圣彼得堡悖论的 python 验证

一．背景

在这学期的《概率论与数理统计》课上，老师提出这样一个问题：

“设定掷出正面或者反面为成功，游戏者如果第一次投掷成功，得奖金 2 元，游戏结束；第一次若不成功，继续投掷，第二次成功得奖金 4 元，游戏结束；这样，游戏者如果投掷不成功就反复继续投掷，直到成功，游戏结束。如果第 n 次投掷成功，得奖金 2 的 n 次方元，游戏结束。”

按照概率期望值的计算方法，将每一个可能结果的得奖值乘以该结果发生的概率即可得到该结果奖值的期望值。游戏的期望值即为所有可能结果的期望值之和。随着 n 的增大，以后的结果虽然概率很小，但是其奖值越来越大，每一个结果的期望值均为 1，所有可能结果的得奖期望值之和，即游戏的期望值，将为“无穷大”。

按照概率的理论，多次试验的结果将会接近于其数学期望。但是实际的投掷结果和计算都表明，多次投掷的结果，其平均值最多也就是几十元。正如 Hacking（1980）所说：**“没有人愿意花 20 元去参加一次这样的游戏。”**

这就是圣彼得堡悖论是数学家丹尼尔・伯努利（Daniel Bernoulli）的堂兄尼古拉・伯努利（Nicolaus Bernoulli）在１７３８提出的一个概率期望值悖论，它来自于一种掷币游戏，即圣彼得堡游戏。

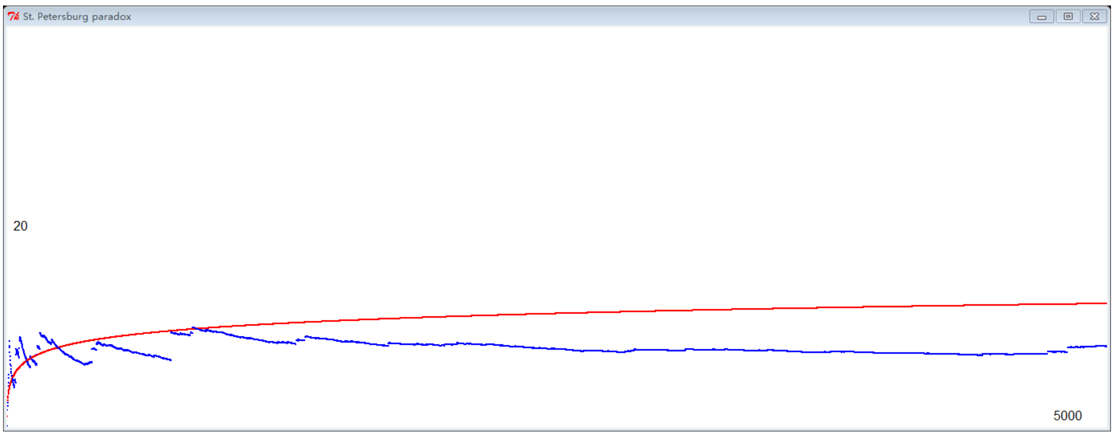
二．程序功能

历史上的模拟试验的结果说明，实际试验的平均值——样本均值是随着实验次数的增加而变化的。在大量实验以后，其实验均值 X 可以近似表示为 X≈logn/log2。我希望利用 python 程序对这一结论进行验证。

本程序就是利用循环结构，模拟多次投掷试验，得出一个拟合曲线（蓝色），和曲线 logn/log2（红色）进行对比分析。并标出了纵坐标“20”，以印证上文 Hacking 的论断，即没有人愿意花 20 元去参加这样一次游戏。

