

Law of Large samples

SSSA - Applied Statistics- Chiara Seghieri and Costanza Tortù

2023-09-26

Idea

Imagine you're a researcher studying public opinion on a new government policy. You want to know what percentage of the population supports this policy. To do this, you decide to conduct a survey.

Now, according to the law of large numbers, as your sample size increases, the sample mean (the proportion of individuals supporting the policy in this case) will converge to the true population mean. In other words, as your sample size approaches infinity, your sample mean will approach the true population mean. In practical terms, this means that with a large enough sample, you are more likely to get an accurate estimate of the proportion of people supporting the policy in the entire population.

So, the law of large numbers is reassuring researchers that, with a sufficiently large sample, the sample proportion becomes a reliable estimate of the population proportion, helping to minimize the impact of random variation.

Recap

WLLN

Weak law of large numbers: the sample mean of a sequence of independent and identically distributed (i.i.d.) random variables converges in probability to the population mean as the sample size increases.

Using the survey example:

Suppose you have a sequence of surveys with increasing sample sizes (50, 500, 5,000). According to the weak law of large numbers, as your sample size increases, the probability that the sample proportion is very different from the true population proportion becomes smaller and smaller. So, with a sufficiently large sample, you can be reasonably confident that your estimate is close to the true proportion, but it doesn't guarantee convergence for every possible outcome.

SLLN

Strong law of large numbers: the sample mean of a sequence of i.i.d. random variables converges almost surely to the population mean as the sample size increases.

Difference between the two

The key distinction between the weak and strong versions lies in the type of convergence. The weak version deals with convergence in probability, while the strong version deals with almost sure convergence, which

is a stronger form of convergence. In practical terms, the weak law is often sufficient for many statistical applications, but the strong law provides a more powerful and stronger guarantee of convergence.

Get the idea through simulations

First step: simulate the data

Randomly generate N realizations of a binary variable X , whose probability of taking the value 1 is fixed and equal to 0.5. X represents the support to a given policy.

```
N <- 10000 # number of observations
x <- sample(0:1, N, replace = T) # generate a binary variable
```

Second step: take the average

```
avg <- cumsum(x)/(1:N)
```

Third step: plot your results

```
plot(avg, ylim=c(.30, .70), type = "l", xlab = "Observations",
      ,ylab = "Probability", lwd = 2)
lines(c(0,N), c(.50,.50),col="red", lwd = 2)
```

