CENTRAL LIMIT THEOREM

Author: Josemari Sarasola

Gizapedia

gizapedia.hirusta.io

General explanations about the Central Limit Theorem

Central Limit Theorem (CLT) is very important in statistics, and that's why we call it *central*. It's frequently used in statistical inference, to deduce the distribution of the sample artimetic mean and other estimators of parameters. It has also interesting applications in practice, as we will in the problems we will solve. Le's go.

Do you remember (or just know) about the reproductivity of the normal distribution? That means that normal plus normal plus normal plus ...., is also a normal distribution (e.g., if daily sales distribute normal, sales along 4 days distribute also normal, if there is independence). So, we have a similar result with CLT but adding other distributions. E.g., uniform plus uniform plus ... distributes normal, but to get that result two important conditions must be held: first, the distributions we add up must be independent (as for reproductivity) and, this is new, the number of distributions we add up must be large; how much large? As a general rule, we will set 30 as the minimum number of distributions to add up in order to apply the CLT. For example, if daily sales distributes uniform (flat or equal probability in an interval), sales along 30 days distribute normal, if daily sales are independent, but we must have at least 30 days for that (actually, for the concrete case of uniform distributions we don't need so many distributions to add up, but generally we will set this minimum number of distributions: 30).

Now we know sum of many independent distributions is normal, but which are the parameters. Here you have the answers: the mean (mu) will be the sum of means, and the standard deviation (sigma) will be the square root of sum of variances. Just the same as in the reproductivity of normal distributions. For example if sales distribute $U(0,6)$, sales along 100 days distributes $N\left(\mu = 3 \times 100, \sigma = \sqrt{\frac{36}{12} \times 100}\right)$ (look for formulas of mean and variances of the continuous uniform). Sometimes, we will have to calculate the mean and deviation for each distribution we add up (problem 130 with solution). For other problems, the mean and deviation will be given (problem 132 with solution). Link for workbook is here.

Problems about CLT may be the same as those for the normal distribution:

- calculating a probability for the sum (problem 129a);
- calculating a maximum or minimum with a given probabilitaty for the sum (problem 129b)
- calculating the number of distributions to be added up in roder to reach a max or min with a given probability (problem 129d)

Students used to have problems with larger and smaller symbols inside the probability. Think deeply about each situation, we don't have a rule of thumb for that.

CLT applies not only for the sum of values (sales, productions, ..) but also for the arithmetic mean (better sample arithmetic mean) of values taken from distributions (or populations), when the sample size (no. of data) is large. For the mean, CLT states this, form the version we have stated above, μ and σ being the mean and deviation of the distributions we add up:

$$\sum_i x_i \sim N(n\mu, \sqrt{n\sigma^2}) \rightarrow \overline{x} \sim N\left(n\mu, \sqrt{\frac{\sigma^2}{n}}\right)$$

For an application of this, see problem 135.

In fact, in Internet you will find mostly the version of CLT according to the mean.

Thank you.