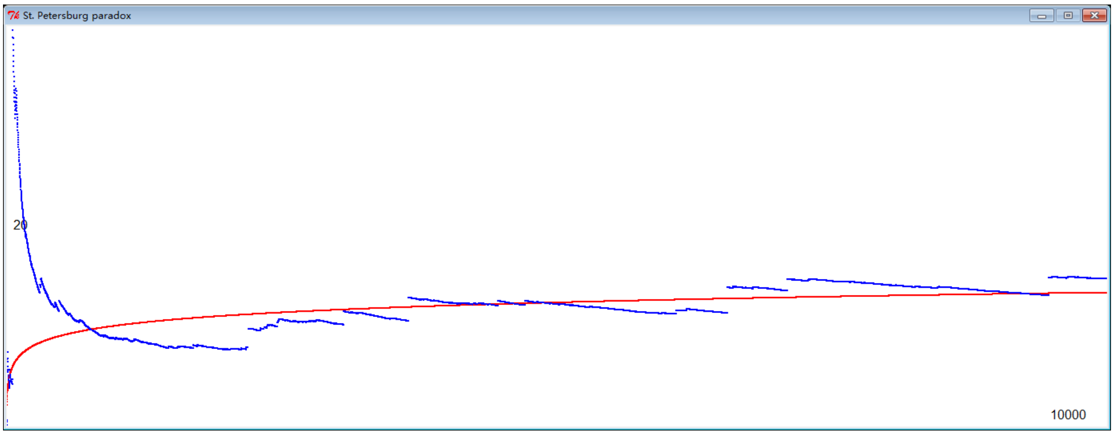
三．结果分析

1. n=5000

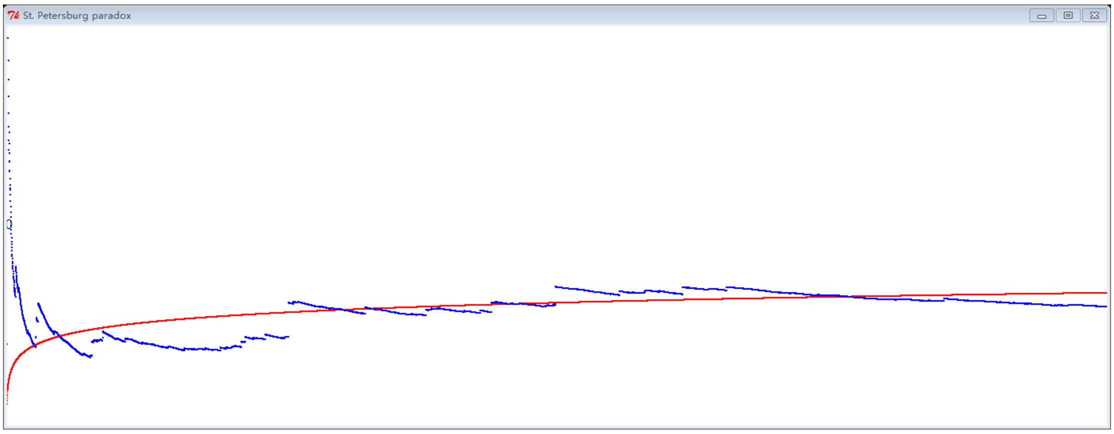


2.n=10000

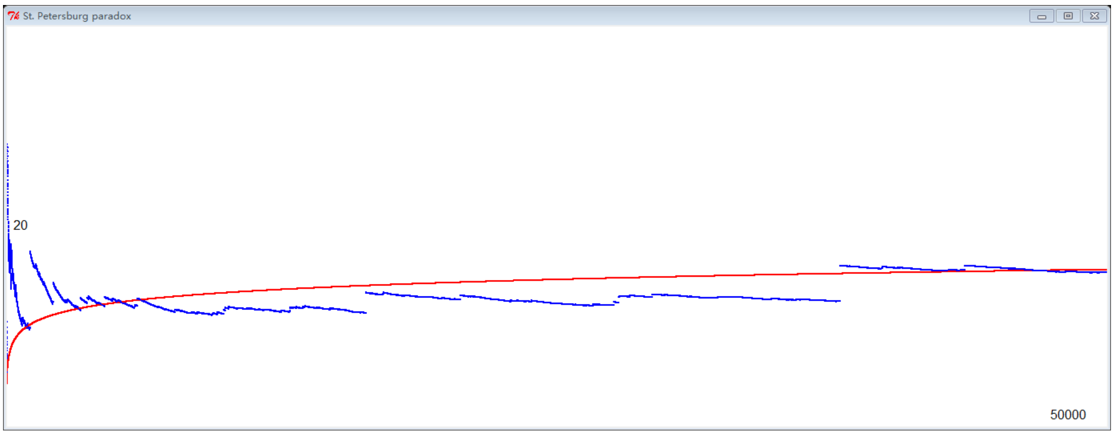
（1）



（2）



3.n=50000



从图像上可以很直观地看出，当 n 较小时，试验结果随机性较大，并无明显规律；而随着 n 逐渐增大，图像越来越趋近于参照函数的图像。

四．存在问题

尽管程序本身并不复杂，但由于要求的样本量较大，一般认为 n>10^6 为有效统计。但是在程序运行中，一般 10^5 的运算时间就会非常大，并不符合线性时间，不知是电脑硬件问题还是程序本身算法还有待改进？