Contents

The Central Limit Theorem

What does this program do?

This program graphically demonstrates the central limit theorem.

It will take any initial distribution on a set of integers (initially from 0 to 6) and will calculate the distribution of the sum of 2 independent variables with this distribution.

It also will calculate the distribution of the sum of 5, 10, and 30 independent variables with the same distribution.

These distributions can be displayed graphically either as histograms or as cumulative distribution functions.

Additionally, together with the histogram the density function of the corresponding normal distribution function can be displayed. Similarly, together with the cumulative distribution function for the user's distribution the corresponding normal cumulative distribution function can be displayed.

How can I use the program?

When the program is started the initial distribution it displays assigns probability 0.1667 to the numbers from 1 to 6 each. Number 0 gets probability 0. These probabilities model a fair dice.

The numbers and probabilities are displayed in the spreadsheet on screen and in the histogram below the spreadsheet.

Pressing the Calculate button will calculate the probability distributions for sums of 2, 5, 10, and 30 independent variables with the same probability distribution and display the histograms for these distributions also.
You can change the probabilities of the initial distribution at any time. Just select any spreadsheet cell for the initial probabilities and change its value. Be careful, though, since the probabilities you enter always have to sum up to 1.
Pressing Calculate again recomputes the probabilities and displays the new histograms. Also double-clicking into any cell of the spreadsheet will recalculate the table and redisplay the histograms.

Pressing Cumulative changes the graphics display from histograms to cumulative distribution functions.

Pressing Histograms changes the graphics display from histograms to cumulative distribution functions.

Pressing Show Normal puts the graphical display of the corresponding normal distribution into all the graphical displays (in addition to the displays of the user's distribution).

Pressing Hide Normal hides the graphs for the normal distribution and only displays the graphs for the user's distribution.
It is possible to enter the maximal value which can be assigned a probability. From the menu choose Modify and then Maximal value. Then you can enter a new number for the maximal value.
It is also possible to do all the operations which can be started by pressing the buttons from the menus contained in the menu bar at the top of the screen.
Here are some probability distributions giving interesting results:

Value					
0	0.2	0.2	0.30	0.4	0.8
1	0.2	0.2	0.15	0.1	0.0
2	0.2	0.2	0.10	0.0	0.0
3	0.3	0.1	0.15	0.0	0.1
4	0.1	0.3	0.30	0.5	0.1

In some cases, especially for distributions with "holes", the density function might seem not to approximate the histogram too well. The user even might think that the graphics displayed contains an error. Switching to the cumulative distribution function display in these cases shows that the central limit theorem still works. The snadard formulation of this theorem refers to cumulatove distribution functions, not to histograms and density functions.

What is the mathematics behind this?

Let us look at a little example. We have 2 loaded (colored) dice giving the numbers 1 to 6 with different probabilities.

Number	Dice A	Dice B
1	0.15	0.20
2	0.15	0.20
3	0.15	0.20
4	0.20	0.20
5	0.20	0.10
6	0.15	0.10

Then the probability for getting a sum of 2 is 0.15*0.20=0.03.

The probability for getting a sum of 6 is obtaining by adding the probabilities for getting 1 and 5 or 2 and or 3 and 3 or 4 and 2 or 5 and 1. So to get the total probability for getting a sum of 6 we have to add 0.15*0.10+0.15*0.20+0.15*0.20+0.20*0.20+0.20*0.20=0.16
Generally to get a sum we have to find all combinations giving the same sum, multiply the probabilities for the parts of the combination, and add all these products.

The method can be generalized for sums of more than 2 terms. The program initially assumes that the basic experiments has the numbers 0, 1, 2, 3, 4, 5, and 6 as possible outcomes and then calculates the probabilities for summing 2 such experiments independently.

Furthermore the program also calculates the probabilities for all possible outcomes when the results of 5, 10, or 30 such basic experiments are added.