《抽奖》解题报告

南京外国语学校 乔明达

1 题目大意

有N个盒子，投一个球落入第i个盒子的概率为p_i，$\sum_{i=1}^{N} p_i = 1$。互相独立地投M次球，每个盒子的分值记为盒内小球个数的平方。求所有盒子分值之和的期望，以及M次投球后非空盒子的期望个数。

2 数据规模和约定

30%的数据满足$N, M \le 1000$，且$p_1 = p_2 = \cdots = p_N = \frac{1}{N}$；
100%的数据满足$N, M \le 100000$。

3 第一问

直接求最终的平方和的期望是比较困难的，不妨考虑每投入一个球对平方和的贡献。

假设我们之前已经投了t个球，现在我们把一个球投入了第i个盒子，这个盒子里有m_i个球。

令Δ_i为投入这个球后分值的变化量，那么$\Delta_i = (m_i + 1)^2 - {m_i}^2 = 2m_i + 1$。

对两边求期望，由于期望具有线性性质，$E[\Delta_i] = E[2m_i + 1] = 2E[m_i] + 1$

另一方面，$E[m_i]$是投了t个球之后，概率为p_i的盒子里球的期望个数。显然有$E[m_i] = t \cdot p_i$，于是$E[\Delta_i] = 2t \cdot p_i + 1$。

令Δ为已经投了t个球之后，再投入一个球分值的变化量。因为$\Delta = \Delta_i$的概率为p_i，所以有：

$$E[\Delta] = \sum_{i=1}^{N} (p_i \cdot E[\Delta_i]) = \sum_{i=1}^{N} \left(2t \cdot {p_i}^2 + p_i\right) = 2t \sum_{i=1}^{N} {p_i}^2 + 1$$

对于$t = 0, 1, \ldots, M-1$时的$E[\Delta]$求和，就可以得到第一问的答案为：

$$\sum_{t=0}^{M-1}\left(2t \cdot \sum_{i=1}^{N} {p_i}^2 + 1\right) = 2\left(\sum_{t=0}^{M-1} t\right)\left(\sum_{i=1}^{N} p_i^2\right) + M = M(M-1)\sum_{i=1}^{N} p_i^2 + M$$

计算上式的时间复杂度为$\Theta(N)$。

4 第二问

用随机变量c_i表示第i个盒子在M次投球之后是否为空，如果为空则$c_i = 0$否则$c_i = 1$。

我们要求的答案为$E[\sum_{i=1}^{N} c_i] = \sum_{i=1}^{N} E[c_i]$

根据c_i的定义，$E[c_i]$数值上等于第i个盒子非空的概率，即$c_i = 1-(1-p_i)^M$。

于是第二问的答案为：

$$\sum_{i=1}^{N}\left[1-(1-p_i)^M\right] = N - \sum_{i=1}^{N}(1-p_i)^M$$

计算上式的时间复杂度为$\Theta(N \log M)$。

5 总结

综合以上两问的解法，本题算法的时间复杂度为$\Theta(N \log M)$。