Shiny Application Help Document

The objectives of this project are to

- Illustrate how common distributions may be simulated by using computers
- Validate the Central Limit Theorem by sampling through observations
- Gain a better understanding of the statistics behind different distributions

What is the Central Limit Theorem (CLT) and why is it important?

- CLT states that the means of sample means always follow a normal distribution regardless of the distribution of original samples.
- CLT is important because it means that we could estimate the mean of any samples, even if we know little about their distribution.

Features

- This application simulates the following common distributions:
  - normal distribution
  - poisson distribution
  - binomial distribution
  - Bernoulli distribution
  - negative binomial distribution
  - uniform distribution
  - exponential distribution
- The user has the ability to control all parameters of the distributions, for example:
  - normal distribution - mean, SD
  - poisson distribution - lambda
  - binomial distribution - trails and probability
  - Bernoulli distribution - probability
  - negative binomial distribution - Dispersion
  - uniform distribution - min and max
  - exponential distribution - rate
- The CLT normally refers to the mean of independent random variables. Here, we will also test if it applies to the variance of the observations.

Tutorials

- The first step is to select a distribution at the left "selection" panel. Only one distribution may be selected at a time.
- A reactive function in the server responds to the selection by updating the parameters specific for each distribution.
- The user then select the parameters, such as the sample size ("observation"), the mean and the SD, etc.
- In response to the user-selected parameters, the server plots the histogram and probability of the distributions at left bottom.

- The red vertical line denotes the mean of the observations. The actual mean and variance are shown at the bottom of the plot.
- To simulate CLT, the user selects the sampling size (default 40, which should be less than the total observations), and sampling times.
- The mean of the sampled observations is updated in the middle panel.
- The variance of the sampled observations is updated at the right panel.
- It is evident that the mean and variance of the original sample population corresponds to the mean and variance of simulated sample mean and variance, respectively.
- While CLT has been related to the mean of observation, it may also applies to the variance of the samples at least for some non-normal distributions.

The following files are available at https://github.com/zhangj5/CLT_simulation

- ui.R
- server.R
- document.pdf

To deploy the server in your own computer, follow steps:

- open R console
- load “shiny” library
- type “runApp()”