Central limit theorem

March 26, 2019

1 The theorem

Let $X_1, X_2, ..., X_n$ be random variables such that:

i Are equally distributed. It means there is a certain distribution D that depends on certain parameters λ_i fixed, such that for all X_i we have that:

$$X_i \sim D$$

(So we know the mean μ and the standard deviation σ for all the variables and it is the same).

ii The random variables are independent.

iii The standard deviation is finite.

Then we have that the variable that is the mean of all the variables:

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

is approximately distributed like a gaussian variable with parameters m and s where:

i $m = \mu$

ii $s = \frac{\sigma}{\sqrt{n}}$

So

$$\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Notes:

- The greater is n, the better is the approximation
- The distribution of all the variables can be any distribution, but it needs to be the same
- Pay attention to the notation. It is not the same $\overline{X}$ (the random variable for the mean), that $\overline{x}$ that is used to note (commonly) the sample mean.

2 Example

The bags of salt packed by a machine have $\mu = 500g$ and $\sigma = 35g$. The bags were packed in boxes of 100 units. Let's assume that each of the bags is distributed as a normal.

1 Calculate the probability that the average weight of the bags in a package is less than 495 g.

2 Calculate the probability that a box 100 of bags weights more than 51 kg.

We have here two different ways to use the CLT.

2.1 Calculate the probability that the average weight of the bags in a package is less than 495 g.

We are taking 100 units (a package) so here we are taking a sample with $n = 100$. We are approximating the average, so we know that the average is approximately a normal distribution with parameters

$$\mu = 500$$

and

$$\sigma = \frac{35}{\sqrt{100}}$$

So we know that the average $\overline{X}$ is, approximately:

$$\overline{X} \sim N(500, 3.5)$$

To know the probability we need:

$$P(\overline{X} < 495) = 0.0764$$

2.2 Calculate the probability that a box 100 of bags weights more than 51 kg

Here we need to define a new random variable: each box will have 100 bags, and we want to treat each box as a single unit. So we define a random variable for the box that will be the sum of 100 random variables, $\overline{Y}$.

Fortunately, we have the CLT to know the distribution of this new random variable. How? Because knowing the mean is the same as knowing the sum, we need just to multiply by n:

$$X_1 + X_2 + ... + X_n = n\frac{X_1 + X_2 + ... + X_n}{n} = n\overline{X}$$

And we "know" that if we multiply a normal

$$Z \sim N(\mu, \sigma)$$

by a single number n, then:

$$nZ \sim N(n\mu, n\sigma)$$

So for the CLT:

$$n\overline{X} \sim N(n\mu, \frac{n\sigma}{\sqrt{n}})$$

so

$$n\overline{X} \sim N(n\mu, \sigma\sqrt{n})$$

For our case:

$$\overline{Y} \sim N(100 \times 500, 35\sqrt{100})$$

and we want:

$$P(\overline{Y} > 51000) = 1 - P(\overline{Y} < 51000) = 0.0021$$