

Lecture 17: The Law of Large Numbers and the Monte-Carlo method

The Law of Large numbers Suppose we perform an experiment and a measurement encoded in the random variable X and that we repeat this experiment n times each time in the same conditions and each time independently of each other. We thus obtain n independent copies of the random variable X which we denote

$$X_1, X_2, \dots, X_n$$

Such a collection of random variable is called a IID sequence of random variables where IID stands for *independent and identically distributed*. This means that the random variables X_i have the same probability distribution. In particular they have all the same means and variance

$$E[X_i] = \mu, \quad \text{var}(X_i) = \sigma^2, \quad i = 1, 2, \dots, n$$

Each time we perform the experiment n times, the X_i provides a (random) measurement and if the average value

$$\frac{X_1 + \dots + X_n}{n}$$

is called the *empirical average*. The Law of Large Numbers states for large n the empirical average is very close to the expected value μ with very high probability

Theorem 1. Let X_1, \dots, X_n IID random variables with $E[X_i] = \mu$ and $\text{var}(X_i)$ for all i . Then we have

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

In particular the right hand side goes to 0 as $n \rightarrow \infty$.

Proof. The proof of the law of large numbers is a simple application from Chebyshev inequality to the random variable $\frac{X_1 + \dots + X_n}{n}$. Indeed by the properties of expectations we have

$$E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n}E[X_1 + \dots + X_n] = \frac{1}{n}(E[X_1] + \dots + E[X_n]) = \frac{1}{n}n\mu = \mu$$

For the variance we use that the X_i are independent and so we have

$$\text{var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2}\text{var}(X_1 + \dots + X_n) = \frac{1}{n^2}(\text{var}(X_1) + \dots + \text{var}(X_n)) = \frac{\sigma^2}{n}$$

By Chebyshev inequality we obtain then

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

■

Coin flip I: Suppose we flip a fair coin 100 times. How likely it is to obtain between 40 and 60 heads? We consider the random variable X which is 1 if the coin lands on head and 0 otherwise. We have $\mu = E[X] = 1/2$ and $\sigma^2 = \text{var}(X) = 1/4$ and by Chebyshev

$$\begin{aligned} P(\text{between 40 and 60 heads}) &= P(40 \leq X_1 + \dots + X_{100} \leq 60) \\ &= P\left(\left|\frac{X_1 + \dots + X_{100}}{100} - \frac{1}{2}\right| \geq \frac{1}{10}\right) \\ &= 1 - P\left(\left|\frac{X_1 + \dots + X_{100}}{100} - \frac{1}{2}\right| \leq \frac{1}{10}\right) \\ &\geq 1 - \frac{1/4}{100(1/10)^2} = \frac{3}{4} \end{aligned} \tag{1}$$

Confidence interval: One can use Chebyshev to build a *confidence interval*. A typical statement is *a certain quantity is equal to x with a precision error of ϵ with a β % confidence*. Typical values for β are 95 % or 99% and the statement means

$$P(\text{value is in } (x - \epsilon, x + \epsilon)) \geq \frac{\beta}{100}$$

In the context of the LLN ϵ is the precision error and we have

$$\frac{\sigma^2}{n\epsilon^2} = 1 - \frac{\beta}{100}$$

From this we obtain the following important consequences

- **Precision error:** With a fixed confidence if we want to the precision error by a factor 10 we need 10^2 as many measurement.
- **Confidence error:** With a fixed precision if we want to the decrease the confidence error by a factor 10 (say go from a 90% to a 99% confidence interval) we need 10 as many measurement.

Coin flip II: I hand you a coin and make the claim that it is biased and that heads comes up only 48% of the times you flip it. Not believing me you decide you test the coin and since you intend to use that coin to cheat in a game you want to be sure with 95% confidence that the coin is biased. How many times should you flip that coin?

To do this you *assume* that the coin is biased and so pick a random X which is 1 if the coin lands on head and 0 otherwise. We have then $\mu = .48$ and $\sigma^2 = .48 \times .52 = .2496$. To test if the coin is biased we flip the coin n times and we choose a precision error of

$$\epsilon = .01$$

so that we are testing if the probability that the coin lands on head is between $(.47, .49)$, in particular it is biased. Using the law of large numbers we want to pick n such that

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - .48\right| \geq .1\right) \leq \frac{.2496}{n(.01)^2}$$

and to obtain a 95% confidence interval we need

$$\frac{.2496}{n(.01)^2} = \frac{1}{20}$$

that is n should be at least $.2496 \times 100^2 \times 20 = 49920$.

