HW1：随机变量及其分布

1. 假设在考试的多项选择中，考生知道正确答案的概率为 p，猜测答案的概率为 1-p，并且假设考生知道正确答案答对题的概率为 1，猜中正确答案的概率为 $1/m$，其中 m 为多选项的数目。那么已知考生答对题目，求他知道正确答案的概率。

2. 假设硬币正面向上的概率为 p。我们抛掷硬币 N 次，令 X 表示正面向上的次数，则 X 为一个二项分布的随机变量。我们直观感觉 X 应该和 Np 很接近。为了验证该结论是否正确，我们重复多次试验，取 X 的平均值，比较 X 的平均值和 Np 的接近程度。比较 $p=0.3, N=10,100,1000$，和 $p=0.03, N=10,100,1000$。给出试验次数 N 与正面向上比率的函数图。

4、理解抽样分布（sampling distribution）。令 $X_1,\ldots,X_N$ 为独立同分布样本（IID），其均值和方差分别为 μ 和 σ^2。则样本均值为 $\bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ 为一统计量，是数据的函数。由于 $\bar{X}_N$ 也是随机变量，因此也可对其进行分布进行描述，该分布称为统计量的抽样分布。请不要将 X_i 的分布函数 p_X 与 $\bar{X}_N$ 的分布 $p_{\bar{X}_N}$ 混淆。为了更清楚地认识到这一点，我们假设 $X_1,\ldots,X_N \sim \text{Unif}[0,1]$，画出 p_X。

(1) 计算理论的 $\mathbb{E}\left(\bar{X}_N\right)$ 和 $\mathbb{V}\left(\bar{X}_N\right)$，分析并画出当 N 变化时二者的变化。

(2) 模拟得到 $\bar{X}_N$ 的分布。取 N= 5, 10, 25, 50, 100，从 $X_1,\ldots,X_N \sim \text{Unif}[0,1]$ 得到 N 个样本，计算 $\bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$，得到 $\bar{X}_N$ 的一个样本。上述过程重复 100 次，可得到 $\bar{X}_N$ 的 100 个样本。计算 100 个 $\bar{X}_N$ 样本的样本均值 $\hat{\mu}_{\bar{X}_N} = \frac{1}{100}\sum_{i=1}^{100} \bar{X}_{Ni}$ 作为 $\mathbb{E}\left(\bar{X}_N\right)$ 的估计，100 个 $\bar{X}_N$ 样本的样本方差 $\hat{\sigma}^2_{\bar{X}_N} = \frac{1}{100}\left(\sum_{i=1}^{100} \bar{X}_{Ni} - \hat{\mu}_{\bar{X}_N}\right)^2$ 作为 $\mathbb{V}\left(\bar{X}_N\right)$ 的估计，观察该估计值与(1)中理论值的差异。当 N 变化时，该差异有何变化规律？

5、完成示例代码中对 Rent Listing Inquries 数据集 bedrooms 特征分布的分析。