数学实验 统计量和 MC 算法 实验报告

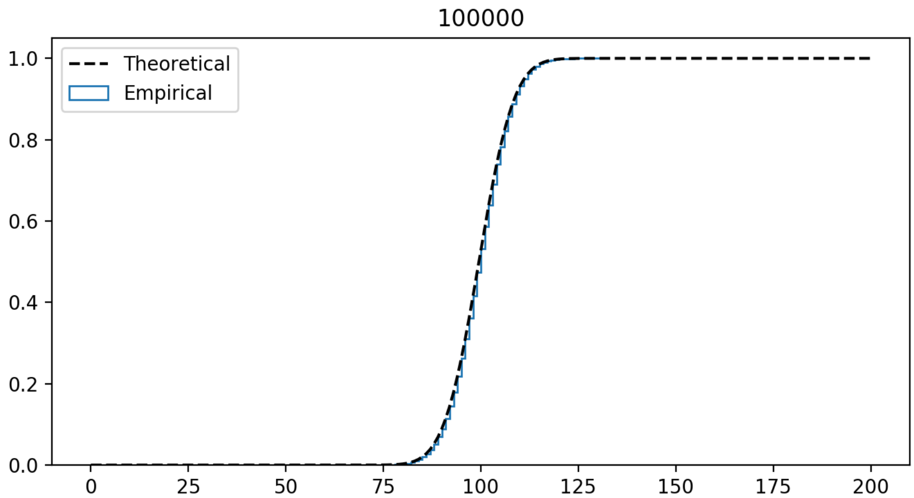
计 65 赖金霖 2016011377

一

1. 独立抛掷 200 次硬币，正面个数的近似分布 p(x=i) 满足：

$$p(x=i)=C_{200}^{i}\times(\frac{1}{2})^{200},\quad i=0,1,...,200$$

可以使用 np.random.randint 函数随机 0 和 1，来模拟投掷硬币。随机进行 100000 次实验（投掷 100000×200 次），得到如下经验分布函数（黑色为近



似分布函数)：

数值上 F(x)有如下的对比：

x	80	90	100	110	120
Theoretical	0.00284	0.08948	0.52817	0.93132	0.99818
Empirical	0.00274	0.08943	0.52927	0.93167	0.99826

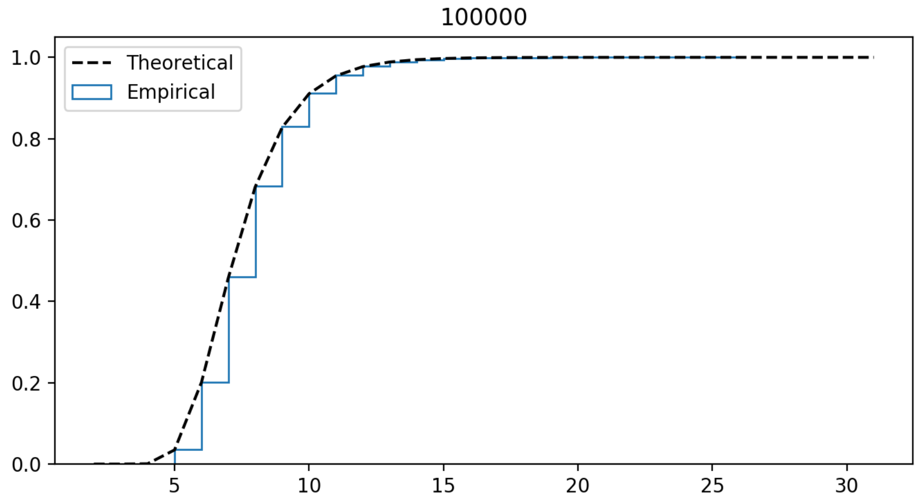
可以看出，经验函数的估计十分准确，误差均不超过 5%。

2. 独立抛掷硬币 k 次，设最长 0 串个数 X 的分布的 cdf 为 $F_k(x)$，最长 0 或 1 串长度 Y 的分布的 cdf 为 $G_k(y)$，其中 x=0,1,…,200，y=1,2,…,200，那么我们有 $G_k(y)=F_{k-1}(y-1)$。
   这是因为长度为 k 的 01 串的最长 0 或 1 串和长度为 k-1 的 01 串的最长 0 串有对应关系：设有一长度为 k 的 01 串 A[1~k]，B[i]=(A[i]等于 A[i+1])，那么 A 中的 0 或 1 串对应 B 中的 0 串；反之，对长度为 k-1 的 01 串 B[1~k-1]，也有对应的 A 串，使 B 的 0 串对应 A 的 0 或 1 串。
   根 据 https://math.stackexchange.com/questions/59738/probability-for-the-length-of-the-longest-run-in-n-bernoulli-trials，$F_k(y)$满足

$$F_k(x)=1+\sum_{j=1}^{\lfloor\frac{k}{x+1}\rfloor}(-1)^j(p+(\frac{n-jx-j+1}{j})(1-p))C_{n-jx-j}^{j-1}(\frac{1}{2})^{jx+2j-1}$$

