概率论与数理统计实验报告

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 班级 |  | 姓名 |  | 学号 |  |
| 实验  名称 | **参加家长会人数问题** | | | | |
| 问题背景描述：  对于一个学生而言，来参加家长会的家长人数是一个随机变量，设一个学生无家长、1名家长、2名家长来参加的概率分别为0.05，0.8，0.15。若学校共有400名学生，各学生参加会议的家长人数是相互独立的，且服从同一分布。   1. 求参加会议的家长人数X超过450的概率； 2. 求一名家长来参加会议的学生人数不少于340的概率。   理论计算后，用软件进行模拟，观察试验与理论结果的差异。 | | | | | |
| 实验目的：   1. 加深对中心极限定理的认识，对背景和应用有直观的理解； 2. 了解软件在模拟仿真中的应用。 | | | | | |
| 实验原理与数学模型：  数学原理主要是**中心极限定理(Central Limit Theorem, CLT)以及由此引申出的正态近似**。  具体而言：   1. **中心极限定理 (CLT)**： 当有大量独立、分布相同、方差有限的随机变量加在一起时，其总和（减去总期望并除以总标准差后）在极大样本量的情况下近似服从标准正态分布。 2. **正态近似**： 利用中心极限定理，以正态分布 来近似人数的分布。通过计算期望以及方差，再将总和标准化：   标准化后的 Z 近似服从标准正态分布 。在此基础上，再进行计算  **问题 (1):** 以 记第 k个学生来参加会议的家长人数。 的分布律为：     1. **期望值** 2. **方差** 3. **总和 的期望与方差** 4. **标准化为标准正态分布 Z** 5. **计算**   即：  **问题 (2):**  以 表示第 k 个学生有恰好一名家长参加会议的指示变量。的定义如下：  **1. 期望值 和方差**  **2. 总和 的期望与方差**  **3. 标准化为标准正态分布 Z**  **4. 计算**  即： | | | | | |
| 实验所用软件及版本：  **Python 环境**  **Python 版本**: 3.10  **相关库**:  NumPy  Matplotlib  SciPy | | | | | |
| 主要内容（要点）：  在实验编程部分，首先通过模拟和理论计算求解了两个概率问题。   1. **计算总出席家长人数 X 超过450的概率**：   对于每个学生，家长出席的数量服从指定的离散分布，分别为0、1、2名家长，概率为0.05、0.8、0.15。通过模拟方法生成多个学生家长出席情况的样本，并对这些样本进行求和，从而得到每次模拟的总家长人数。计算出总出席家长人数超过450的经验概率，并通过中心极限定理，使用正态近似方法计算理论概率。通过绘制模拟结果的直方图和正态分布曲线进行可视化对比，展示模拟结果与理论结果的接近程度。   1. **计算恰有一名家长出席的学生人数 Y 不少于340的概率**：   模拟每个学生是否有一名家长出席，并通过多次模拟得到满足条件的学生人数总和。随后，计算出这些模拟中学生人数大于等于340的比例，作为经验概率。同时，通过正态近似计算理论概率，并将其与模拟结果进行比较。使用直方图和正态分布曲线的叠加展示模拟与理论的匹配情况。  通过这两个模拟过程，验证中心极限定理的应用，帮助加深了对正态近似方法在大样本情况下的理解和应用。 | | | | | |
| 实验过程记录（含：基本步骤、主要程序清单及异常情况记录等）：   1. **基本步骤**：    * 利用NumPy的 random.choice 根据给定概率生成大小为 的矩阵，每行模拟一次400个学生家长出席数情况。    * 对每次模拟求和，统计X的分布情况，再估计 P(X>450) 的模拟值。    * 对 Y 的情况使用伯努利试验 (binomial) 模拟Y 的分布，估计 P(Y≥340)。    * 利用Matplotlib绘制直方图和正态近似曲线，以进行对比和可视化。    * 运行程序无明显异常，结果与理论值较为接近。 2. **主要程序清单**：    * 1.py模拟 X 并计算 P(X>450)。    * 2.py第二部分代码模拟 Y 并计算 P(Y≥340)。 3. **程序使用说明**    * 两个python文件直接执行即可 4. **异常情况记录**：    * 无明显报错或异常。    * 数值结果随随机种子变化有所波动，但都在可接受范围内。 | | | | | |
| 实验结果报告与实验总结：  **实验结果报告**  **对于 X 的结果：**   * 理论：P(X>450)≈0.1257 (约12.57%) * 实际模拟结果：11.49%     二者接近。  **对于 Y 的结果：**   * 理论：P(Y≥340)≈0.0062 (约0.62%) * 实际模拟结果： 0.61%     模拟值略有微小差别，是误差正常范围内。  **实验总结**  本次实验，验证了使用正态近似对大样本独立同分布随机变量和的分布进行估计的有效性。理论结果与模拟结果接近，展示了中心极限定理。对于400这样的样本量，分布相对集中，其和趋近于正态分布是合理的。  由于正态分布近似模拟，所以存在一定偏差，属于正常现象。 | | | | | |
| 思考与深入：   1. **中心极限定理的应用场景**：本实验展示了当样本量较大时，可以用正态分布对求和分布近似，从而简化计算过程。在实际问题中，若要计算大数量独立随机变量之和的分布，正态近似是常用手段。 2. **误差来源**：    * 正态近似本身会有一定误差，尤其对分布较偏或尾部概率较小的事件。    * 模拟结果中的随机性导致测量误差。增加模拟次数可减小这种误差。 3. **改进方向**：    * 对于 Y 的问题，由于它是简单的二项分布 ，可直接用Scipy的 binom.pmf 或 cdf 函数计算精确概率，与正态近似做对比，进一步检验近似精度。    * 对于 X 的问题，可以考虑分布的确切形式（它是多个非对称分布的和），或者尝试泊松近似（如果合适）进行比较。   本次实验为中心极限定理和正态近似的实际应用提供了一个生动的案例。 | | | | | |