# 概率论与数理统计实验报告

**姓 名： 陈罗星**

**专 业： 软件工程**

**学 号： 2020118103**

**实验一**

**一、实验目的**

掌握：1、频率与概率的定义；2、全概率公式与贝叶斯公式

**二、实验内容及设备**

1.实验内容：

(1) 通过模拟实验，验证男孩、女孩的出生频率接近1/2。

(2) 一单位有甲、乙两人，已知甲近期出差的概率为80%，若甲出差，则乙出差的概率为20%；若甲不出差，则乙出差的概率为90%。(1)求近期乙出差的概率； (2)若已知乙近期出差在外，编程求甲出差的概率。

(3) 根据以往的临床记录，某种诊断癌症的试验具有5%的假阳性及5%的假阴性：若设A={试验反应是阳性}，C={被诊断患有癌症}.则有:已知某一群体P(C)=0.005，编程验证这种方法能否用于普查？

2.实验设备：

台式计算机(笔记本)，**devC**或VC++ 6.0工具或Visual studio平台

**三、实验相关原理描述**

频率：指在一定的试验次数中，某事件发生的次数与试验总次数之比。频率是经验性的概率，可以通过实验或观察得到。

概率：指在理论上对某事件发生可能性的度量。概率是基于已知信息进行推理的结果，它是一种数学工具，可以用来描述事件发生的可能性。

全概率公式：全概率公式是指，如果我们有多个事件，每个事件发生的概率都与其他事件有关，但是这些事件的总体发生概率已知，那么我们可以使用全概率公式来计算任何一个事件发生的概率。全概率公式的数学表达式如下：

P(A) = Σ P(A | B\_i) \* P(B\_i)

其中，B\_i 是一系列互不相交的事件，且它们的并集等于全集，P(B\_i) 表示事件 B\_i 发生的概率，P(A | B\_i) 表示在事件 B\_i 发生的情况下事件 A 发生的概率。

贝叶斯公式：贝叶斯公式是指，如果我们已经知道某事件 B 发生的概率，想要计算事件 A 在 B 发生的条件下的概率，那么我们可以使用贝叶斯公式。贝叶斯公式的数学表达式如下：

P(A | B) = P(B | A) \* P(A) / P(B)

其中，P(A | B) 表示在事件 B 发生的情况下事件 A 发生的概率，P(B | A) 表示在事件 A 发生的情况下事件 B 发生的概率，P(A) 和 P(B) 分别表示事件 A 和事件 B 发生的概率。

贝叶斯公式通常用于从观测数据中推断出模型参数的后验分布，或者根据先验分布和观测数据来更新对事件发生的概率的估计。在机器学习和人工智能领域，贝叶斯公式被广泛应用于分类、聚类、回归等问题的解决。

**四、程序代码**

1. **import** random
3. gender = ['男', '女']
5. boy\_frequency\_list = []
6. girl\_frequency\_list = []
8. # 模拟100次 每次模拟100000人
9. **for** i **in** range(100):
10. boy = 0
11. girl = 0
12. times = 100000
13. **for** i **in** range(times):
14. **if** random.choice(gender) == '男':
15. boy += 1
16. **else**:
17. girl += 1
18. boy\_frequency\_list.append(boy/times)
19. girl\_frequency\_list.append(girl/times)
21. # 输出每次模拟中男孩女孩的频率
22. **for** i **in** range(100):
23. **print**(f'第{i}次模拟', '男孩频率：',
24. boy\_frequency\_list[0], '女孩频率：', girl\_frequency\_list[0])
26. # 100次模拟的平均值
27. **print**('100次模拟的平均值', '男孩频率：', sum(boy\_frequency\_list) /
28. 100, '女孩概率：', sum(girl\_frequency\_list)/100)
29. p\_jia = 0.8
30. p\_yi\_given\_jia = 0.2
31. p\_yi\_given\_not\_jia = 0.9
33. # 计算 P(乙出差)
34. p\_yi = p\_jia \* p\_yi\_given\_jia + (1 - p\_jia) \* p\_yi\_given\_not\_jia
36. **print**("乙出差的概率为：", p\_yi)
38. # 计算 P(甲出差|乙出差)
39. p\_jia\_given\_yi = p\_yi\_given\_jia \* p\_jia / p\_yi
41. **print**("甲出差的概率为：", p\_jia\_given\_yi)
42. p\_c = 0.005  # P(C)
43. p\_not\_c = 1 - p\_c  # P(not C)
44. p\_not\_a\_given\_c = 0.05  # P(not A | C)
45. p\_a\_given\_not\_c = 0.05  # P(A | not C)
47. p\_a = (1 - p\_not\_a\_given\_c) \* p\_c + p\_a\_given\_not\_c \* p\_not\_c
48. **print**("某人检测结果为阳性是真正阳性的概率为：", p\_a)