   我们可以直接算出 $G_k(x)$的分布函数。

当 k=200 时，随机进行 100000 次实验，最长 0 或 1 串的分布如下所示



根据 Dvoretzky - Kiefer - Wolfowitz 不等式，在样本数为 n 的情况下，经验 cdf $F^*_n(x)$ 和理论 cdf F(x) 的差距满足

$$Pr(\sqrt{n} \parallel F_n^* - F \parallel_\infty > z) \le 2e^{-2z^2}$$

在实验中，若取 z^2=0.1，那么我们有

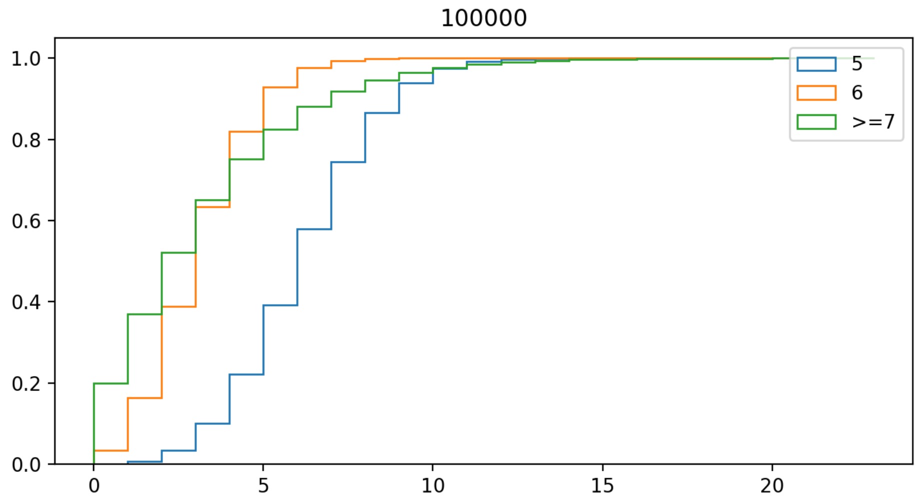
$$Pr(\parallel F_n^* - F \parallel_\infty > 0.01) \le 4.12 \times 10^{-9}$$

$G_{200}(x)$ 的实验分布和理论分布数值如下

x	6	7	8	9	10
Theoretical	0.20072	0.45813	0.68165	0.82760	0.91053
Empirical	0.20167	0.46006	0.68246	0.82683	0.91042

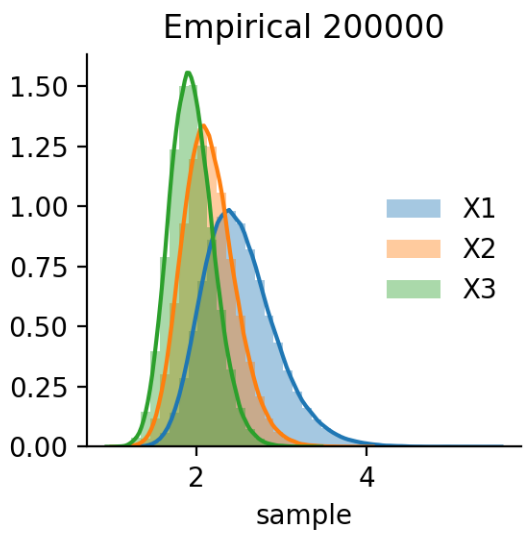
结果在理论上和时间上都特别准确。

3. 和 1，2 类似，进行 100000 次实验，长度为 5、6 和不小于 7 的 0 或 1 串的个数的经验分布如下所示（000000 里有 2 个长度为 5 的 0 串）



可以明显看出，在连续 200 次抛掷硬币的实验里，有很大的概率出现连续 5 个或以上的连续相同串。

二、使用程序随机抽取 100 个标准正态分布并排序，X(1)，X(2)，X(3)是最大的三个数，可以看做一次实验。进行 200000 次实验，可以得到如下的的经验分布（pdf）



明显可以看出，X(1)、X(2)、X(3)三者，均值越来越小，方差越来越大。

有如下的数值（方差经过修正）：

X	1	2	3
均值	2.50869	2.14869	1.94711
方差	0.42952	0.30963	0.26186

三、

5. 炮弹落点 X 满足

$$X \sim N(\mu, \Sigma) \quad \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_x^2 & r\sigma_x\sigma_y \\ r\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

带入数字，使用 np.random.multivariate_normal 进行 1000000 次采样，在|X|≤100 时视为炮弹落入圆形区域，有炮弹命中概率为 0.698834。

6. 容易算出，球面体积为 2/3π，而对锥面体积，使用积分方法有

$$V = \pi - \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_0^1 \sqrt{x^2+y^2}\,dxdy$$
$$= \int_0^{2\pi} \int_0^1 \rho^2 d\rho d\phi = \frac{1}{3}\pi$$

所以理论上体积为π。

使用 MC 方法，则可以随机在[-1,1]×[-1,1]×[0,2]空间上采点，其在冰淇淋内的频率×空间体积 8，即为冰淇淋的体积。进行 1000000 次采样，可以算出冰淇淋的体积为 3.139952，与理论数值有一定差距。

四、参考资料

https://en.wikipedia.org/wiki/Empirical_distribution_function
https://math.stackexchange.com/questions/59738/probability-for-the-

length-of-the-longest-run-in-n-bernoulli-trials

五、代码

代码路径：

https://github.com/lll6924/math_exp/blob/master/exp7/statistics_mc.py