**深 圳 大 学 实 验 报 告**

**课程名称：­ 概率论与数理统计**

**实验项目名称:** Application of Central Limit Theorem

**学院： 电子与信息工程学院**

**专业： 电子信息工程**

**指导教师： 陈昌盛**

**报告人： 杨烨 学号**： **2022280380**

**班级： 06**

**实验时间： 2023年11月10日**

**实验报告提交时间： 2023年11月30日**

**教务处制**

|  |
| --- |
| Aim of Experiment:  1. Familiar with the central limit theorem.  2. Understand the implementation of the central limit theorem in python.  3. Know how to visualize data in different distributions.  4. Familiar with seaborn, a powerful visual database in python; |
| Experiment Content:  The Central Limit Theorem (CLT) is often referred to as one of the most important theorems, not only in probability & statistics but also in the sciences as a whole.  Try a Python simulation to understand the nature of the central limit theorem. |
| Experiment Process：  Firstly, we should be familiar with the relevant background knowledge of the Central Limit Theorem and the knowledge points needed for this experiment. Samples and the Sampling Distribution, Central Limit Theorem - Statement & Assumptions    Fig1 the Central Limit Theorem  This experiment is divided into **three sections:**  (1) **Demonstration of CLT in action using simulations in Python；**  ①Experiment 1 - Exponentially distributed population  ②Experiment 2 - Binomially distributed population  (2) **An Application of CLT in Investing/Trading；**  (3) **More visualization exercises；**  ①Hourly temperature records in Detroit；  ②Hourly humidity, temperature, air pressure, and wind speed records in Detroit；  ③Temperature change curves of four cities；  **The following is the specific process of this experiment：**   1. **Demonstration of CLT in action using simulations in Python:**   ①Experiment 1 - **Exponentially distributed** population:    Fig2 Density function of exponential distribution  Suppose that **θ=4**；Calculate the mean and variance of the exponential distribution.  Implementation code：    Fig3 Code to calculate the mean and variance  Keep the original form and use the **Dataframe** data type in the pandas library to implement the operation, Plot of an exponential distribution sample with a sample size of 500.    Fig4 Code to calculate the mean and variance  ②Experiment 2 - **Binomially distributed** population:    Fig5 Binomially distributed  Suppose that k=30 and p=0.9；  Calculate the mean and variance of the exponential distribution：  Implementation code：    Fig6 Code to calculate the mean and variance   1. **An Application of CLT in Investing/Trading:**   In financial models, we often use relevant knowledge of central limit theorem for investment and return. For analysis, we first import some standard Python libraries and obtain daily closing price ITC stock data from yfinance library.  Use 1\*2 subgraphs to plot the result:    Fig7 Implement 1\*2 subgraphs to present data visualization code   1. **More visualization exercises；**   A continuous random variable represents an infinite number of possible outcomes. For a sample of a continuous random variable 𝑋, we may not be able to cover the entire sample space. Since the sample space has an infinite number of observations, we cannot accurately estimate its exact distribution, so a parameterized continuous distribution can be used to approximate the observed distribution. In this task, we will use hourly weather data sets for the city of Detroit to approximate temperature records in a continuous distribution.  ①Hourly temperature records in **Detroit**；    Fig8 A code that presents hourly temperature data for Detroit  ②Hourly humidity, temperature, air pressure, and wind speed records in Detroit；    Fig 9 A code that presents hourly humidity, temperature, air pressure,  and wind speed data for Detroit  ③Temperature change curves of four cities；    Fig10 Code that shows the temperature curve of four cities |
| Data Logging and Processing:  **(1) Demonstration of CLT in action using simulations in Python:**  ①Experiment 1 - Exponentially distributed population    Fig11 Exponential distribution of small sample size (Size:2)    Fig12 Exponential distribution of large sample size (Size: 500)  We can observe that the mean of all the sample means is very close to the population: mean (μ=4):    Fig13 The result of the calculation of the average of the first five samples  Similarly, we can observe that the standard deviation of the 50 sample means is quite close to the value stated by the CLT, :    Fig14 The result of standard deviation calculation  ②Experiment 2 - Binomially distributed population:  IMG_256  Fig15 Binomially distributed of large sample size (Size: 500)  The sampling distribution should be approximately normal with mean population:  **Theoretical value:**  mean = 27;  standard deviation =0.0734;  **Actual value:**  Mean of sample means: 26.99712;  Standard deviation of sampling distribution: 0.0779756; **(2)An Application of CLT in Investing/Trading:** Taking a peek at the fetched data:    Fig16 First five data  Now that we have the daily log returns for ITC. Visualize both the returns and their distribution according to the diagrams below:  IMG_256  Fig17 Daily Closing price - Visualization of closing price ITC stock data  **(3)More visualization exercises；**  ①Hourly temperature records in Detroit；  IMG_256  Fig18 Hourly temperature records in Detroit；  ②Hourly humidity, temperature, air pressure, and wind speed records in Detroit；  IMG_256  Fig19 Hourly humidity, temperature, air pressure, and wind speed records in Detroit；  ③Temperature change curves of four cities；  IMG_256  Fig 20 Temperature change curves of four cities |
| Experimental Results and Analysis:  The experimental results are as follows:  In this experiment, I successfully realized the sampling analysis of exponential distribution and binomial distribution, drew the histogram of relative frequency distribution and calculated the corresponding mean and variance; We also successfully completed the data analysis of financial models and the visualization of different weather indicators for various cities in the United States.  Analysis:  There are also some parts of this experiment that can be improved:  Deeper theoretical exploration: For each distribution and model, the mathematical and statistical theory behind it can be further explored to better understand its application scenarios and limitations.  More comprehensive data analysis: In terms of financial models and weather indicator visualization, more data sets and variables can be explored and more comprehensive data analysis and visual presentation can be performed.  More detailed experimental design: When sampling analysis is performed, more experimental sessions and different parameter combinations can be designed to more fully explore the nature and characteristics of these distributions. |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内。