

基于MATLAB的抽样分布理论实验开发

任梦茹, 张子怡, 黄梦瑶, 皇甫佳欣, 郭海刚

河南科技大学数学与统计学院, 河南 洛阳

收稿日期: 2024年12月20日; 录用日期: 2025年1月11日; 发布日期: 2025年1月23日

摘要

本文利用MATLAB的界面设计功能对抽样分布理论构建了一个模拟实验界面, 以可视化动态演示的形式呈现抽象的理论, 将几种常见抽样分布的频数直方图与概率密度函数进行比较。在该界面中, 学生可以自主进行实验操作, 修改参数, 观察多种情况下的实验结果。本项目实现了用户与界面的互动, 省略了复杂的证明过程, 从而帮助学生更好地理解抽样分布理论。

关键词

抽样分布, 卡方分布, 动态演示, 检验

Experimental Development of Sampling Distribution Theory Based on MATLAB

Mengru Ren, Ziyi Zhang, Mengyao Huang, Jiaxin Huangfu, Haigang Guo

School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang Henan

Received: Dec. 20th, 2024; accepted: Jan. 11th, 2025; published: Jan. 23rd, 2025

Abstract

This paper uses the interface design function of MATLAB to build a simulation experiment interface for the theory of sampling distribution, presenting the abstract theory in a visual and dynamic demonstration form. It compares the frequency histograms and probability density functions of several common sampling distributions. In this interface, students can conduct experimental operations autonomously, modify parameters, and observe the experimental results under various conditions. This project achieves interaction between users and interfaces, omits complex proof processes, and helps students better understand the theory of sampling distribution.

文章引用: 任梦茹, 张子怡, 黄梦瑶, 皇甫佳欣, 郭海刚. 基于 MATLAB 的抽样分布理论实验开发[J]. 统计学与应用, 2025, 14(1): 194-201. DOI: 10.12677/sa.2025.141019

Keywords

Sampling Distribution, Chi Square Distribution, Dynamic Demonstration, Test

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).

<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

1. 引言

抽样分布理论作为统计学的基础理论之一，是参数估计和假设检验的基础，对于实际数据的分析和推断具有重要意义。然而，在传统的统计学教育中，抽样分布理论的学习往往面临着直观理解上的困难。由于该理论涉及到复杂的数学推导和较为抽象的概念理解，学生往往难以将理论知识与实际问题相结合，进而影响了他们学习的积极性。在以往的研究中，李秀敏、徐凌云学者在文献中利用 R 软件对服从正态总体分布的均值进行分析，作出直方图，并与抽样理论进行对比[1]；彭维湘学者在文献中利用 excel 生成随机数，抽取样本，计算样本统计量得出样本统计量分布，与理论分布对比[2]。而在本项目中，我们利用 MATLAB 的 App Designer 工具，设计了一款直观的界面来辅助学生理解和应用抽样分布理论，该界面会对实验结果给出生动的动态演示。此外，在该界面中，学生可以自主进行实验操作，通过设置不同的参数观察不同情况下的实验结果，对抽样分布理论有更深刻的理解。这能够省略抽样分布理论复杂的证明过程，从而降低了学生对抽样分布理论学习的难度。

本项目的核心工作是构建一个基于蒙特卡洛方法的模拟界面，利用 App Designer 设计用户友好的图形绘制界面，该界面可以绘制样本统计量随着实验次数变化的动态图像、样本统计量的频数直方图和对理论分布的概率密度函数，用户可调整抽样数据参数，观察不同情况下理论图像与实验数据图像的差异，从而直观地理解抽样分布的特性。此外，程序还能对分布进行检验，为学生提供更加全面的数据分析视角。

2. 实验设计

2.1. 样本统计量的分布

在统计学中，把研究对象的全体称为“总体”，把组成总体的每一个基本单位称为个体。“样本”是指从总体中选取出来的一部分个体，用于代表和研究总体的特征。“统计量”是样本的函数，它是对样本信息进行加工、提炼后得到的量，用于描述样本的特征，进而推断总体的性质。例如样本均值、样本方差等都是常见的统计量[3]。

