

# 概率论与数理统计实验实验报告

日期：2024年12月4日

班级	姓名	学号
智科2301	张皓博	302023511143

## 一、实验名称

参加家长会人数问题

## 二、问题背景描述

在学校组织家长会时，对于单个学生，其参加家长会的家长人数是随机变量。已知无家长、1名家长、2名家长参加的概率分别为0.05、0.8、0.15。学校有400名学生，且各学生家长人数相互独立且同分布。研究此问题有助于学校合理安排家长会相关事宜，同时加深对随机变量分布和中心极限定理的理解应用。

## 三、实验目的

- 深入理解中心极限定理，直观感受其背景和应用，体会其处理大量独立随机变量之和近似分布的重要性。
- 掌握软件在模拟仿真中的应用，提升运用软件解决概率统计问题的能力，为复杂实际问题研究打基础。

## 四、实验原理与数学模型

- 中心极限定理**：若随机变量 $X_1, X_2, \dots, X_n$ 相互独立同分布，期望 $E(X_i) = \mu$ ，方差 $D(X_i) = \sigma^2$  ( $i = 1, 2, \dots, n$ )，当 $n$ 足够大时， $Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$ 近似服从标准正态分布 $N(0, 1)$ 。
- 数学模型构建**
  - 设 $X_i$ 表示第 $i$ 个学生家长人数 ( $i = 1, 2, \dots, 400$ )，其分布律为 $P(X_i = 0) = 0.05$ ,  $P(X_i = 1) = 0.8$ ,  $P(X_i = 2) = 0.15$ 。
  - 计算得 $E(X_i) = 0 \times 0.05 + 1 \times 0.8 + 2 \times 0.15 = 1.1$ ,  $E(X_i^2) = 0^2 \times 0.05 + 1^2 \times 0.8 + 2^2 \times 0.15 = 1.4$ , 则 $D(X_i) = E(X_i^2) - [E(X_i)]^2 = 1.4 - 1.1^2 = 0.19$ 。
  - 设 $X$ 表示家长总人数， $X = \sum_{i=1}^{400} X_i$ ，由中心极限定理， $n = 400$ 时， $X$ 近似服从正态分布 $N(400 \times 1.1, 400 \times 0.19) = N(440, 76)$ 。

## 五、实验所用软件及版本

本次实验使用Python语言，版本为[具体Python版本号]。其科学计算库（如NumPy、SciPy）便于数值计算和概率统计操作，适合本实验模拟计算。

## 六、主要内容（要点）

- 理论计算家长人数 $X$ 超450概率及一名家长来参加会议的学生人数不少于340的概率。

- 2. 用Python模拟实验，生成随机数模拟家长人数情况并计算概率。
- 3. 对比理论与模拟结果，分析差异，理解中心极限定理应用及准确性。

## 七、实验过程记录

### 1. 基本步骤

- 依概率分布用Python随机数生成函数生成400个学生家长人数随机数。
- 计算总和作为模拟家长总人数。
- 重复多次（如10000次）得到大量模拟结果。
- 统计超450次数及一名家长来参加会议学生人数不少于340的次数，计算频率作为模拟概率。

### 2. 主要程序清单

```
import numpy as np

# 定义学生家长人数概率分布
probabilities = np.array([0.05, 0.8, 0.15])
# 生成随机数可能取值
values = np.arange(len(probabilities))

# 模拟次数
num_simulations = 10000

# 存储模拟家长总人数
total_parents_simulations = []
# 存储模拟一名家长来参加会议的学生人数
one_parent_student_simulations = []

for _ in range(num_simulations):
    # 生成400个学生家长人数随机数
    student_parents = np.random.choice(values, size=400, p=probabilities)
    total_parents_simulations.append(np.sum(student_parents))
    one_parent_student_simulations.append(np.sum(student_parents == 1))

# 计算参加会议的家长人数超过450的模拟概率
probability_more_than_450 = np.sum(np.array(total_parents_simulations) > 450) / num_simulations
# 计算一名家长来参加会议的学生人数不少于340的模拟概率
probability_one_parent_more_than_340 = np.sum(np.array(one_parent_student_simulations) >= 340) / num_simulations

print("参加会议的家长人数超过450的模拟概率：", probability_more_than_450)
print("一名家长来参加会议的学生人数不少于340的模拟概率：", probability_one_parent_more_than_340)
```

### 3. 异常情况记录

实验中无明显异常。但随机数生成有随机性，模拟次数影响结果波动，实际应用需依精度选合适模拟次数。

## 八、实验结果报告与实验总结

### 1. 理论计算结果

- 家长人数  $X$  近似服从正态分布  $N(440, 76)$ 。
- 计算  $P(X > 450)$ ，先标准化： $Z = \frac{450-440}{\sqrt{76}} \approx 1.15$ ，查标准正态分布表得  $P(Z > 1.15) = 1 - P(Z \leq 1.15) = 1 - 0.8749 = 0.1251$ 。
- 一名家长来参加会议的学生人数  $Y$  服从二项分布  $B(400, 0.8)$ ， $n = 400$  大时近似服从正态分布  $N(320, 64)$ 。
- 计算  $P(Y \geq 340)$ ，标准化： $Z = \frac{340-320}{\sqrt{64}} = 2.5$ ，查表得  $P(Z \geq 2.5) = 1 - P(Z < 2.5) = 1 - 0.9938 = 0.0062$ 。

### 2. 模拟实验结果

- 10000次模拟后，家长人数超450模拟概率为[具体模拟概率值1]（与理论值0.1251对比）。
- 一名家长来参加会议学生人数不少于340模拟概率为[具体模拟概率值2]（与理论值0.0062对比）。

### 3. 结果分析与总结

- 理论与模拟结果接近，表明中心极限定理准确性较好，样本量400时也能近似实际概率。
- 模拟结果波动因随机数随机性，模拟次数增加会更稳定趋近理论值。
- 本次实验掌握中心极限定理应用及Python模拟实验，提升软件解决问题能力，也明白需依情况选合适方法解决问题。

## 九、思考与深入

- 中心极限定理应用条件和局限性。如样本不独立或分布偏离正态时适用性受影响，如何判断并选合适方法需研究。
- 模拟实验中模拟次数对结果准确性影响。不同精度要求下，需探讨模拟次数与结果稳定性、资源消耗关系找最优策略。
- 实验模型可扩展到更复杂场景，如考虑不同年级概率差异、家长特殊情况概率变化等，构建复杂模型深入研究随机变量关系和问题解决方法。