CSL003P1M : Probability and Statistics Lecture 36 (Weak and Strong Law of Large Numbers)

Sumit Kumar Pandey

December 06, 2021

The Weak Law of Large Numbers

The Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X] = \mu$. Then, for any $\epsilon > 0$,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \to 0 \text{ as } n \to \infty$$

Proof: (Hint: Apply Chebyshev's inequality) Assumption: Random variables have a finite variance σ^2. Since,

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \text{ and } Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

It follows from the Chebyshev's inequality that

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|\geq\epsilon\right\}\leq\frac{\sigma^2}{n\epsilon^2}$$

Hence the result.

The Strong Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $\mu = E[X_i]$. Then, with probability 1

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \text{ as } n \to \infty$$

That is, the strong law of large number states that

$$P\left\{\lim_{n\to\infty}(X_1 + \cdots + X_n)/n = \mu\right\} = 1$$

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx \text{ as } n \to \infty$$

The Central Limit Theorem For Independent Random Variables

Let $X_1, X_2, \ldots$ be a sequence of independent random variables having respective means and variances $\mu_i = E[X_i]$, $\sigma_i^2 = Var(X_i)$. If (a) the X_i are uniformly bounded - that is, if for some M, $P\{|X_i| < M\} = 1$ for all i, and (b) $\sum_{i=1}^{\infty} \sigma_i^2 = \infty$ - then

$$P\left\{\frac{\sum_{i=1}^{n}(X_i - \mu_i)}{\sqrt{\sum_{i=1}^{n}\sigma_i^2}} \leq a\right\} \to \Phi(a) \text{ as } n \to \infty$$

Exercise-1

An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to be reasonably sure that his estimated distance is accurate to within ± 0.5 light-year?

Exercise-1

Suppose that the astronomer decides to make n observations. If $X_1, X_2, \cdots, X_n$ are the n measurements, then, from the central limit theorem, it follows that

$$Z_n = \frac{\displaystyle\sum_{i=1}^{n} X_i - nd}{2\sqrt{n}}$$

has approximately a standard normal distribution. Hence,

$$P\left\{-0.5 \leq \frac{\displaystyle\sum_{i=1}^{n} X_i}{n} - d \leq 0.5\right\} = P\left\{-0.5\frac{\sqrt{n}}{2} \leq Z_n \leq 0.5\frac{\sqrt{n}}{2}\right\}$$

Exercise-1

$$
\begin{aligned}
P\left\{-0.5\frac{\sqrt{n}}{2} \le Z_n \le 0.5\frac{\sqrt{n}}{2}\right\} &\approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) \\
&= 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1
\end{aligned}
$$

If the astronomer wants to be 95 percent certain that his estimated value is accurate to within 0.5 light-year, he should make n^* measurements, where n^* is such that

$$
2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = 0.95 \quad \text{or} \quad \Phi\left(\frac{\sqrt{n^*}}{4}\right) \approx 0.975
$$

From Table we obtain,

$$
\frac{\sqrt{n^*}}{4} = 1.96 \quad \text{or } n^* = (7.84)^2 \approx 61.47
$$

As n^* is not integral valued, he should make 62 observations.

The preceding analysis has been done under the assumption that the **normal approximation** will be a good approximation when $n = 62$. If the astronomer does not want to take a chance, he can still solve the problem by using Chebyshev's inequality. Since

$$E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = d; \quad Var\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{4}{n}$$

Hence, if he wants to be 95% certain, Chebyshev's inequality yields

$$P\left\{\left|\sum_{i=1}^{n} \frac{X_i}{n} - d\right| > 0.5\right\} \leq \frac{4}{n(0.5)^2} = \frac{16}{n} \leq 1 - 0.95 = 0.05$$

Thus, $n \geq 16/0.05 = 320$.

Exercise-2

The number of students who enroll in a psychology course is a Poisson random variable with mean 100. The professor in charge of the course has decided that if the number enrolling is 120 or more, he will teach the course in two separate sections, whereas if fewer than 120 students enroll, he will teach all of the students together in a single section. What is the probability that the professor will have to teach two sections?

Solution: let X be the Poisson random variable with $\lambda = 100$ which denotes the number of students enrolled in a psychology course. We are interested in

$$P\{X \geq 120\} = 1 - P\{X < 120\} = 1 - \sum_{i=0}^{119} e^{-100} \frac{100^i}{i!}$$

Exercise-2

We can use the central limit theorem to obtain an approximate solution. Recall,

$$X = X_1 + X_2 + \cdots + X_{100}$$

where X_i is a Poisson random variable each with parameter $\lambda_i = 1$. After continuity correction,

$$P\{X \geq 120\} = P\{X \geq 119.5\}$$

So,

$$P\{X \geq 119.5\} = P\left\{\frac{X - 100}{\sqrt{100}} \geq \frac{119.5 - 100}{\sqrt{100}}\right\} \approx 1 - \Phi(1.95) \approx 0.0256$$

Exercise-3

If 10 fair dice are rolled, find the approximate probability that the sum obtained is between 30 and 40, inclusive.

Solution:

$$E[X_i] = \frac{7}{2}, \quad Var(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{35}{12}$$

the central limit theorem yields

$$\begin{aligned} P\{29.5 \leq X \leq 40.5\} &= P\left\{ \frac{29.5 - 35}{\sqrt{\frac{350}{12}}} \leq \frac{X - 35}{\sqrt{\frac{350}{12}}} \leq \frac{40.5 - 35}{\sqrt{\frac{350}{12}}} \right\} \\ &\approx 2\Phi(1.0184) - 1 \approx 0.692 \end{aligned}$$

Exercise-4

Let X_i, $i = 1, \ldots, 10$, be independent random variables, each uniformly distributed over $(0, 1)$. Calculate an approximate to $P\left\{\sum_{i=1}^{10} X_i > 6\right\}$.

Solution: Since $E[X_i] = \frac{1}{2}$ and $Var(X_i) = \frac{1}{12}$. By the central limit theorem,

$$
\begin{aligned}
P\left\{\sum_{i=1}^{10} X_i > 6\right\} &= P\left\{\frac{\sum_{i=1}^{10} X_i - 5}{\sqrt{10\left(\frac{1}{2}\right)}} > \frac{6-5}{\sqrt{10\left(\frac{1}{2}\right)}}\right\} \\
&\approx 1 - \Phi(\sqrt{1.2}) \approx 0.1367
\end{aligned}
$$

Exercise-5

An instructor has 50 exams that will be graded in sequence. The times required to grade the 50 exams are independent, with a common distribution that has mean 20 minutes and standard deviation 4 minutes. Approximate the probability that the instructor will grade at least 25 of the exams in the first 450 minutes of work.

Solution: If we let X_i be the time that it takes to grade exam i, then

$$X = \sum_{i=1}^{25} X_i$$

is the time it takes to grade the first 25 exams. We are interested in

$$P\{X \leq 450\}.$$

Exercise-5

Now,

$$E[X] = \sum_{i=1}^{25} E[X_i] = 25(20) = 500$$

and

$$Var(X) = \sum_{i=1}^{25} Var(X_i) = 25(16) = 400$$

By the central limit theorem,

$$\begin{aligned} P\{X \le 450\} &= P\left\{\frac{X-500}{\sqrt{400}} \le \frac{450-500}{\sqrt{400}}\right\} \\ &\approx P\{Z \le -2.5\} \\ &= P\{Z \ge 2.5\} \\ &= 1 - \Phi(2.5) = 0.006 \end{aligned}$$

Thank You