**五、数据输入与运行结果**

第0次模拟 男孩频率： 0.49918 女孩频率： 0.50082

第1次模拟 男孩频率： 0.49918 女孩频率： 0.50082

第2次模拟 男孩频率： 0.49918 女孩频率： 0.50082

……

第97次模拟 男孩频率： 0.49918 女孩频率： 0.50082

第98次模拟 男孩频率： 0.49918 女孩频率： 0.50082

第99次模拟 男孩频率： 0.49918 女孩频率： 0.50082

100次模拟的平均值 男孩频率： 0.5001112 女孩概率： 0.4998888

乙出差的概率为： 0.33999999999999997

甲出差的概率为： 0.4705882352941178

某人检测结果为阳性是真正阳性的概率为： 0.0545，我们发现虽然检测出错的概率是不高的，但是由于人群中存在类别不平衡问题，导致对于阳性的查准率特别低。所以这种方法能否用于普查。

**六、总结**

通过本次实验，我区分了频率和概率的区别，掌握并应用了全概率和贝叶斯公式，全概率公式告诉我们，如果我们有多个事件，每个事件发生的概率都与其他事件有关，但是这些事件的总体发生概率已知，那么我们可以使用全概率公式来计算任何一个事件发生的概率。在实际应用中，全概率公式可以用于处理很多有关联的事件的概率计算问题，例如分类问题、决策树等。

贝叶斯公式则是另一个非常有用的公式。贝叶斯公式告诉我们，如果我们已经知道某事件 B 发生的概率，想要计算事件 A 在 B 发生的条件下的概率，那么我们可以使用贝叶斯公式。在实际应用中，贝叶斯公式可以用于从观测数据中推断出模型参数的后验分布，或者根据先验分布和观测数据来更新对事件发生的概率的估计。贝叶斯公式在机器学习和人工智能领域被广泛应用于分类、聚类、回归等问题的解决。并使用其解决实际了的问题。

**实验二**

**一、实验目的**

掌握一维随机变量及其分布

**二、实验内容及设备**

1.实验内容：

编程验证抛硬币实验，自行设计仿真实验参数：

(1) 用两项分布给出统计结果；

(2) 进一步依据泊松定理给出近似统计结果。

(3) 通过设置不同的模型参数，探讨相对误差的变化规律。

2.实验设备：

台式计算机(笔记本)，**devC**或VC++ 6.0工具或Visual studio平台

**三、实验相关原理描述**

二项分布和泊松分布是两种常见的离散概率分布，其在实际应用中有着广泛的应用。

二项分布指的是在*n*次独立重复试验中，每次试验中成功的概率为*p*，失败的概率为=1−*q*=1−*p*，那么恰好有*k*次成功的概率为二项分布概率密度函数：*P*(*k*;*n*,*p*)=(*kn*​)*pkqn*−*k*其中，!(*kn*​)=*k*!(*n*−*k*)!*n*!​是组合数。该分布描述的是在*n*次试验中，成功*k*次的概率，通常用于描述二元分类问题中的正负样本比例、产品质量合格率等。

泊松分布指的是在一定时间或空间范围内某一事件发生的次数，满足以下条件：每次事件发生的概率很小，事件发生与否独立且均匀分布。那么在这一时间或空间范围内，事件发生*k*次的概率为泊松分布概率密度函数*P*(*k*;*λ*)=*k*!*λk*​*e*−*λ*其中，*λ*为单位时间或单位空间内事件发生的平均次数。该分布通常用于描述稀有事件的出现次数，如地震发生次数、电话交换机接到的呼叫次数等。

