Central Limit Theorem

The central limit theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed. It means whatever the distribution of the population under study if you take the mean of samples then that "means" distribution will be normal.

For a while this theorem dominated statistics because it became almost a "straight-jacket" to statistical exploration – why bother if everything can be reduced to a "normal" distribution. But what does it mean in practice?

One day I'll have to try this experiment – there are dangers simulating because "random" number aren't actually random and contain within them deep patterns.

Let's suppose I have 10 bowls containing piles of tiles with the successive labels 1, 2, 3, 4, 5 6, 7, 8, 9, and 10. Now for each bowl I throw a die and label the bowl with the result – a number between one and six. Finally I get a giant bowl and scoop up one thousand tiles from each bowl **but multiplied by** the die throw result.

My final giant bowl might have three thousand 0's, eight thousand 1's, five thousand 2's etc. It will contain around thirty five thousand tiles which I mix up really well but who knows what actual distribution – not "normal", not "rectangular", not anything we can define.

Now the final part. I pull out ten tiles at a time, calculate the mean and record that. If I perform this experiment many times the central limit theorem tells me the distribution will approximate more and more to the "normal".

Does that conclusion "make sense"?

Suppose in the whole population I was to analyse the results of GCSE History by Grade. I am confident that the distribution would be "normal". But if I focus on one student I ask the question "Why did that student get that particular grade – let's say a grade 2". What reasons might explain that? I could come up with a very long list.

They spent too long playing or watching football
The parents aren't very bright and/or didn't encourage them.
They didn't need History to get their next career choice and didn't bother studying

Then there might be reasons like

Their pet dog died the day before
There was a house fire and the family were homeless for the previous month

So there are many reasons, each reason having a different contribution from none to significant on the individual but each of those reasons following some other distribution of occurrence which themselves each have multiple reasons for occurring. What a total mess YET the net effect is that the outcomes when themselves grouped in samples approximately follow a "normal" distribution

It's a very deep theorem. Out of chaos naturally arises some precise mathematical order. Is this what fundamentally drives the whole Universe from a state of chaos to a state of order?

This phenomenon was first discovered by Francis Galton, a cousin of Charles Darwin and a pioneer of statistics. He called it the "supreme law of unreason" and wrote:

"Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along."

The central limit theorem has profound implications for science, engineering, and everyday life. It explains why we can make reliable predictions and inferences based on limited data, such as polls, surveys, experiments, or measurements. It also shows how randomness can give rise to order and stability, such as in the case of coin flips, dice rolls, or lottery draws. As Pearl and Mackenzie[#] put it: "This is the same law that makes insurance companies so profitable, despite the uncertainties in human affairs."

rg

Judea Pearly and Dana Mackenzie "The Book of Why : The New Science of Cause and Effect"

Central Limit Theorem

The central limit theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed. It means whatever the distribution of the population under study if you take the mean of samples then that "means" distribution will be normal.

For a while this theorem dominated statistics because it became almost a "straight-jacket" to statistical exploration – why bother if everything can be reduced to a "normal" distribution. But what does it mean in practice?

One day I'll have to try this experiment – there are dangers simulating because "random" number aren't actually random and contain within them deep patterns.

Let's suppose I have 10 bowls containing piles of tiles with the successive labels 1, 2, 3, 4, 5 6, 7, 8, 9, and 10. Now for each bowl I throw a die and label the bowl with the result – a number between one and six. Finally I get a giant bowl and scoop up one thousand tiles from each bowl **but multiplied by** the die throw result.

My final giant bowl might have three thousand 0's, eight thousand 1's, five thousand 2's etc. It will contain around thirty five thousand tiles which I mix up really well but who knows what actual distribution – not "normal", not "rectangular", not anything we can define.

Now the final part. I pull out ten tiles at a time, calculate the mean and record that. If I perform this experiment many times the central limit theorem tells me the distribution will approximate more and more to the "normal".

Does that conclusion "make sense"?

Suppose in the whole population I was to analyse the results of GCSE History by Grade. I am confident that the distribution would be "normal". But if I focus on one student I ask the question "Why did that student get that particular grade – let's say a grade 2". What reasons might explain that? I could come up with a very long list.

They spent too long playing or watching football
The parents aren't very bright and/or didn't encourage them.
They didn't need History to get their next career choice and didn't bother studying

Then there might be reasons like

Their pet dog died the day before
There was a house fire and the family were homeless for the previous month

So there are many reasons, each reason having a different contribution from none to significant on the individual but each of those reasons following some other distribution of occurrence which themselves each have multiple reasons for occurring. What a total mess YET the net effect is that the outcomes when themselves grouped in samples approximately follow a "normal" distribution

It's a very deep theorem. Out of chaos naturally arises some precise mathematical order. Is this what fundamentally drives the whole Universe from a state of chaos to a state of order?

This phenomenon was first discovered by Francis Galton, a cousin of Charles Darwin and a pioneer of statistics. He called it the "supreme law of unreason" and wrote:

"Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along."

The central limit theorem has profound implications for science, engineering, and everyday life. It explains why we can make reliable predictions and inferences based on limited data, such as polls, surveys, experiments, or measurements. It also shows how randomness can give rise to order and stability, such as in the case of coin flips, dice rolls, or lottery draws. As Pearl and Mackenzie[#] put it: "This is the same law that makes insurance companies so profitable, despite the uncertainties in human affairs."

rg

Judea Pearly and Dana Mackenzie "The Book of Why : The New Science of Cause and Effect"