统计量的分布称为抽样分布，主要讨论正态总体下样本均值、样本方差等统计量的分布。常用的抽样分布有正态分布、卡方分布、t 分布以及 F 分布等，它们是各种参数估计、推断和假设检验的重要理论基础。抽样分布理论如下[4]：

卡方分布 设 $X_1, X_2, \dots, X_n$ 是一个来自正态总体 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ 和 } S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$$

则

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

t 分布 设 $X_1, X_2, \dots, X_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的一个样本, $\bar{X}$ 与 S^2 分别是该样本的样本均值与样本方差, 则有

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1)$$

F 分布 设 $X_1, X_2, \dots, X_m$ 是来自 $N(\mu_1, \sigma_1^2)$ 的样本, $Y_1, Y_2, \dots, Y_n$ 是来自 $N(\mu_2, \sigma_2^2)$ 的样本, 且此两样本相互独立, 记

$$S_x^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \quad S_y^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2$$

其中

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j$$

则有

$$F = \frac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2} \sim F(m-1, n-1)$$

特别地, 若 $\sigma_1^2 = \sigma_2^2$, 则

$$F = S_x^2/S_y^2 \sim F(m-1, n-1)$$

2.2. 实验设计思路

由于各抽样分布理论本质相同, 实验设计思路相似, 本文仅以卡方分布为例, 简述其随机模拟实验思路[4]:

首先, 从正态总体中抽取样本, 计算样本均值和样本方差, 构造样本统计量, 重复该步骤, 得到多个样本统计量的值; 其次, 绘制统计量的频数直方图并与对应理论的概率密度图进行比较; 最后, 对实验结果进行检验, 以确保实验的准确性。模拟实验思路图如图 1 所示。

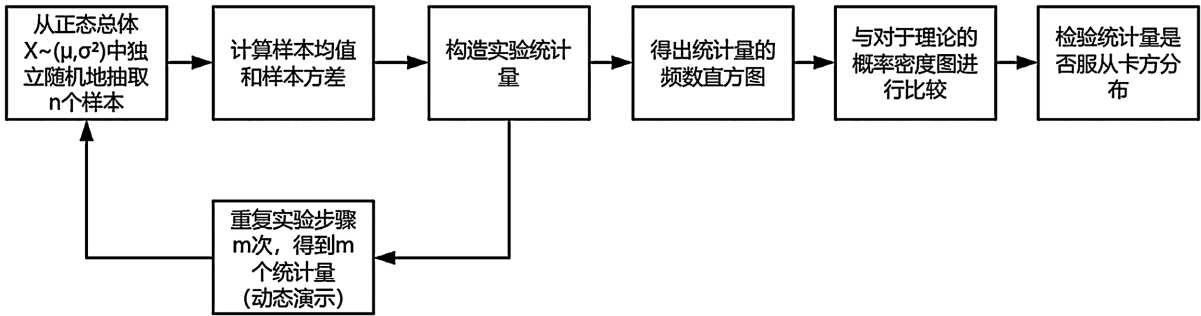


Figure 1. Design idea diagram of Chi-square distribution simulation experiment
图 1. 卡方分布模拟实验设计思路图

2.3. 软件实现

本文的计算机随机模拟算法由 MATLAB 数学软件实现, 软件的设计过程主要包括两大部分: 用户界

面(App Designer)设计和回调函数设计[5]。

2.3.1. 用户界面设计

本文运用 MATLAB 的 App Designer 函数设计功能, 设计了一个简易的操作界面, 实现用户与界面的交互性。界面设计主要分三个模块: ①菜单封面; ②理论界面; ③可视化操作及效果图。

(1) 菜单界面

该界面主要是由一个图像组件、一个文本框和五个按钮组件构成, 如图 2 所示。菜单封面主要展示了抽样分布的分布类型, 主要有四种: 正态分布、t 分布、F 分布和卡方分布, 通过点击对应的分布类型, 跳转到相应的理论界面。

(2) 卡方分布界面

以卡方分布为例, 该界面由一个图像组件和三个按钮组件构成。图像组件展示了卡方分布的理论分布, 三个按钮分别命名为理论、演示和返回。点击理论按钮可展示定理内容如图 3(b)所示, 点击演示按钮可跳转至可视化操作及效果图界面, 点击返回按钮转至菜单界面。