Polling: You read in the newspaper that after polling 1500 people it was found that 38% percent of the voting population prefer candidate A while 62% prefer candidate B . Estimate the margin of error of such poll with a 90% confidence.

We pick a random variable with $\mu = .38$ and so $\sigma^2 = .38 \times .62 = .2356$. By the LLN with $n = 1500$ we have

$$P\left(\left|\frac{X_1 + \cdots + X_{1500}}{1500} - 0.38\right| \geq \epsilon\right) \leq \frac{.2356}{1500\epsilon^2}$$

So if we set $\frac{1}{10} = \frac{.2356}{1500\epsilon^2}$ we find

$$\epsilon = \sqrt{\frac{10 \times .2356}{1500}} = 0.039$$

So with the margin of error is about 4% with a 90% confidence.

The Monte-Carlo method: Under the name Monte-Carlo methods, one understands an algorithm which uses randomness and the LLN to compute a certain quantity which might have nothing to do with randomness. Such algorithm are becoming ubiquitous

in many applications in statistics, computer science, physics and engineering. We will illustrate the ideas here with some very simple test examples (toy models).

Random number: A computer comes equipped with a *random number generator*, (usually the command **rand**, which produces a number which is uniformly distributed in $[0, 1]$. We call such a number U and such a number is characterized by the fact that

$$P(U \in [a, b]) = b - a \quad \text{for any interval } [a, b] \subset [0, 1].$$

Every Monte-Carlo method should be in principle constructed with Random number so as to be easily implementable. For example we can generate a 0 – 1 random variable X with $P(X = 1) = p$ and $P(X = 0) = 1 - p$ by using a random number. We simply set

$$X = \begin{cases} 1 & \text{if } U \leq p \\ 0 & \text{if } U > p \end{cases}$$

Then we have

$$P(X = 1) = P(U \in [0, p]) = p.$$

An algorithm to compute the number π : To compute the number π we draw a square with side length 1 and inscribe in it a circle of radius 1/2. The area of the square of 1 while the area of the circle is $\pi/4$. To compute π we generate a random point in the square. If the point generated is inside the circle we accept it, while if it is outside we reject it. The we repeat the same experiment many times and expect by the LLN to have a proportion of accepted points equal to $\pi/4$

More precisely the algorithm now goes as follows

- Generate two random numbers U and V , this is the same as generating a random point in the square $[0, 1] \times [0, 1]$.
- If $U^2 + V^2 \leq 1$ then set $X_1 = 1$ while if $U^2 + V^2 > 1$ set $X_1 = 0$.
- Repeat to the two previous steps to generate X_2, X_3, \dots, X_n .

We have

$$P(X = 1) = P(U_1^2 + U_2^2 \leq 1) = \frac{\text{area of circle}}{\text{area of the square}} = \frac{\pi/4}{1}$$

and $P(X = 0) = 1 - \pi/4$. We have then

$$E[X] = \mu = \pi/4 \quad \text{var}(X) = \sigma^2 = \pi/4(1 - \pi/4)$$

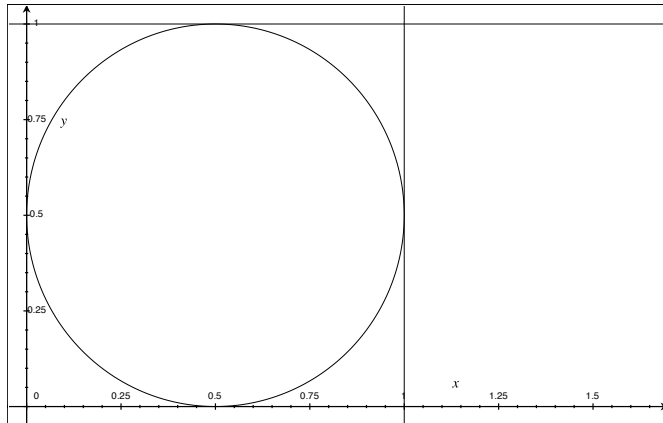


Figure 1:

So using the LLN we have

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\pi}{4}\right| \geq \epsilon\right) \leq \frac{\pi/4(1 - \pi/4)}{n\epsilon^2}$$

Suppose we want to compute π with a 4 digit accuracy and a 99% confidence. That is we want $\pi \pm .001$ that is $\pi \pm \frac{1}{1000}$ or $\frac{\pi}{4} \pm \frac{1}{4000}$. That is we take $\epsilon = \frac{1}{4000}$. To find an appropriate n we need then

$$\frac{\pi/4(1 - \pi/4)}{n\epsilon^2} \leq \frac{1}{100}.$$

Note that we don't know π so σ^2 is not known. But we note that the function $p(1 - p)$ on $[0, 1]$ has its maximum at $p = 1/2$ and the maximum is $1/4$. So we can take $\sigma^2 \leq 1/4$ and then we pick n such that

$$\frac{1/4}{n(1/4000)^2} \leq \frac{1}{100} \quad \text{or} \quad n \geq 400 \times 10^6$$

The Monte-carlo method to compute the integral $\int_a^b f(x) dx$. We consider a function f on the interval $[a, b]$ and we wish to compute

$$I = \int_a^b f(x) dx.$$

for some bounded function f . Without loss of generality we can assume that $f \geq 0$, otherwise we replace f by $f + c$ for some constant c . Next we can also assume that $f \leq 1$ otherwise we replace f by cf for a sufficiently small c . Finally we may assume that $a = 0$

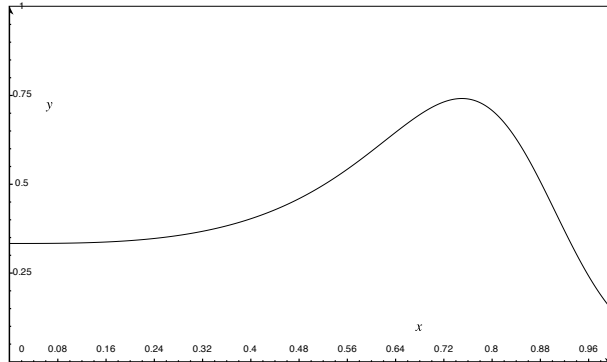


Figure 2: The graph of the function $y = \frac{e^{\sin(x^3)}}{3(1 + 5x^8)}$

and $b = 1$ otherwise we make the change of variable $y = (x - a)/(b - a)$. For example suppose we want to compute the integral

$$\int_0^1 \frac{e^{\sin(x^3)}}{3(1 + 5x^8)} dx$$

see figure

This cannot be done by hand and so we need a numerical method. A standard method would be to use a Riemann sum, i.e. we divide the interval $[0, 1]$ in subinterval and set $x_i = \frac{i}{n}$ then we can approximate the integral by

$$\int f(x)dx \approx \frac{1}{n} \sum_{i=1}^n f(x_i)$$

that is we approximate the area under the graph of f by the sum of areas of rectangles of base length $1/n$ and height $f(i/n)$.

We use instead a Monte-Carlo method. We note that

$$I = \text{Area under the graph of } f$$

and we construct a $0 - 1$ random variable X so that $E[X] = I$. We proceed as for computing π .

More precisely the algorithm now goes as follows

- Generate two random numbers U and V , this is the same as generating a random point in the square $[0, 1] \times [0, 1]$.
- If $V \leq f(U)$ then set $X_1 = 1$ while if $V > f(U)$ set $X_1 = 0$.

- Repeat to the two previous steps to generate X_2, X_3, \dots, X_n .

We have

$$P(X = 1) = P(V \leq f(U)) = \frac{\text{Area under the graph of } f}{\text{Area of } [0, 1] \times [0, 1]} = I = \int_0^1 f(x) dx$$

and so $E[X] = I$ and $\text{var}(X) = I(1 - I)$.

Exercise 1: How many people you include in a poll about preferences between two candidates if you want a precision of 1%? What about 0.5%?

Exercise 2: Suppose that you run the Monte-Carlo algorithm to compute π 10'000 times and observe 7932 points inside the circle. Find your estimation for the number π ? What is the accuracy of your observation with a 95% confidence? The answer should be in the form: based on my computation the number π belong to the interval $[a, b]$ with probability .95.

Exercise 3: Invent a Monte-carlo algorithm to compute the number e . You may consider for example a suitable integral