Central Limit Theorem

David Carlson

Feb. 28, 2022

CLT Overview

- Incredibly important, fundamental theorem in probability theory

CLT Overview

- Incredibly important, fundamental theorem in probability theory
- Used in many cases, perhaps most importantly in regression

CLT Overview

- Incredibly important, fundamental theorem in probability theory
- Used in many cases, perhaps most importantly in regression
- Establishes that when independent random variables are summed, their properly normalized sum tends towards a normal distribution *irrelevant* of the variables' distributions

CLT Overview

- Incredibly important, fundamental theorem in probability theory
- Used in many cases, perhaps most importantly in regression
- Establishes that when independent random variables are summed, their properly normalized sum tends towards a normal distribution *irrelevant* of the variables' distributions
- Implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions

More Formally

- Let $X_1, X_2, \ldots, X_n$ are n random samples drawn from a population with overall mean μ and *finite* variance σ^2

More Formally

- Let $X_1, X_2, \ldots, X_n$ are n random samples drawn from a population with overall mean μ and *finite* variance σ^2
- Let $\bar{X}_n$ be the sample mean

More Formally

- Let $X_1, X_2, \ldots, X_n$ are n random samples drawn from a population with overall mean μ and *finite* variance σ^2
- Let $\bar{X}_n$ be the sample mean
- Let $Z = \lim_{n\to\infty} \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$

More Formally

- Let $X_1, X_2, \ldots, X_n$ are n random samples drawn from a population with overall mean μ and *finite* variance σ^2
- Let $\bar{X}_n$ be the sample mean
- Let $Z = \lim_{n\to\infty} \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$
- Z is a *standard* normal distribution

Example

- Suppose that a sample is obtained containing many observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic mean of the observed values is computed

Example

- Suppose that a sample is obtained containing many observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic mean of the observed values is computed
- If this procedure is performed many times, the central limit theorem says that the probability distribution of the average will closely approximate a normal distribution

Example

- Suppose that a sample is obtained containing many observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic mean of the observed values is computed
- If this procedure is performed many times, the central limit theorem says that the probability distribution of the average will closely approximate a normal distribution
- A simple example of this is that if one flips a coin many times, the probability of getting a given number of heads will approach a normal distribution, with the mean equal to half the total number of flips

Example

- Suppose that a sample is obtained containing many observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic mean of the observed values is computed
- If this procedure is performed many times, the central limit theorem says that the probability distribution of the average will closely approximate a normal distribution
- A simple example of this is that if one flips a coin many times, the probability of getting a given number of heads will approach a normal distribution, with the mean equal to half the total number of flips
- At the limit of an infinite number of flips, it will equal a normal distribution

Regression and CLT

- Regression analysis and in particular ordinary least squares specifies that a dependent variable depends according to some function upon one or more independent variables, with an additive error term

Regression and CLT

- Regression analysis and in particular ordinary least squares specifies that a dependent variable depends according to some function upon one or more independent variables, with an additive error term
- Various types of statistical inference on the regression assume that the error term is normally distributed

Regression and CLT

- Regression analysis and in particular ordinary least squares specifies that a dependent variable depends according to some function upon one or more independent variables, with an additive error term
- Various types of statistical inference on the regression assume that the error term is normally distributed
- This assumption can be justified by assuming that the error term is actually the sum of many independent error terms; even if the individual error terms are not normally distributed, by the central limit theorem their sum can be well approximated by a normal distribution

Illustration

http://195.134.76.37/applets/AppletCentralLimit/Appl_CentralLimit2.html