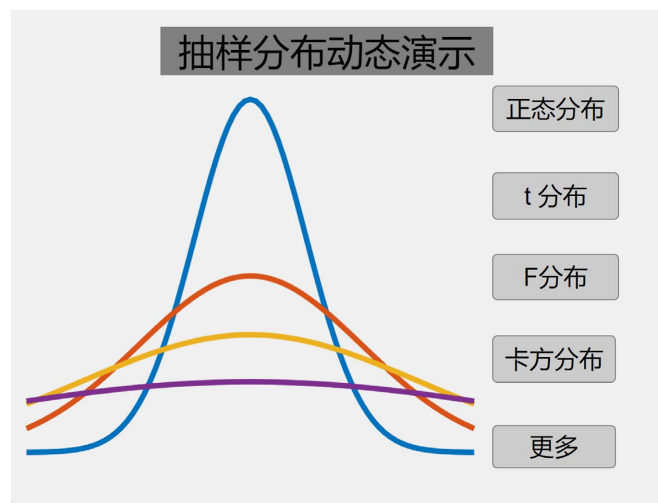


Figure 2. Sampling distribution interface

图 2. 抽样分布界面

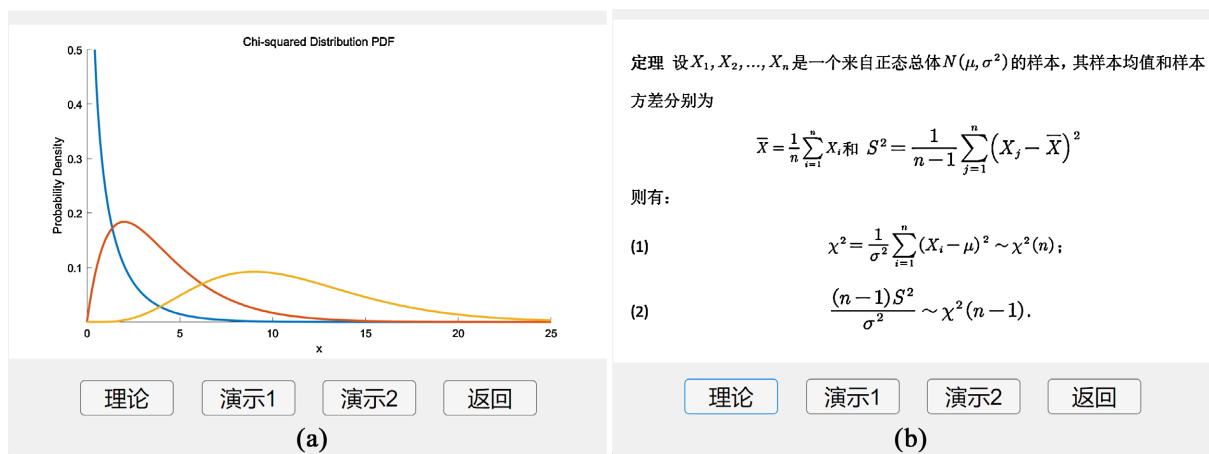


Figure 3. Interface of Chi-square distribution theory

图 3. 卡方分布理论界面

(3) 卡方分布实验界面

设计实验界面，以卡方分布为例，运行界面设计样式如图 4。

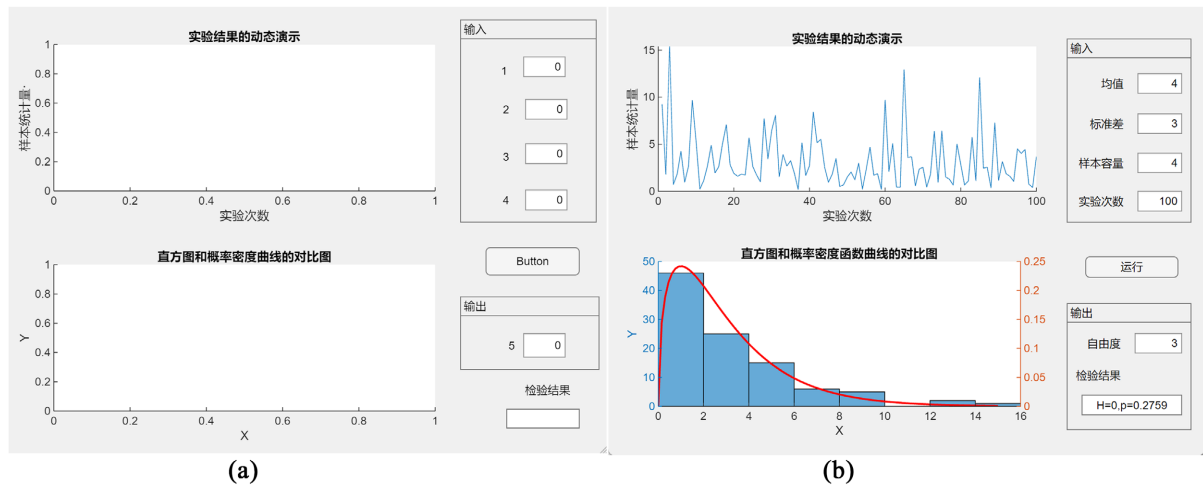


Figure 4. Visual operation interface

图 4. 可视化操作界面

该界面由两个坐标轴对象、两个面板、一个按钮组件和五个编辑字段组件构成，如图 4(a)所示。两个坐标区分别用于显示样本统计量随实验次数增加而变化的动态图像和频数直方图与对应理论概率密度曲线的对比图。将编辑字段组件编号为 1、2、3、4、5，分别匹配用于输入样本均值，标准差，样本容量，实验次数，自由度。两个面板将组件分组为输入、输出。将按钮组件命名为“运行”，用于运行程序绘图。

在实验界面中，界面会给出初始值，可以直接单击“运行”按钮，绘制动态图形，直方图和概率密度曲线，输出自由度。也可以在编辑框中进行修改，分别输入样本均值、标准差、样本容量、实验次数，点击按钮，运行程序。运行按钮的部分回调函数如下：

```
.....
if n>=8    %n 为样本容量
errorDlg("样本容量建议小于 8","错误提示','error'); %错误提示
end
for i = 1:k