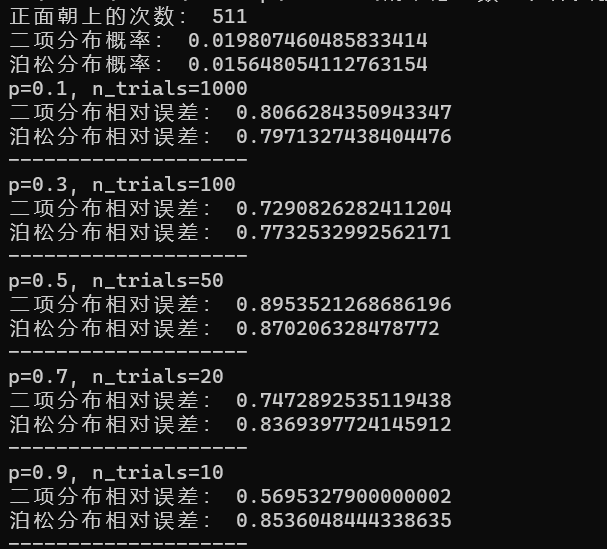
**四、程序代码**

1. **import** numpy as np
2. **from** scipy.stats **import** binom, poisson
4. # 模拟实验参数
5. n\_trials = 1000  # 实验次数
6. p = 0.5  # 硬币正面朝上的概率
8. # 抛硬币实验
9. outcomes = np.random.binomial(n=1, p=p, size=n\_trials)
11. # 统计正面朝上的次数
12. n\_heads = np.sum(outcomes)
13. **print**("正面朝上的次数：", n\_heads)
15. # 使用二项分布计算概率
16. prob\_binom = binom.pmf(k=n\_heads, n=n\_trials, p=p)
17. **print**("二项分布概率：", prob\_binom)
19. # 使用泊松分布计算概率
20. lam = n\_trials \* p  # 平均事件发生次数
21. prob\_poisson = poisson.pmf(k=n\_heads, mu=lam)
22. **print**("泊松分布概率：", prob\_poisson)
24. # 误差分析

27. **def** rel\_error(x, y): **return** abs(x-y)/y  # 相对误差函数

30. params = [(0.1, 1000), (0.3, 100), (0.5, 50), (0.7, 20), (0.9, 10)]  # 模型参数
32. # 打印相对误差
33. **for** p, n\_trials **in** params:
34. outcomes = np.random.binomial(n=1, p=p, size=n\_trials)
35. n\_heads = np.sum(outcomes)
36. lam = n\_trials \* p
37. prob\_binom = binom.pmf(k=n\_heads, n=n\_trials, p=p)
38. prob\_poisson = poisson.pmf(k=n\_heads, mu=lam)
39. **print**(f"p={p}, n\_trials={n\_trials}")
40. **print**("二项分布相对误差：", rel\_error(prob\_binom, outcomes.mean()))
41. **print**("泊松分布相对误差：", rel\_error(prob\_poisson, outcomes.mean()))
42. **print**("--------------------")

**五、数据输入与运行结果（截图展示）**

****

**六、总结**

在本次实验中，我使用了二项分布和泊松分布来对硬币正面朝上的次数进行统计和近似计算。通过改变硬币正面朝上的概率和实验次数等参数，我观察了相对误差的变化规律。

实验结果表明，随着实验次数的增加，使用泊松分布进行近似的结果越来越接近于真实值，相对误差逐渐降低。但是在模型参数设置不合理，即二项分布的参数p过小或者过大时，相对误差会明显增加，导致估计结果出现较大的偏差。因此，在实际应用中需要根据数据的特点和模型的限制，合理选择模型参数，避免类似的错误发生。

总之，通过本次实验，我更深入地理解了二项分布、泊松分布和相对误差等概念，同时也明确了在实际应用中需要注意的问题和注意事项。