large. By Markov,

$Pr(|S_n| \geq n\epsilon) \leq \frac{1}{(n\epsilon)^4} E[S_n^4] \leq \frac{C}{\epsilon^4 n^2}$,

for n sufficiently large, and therefore this series is summable. Since this holds for any $\epsilon > 0$, we have established the Strong LLN.

4 Simulations



An illustration of the law of large numbers using a particular run of rolls of a single dice. As the number of rolls in this run increases, the average of the values of all the results approaches 3.5. Although each run would show a distinctive shape over a small number of throws (at the left), over a large number of rolls (to the right) the shapes would be extremely similar.

References

[1] https://en.wikipedia.org/wiki/Law_of_large_numbers#Proof_of_the_weak_law

[2] https://www.statlect.com/asymptotic-theory/law-of-large-numbers