samples = normrnd(mu, sigma, [n, 1]);
v(i) = var(samples);
y(i)=(n-1)*v(i)/(sigma*sigma);
plot(app.UIAxes,y(1:i)) %实验结果动态演示
drawnow
end
yyaxis(app.UIAxes_2,"left")
histogram(app.UIAxes_2, y,'normalization','count')%绘制直方图
n = app.ss.Value-1;
x = linspace(0, 3*n); %计算 x 范围
pdf_values = chi2pdf(x, n);
```

```

yyaxis(app.UIAxes_2,"right")
plot(app.UIAxes_2, x, pdf_values,'r','LineWidth',1.5); %绘制卡方分布概率密度函数图像
xlim(app.UIAxes_2,[0,inf]); %设置 x 轴范围
app.DF.Value=n; %输出自由度
.....

```

2.3.2. 结果分析及检验

当总体 X 服从均值为 4，标准差为 3 的正态分布，取样本容量 n 等于 4，实验次数为 100，运行程序实验结果如图 4(b)所示。上方坐标区中展示了每次实验卡方统计量的值，下方坐标框中红色曲线表示自由度为 3 的卡方分布的概率密度函数图像，蓝色直方图表示卡方统计量的频数分布，由图可知二者分布基本吻合。

经多次测试，可得实验次数越大，样本统计量的频数直方图和对应该理论分布的概率密度曲线吻合程度越高，如图 5(a)；为区别卡方分布正态分布概率密度函数图像，当输入的样本容量大于等于 8 时，会弹出提示框，如图 5(b)所示。

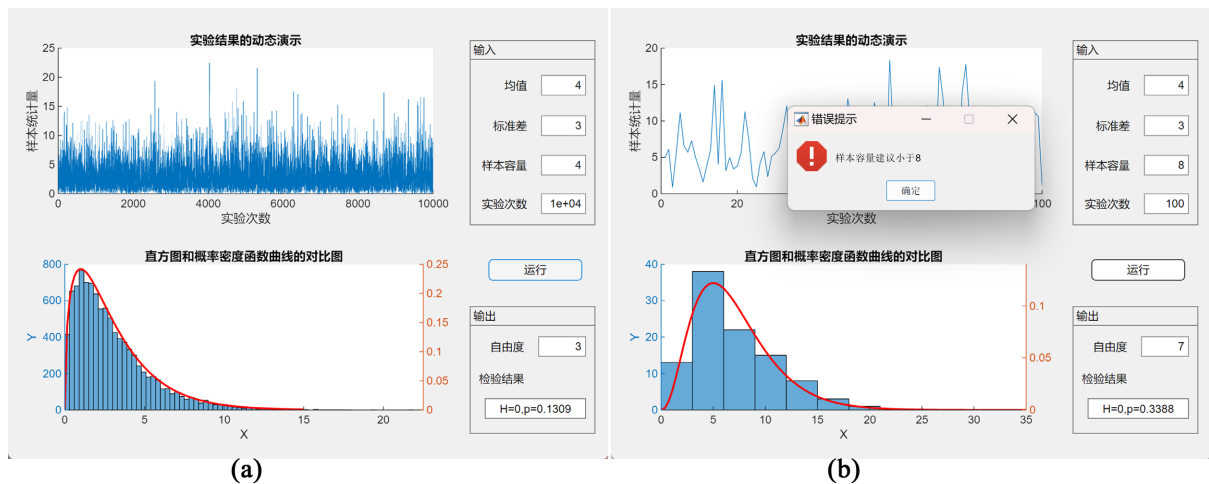


Figure 5. Analysis of results
图 5. 结果分析

通过图像可以直观判断样本统计量大致服从卡方分布，但并不严谨，为了确保该实验的准确性，进一步地，对实验结果进行卡方检验，部分回调函数如下：

```

.....
[H,P,stats]=chi2gof(y,'CDF',@(x) chi2cdf(x,n2),'NParams',1);
app.Label_3.Text=sprintf('H=%f,p=%f',H,P);
if n<8
if H==0
errordlg('服从卡方分布','检验结果','error');
else
errordlg('不服从卡方分布','检验结果','error');
end
end
end

```

.....

在回调函数中,我们利用 `chi2gof` 函数对样本统计量进行卡方检验,默认显著性水平 α 为 0.05,然后将检验的 p 值与显著性水平进行比较,判断数据是否服从卡方分布。

在给定的条件下运行一次的判断结果如图 6 所示,此时 MATLAB 得出的 H 值为 0, p 值为 0.2759, p 值大于显著性水平,所以接受原假设,即数据服从卡方分布。

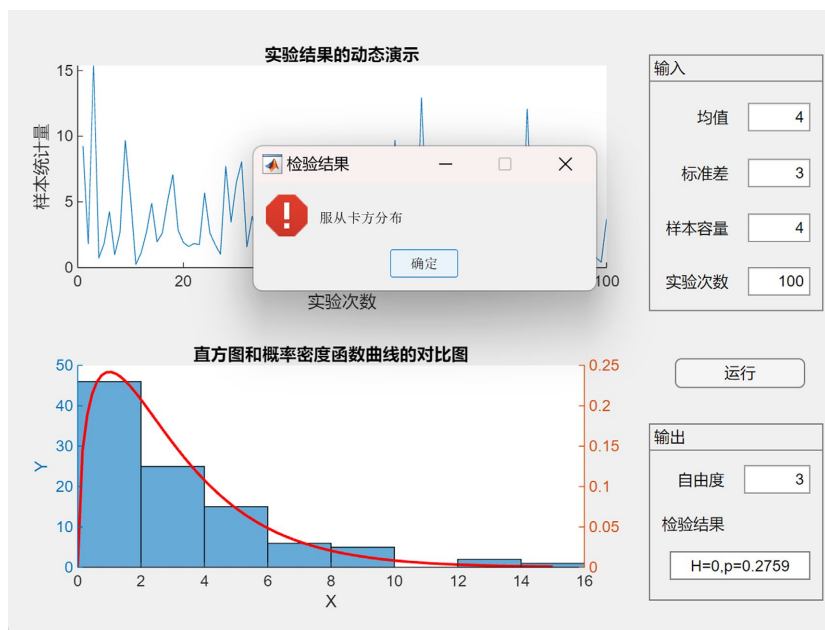


Figure 6. Chi-square test results
图 6. 卡方检验结果

除卡方检验外,我们还运用了 Kolmogorov-Smirnov 检验方法对卡方分布的实验结果进行检验,部分回调函数如下:

.....

```
cdf= chi2cdf(y,n2);
[H1,p1]=kstest(y,[y, cdf] );
app.Label_4.Text=sprintf('H=%f,p=%4f',H1,p1);
```

.....

在回调函数中,我们利用 `kstest` 函数对样本统计量进行 Kolmogorov-Smirnov 检验,默认显著性水平 α 为 0.05,然后将检验的 p 值与显著性水平进行比较,判断数据是否服从卡方分布。

在给定的条件下运行一次的判断结果如图 7 所示,此时 MATLAB 得出的 H 值为 0, p 值为 0.5216, p 值大于显著性水平,所以接受原假设,即数据服从卡方分布。

3. 结语

本文利用 MATLAB 的 App Designer 直观展示了抽样分布中重要分布之一卡方分布的一个定理,利用蒙特卡洛方法,借助 MATLAB 的界面设计功能直观地呈现了较为抽象的理论内容。通过调整实验参数,观察不同条件下抽样分布的变化规律,这对于学生深入理解抽样分布理论具有重要意义,能帮助学生进一步掌握相关概念、知识点,增加统计学习的趣味性与互动性,调动学生学习的积极性,同时这一

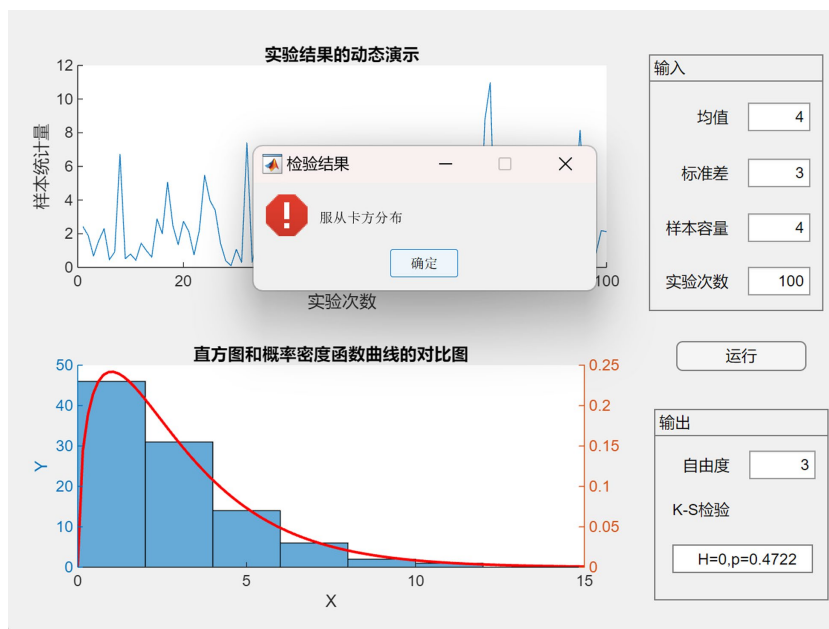


Figure 7. Kolmogorov-Smirnov test results

图 7. Kolmogorov-Smirnov 检验结果

教学方法的创新应用,也为教师的课堂教学提供了更为便捷且高效的工具,有助于提升教学质量与效率。

基金项目

河南科技大学大学生创新创业训练计划项目(2024236),河南科技大学实验技术开发基金(SY2324041)。

参考文献

- [1] 李秀敏, 徐凌云. 用随机模拟方法研究抽样分布问题[J]. 高师理科学刊, 2018, 38(3): 62-65.
- [2] 彭维湘. 抽样分布的模拟试验方法与应用[J]. 统计与决策, 2008(11): 21-22.
- [3] 武新乾. 概率论与数理统计[M]. 第3版. 北京: 科学出版社, 2023: 153-156.
- [4] 茆诗松. 概率论与数理统计教程[M]. 第3版. 北京: 高等教育出版社, 2019: 232-234.
- [5] 林小苹, 吴文杰. 用 MATLAB 模拟大数定律和中心极限定理[J]. 汕头大学学报(自然科学版), 2005(2): 12-18+80.