**[快速排序里的学问：再看看称球问题](http://www.nowamagic.net/librarys/veda/detail/2388)**

前面用猜数字游戏说明了二分的思想，这里再看一个常见的思维题：

12个小球，其中有一个是坏球。有一架天平。需要你用最少的称次数来确定哪个小球是坏的并且它到底是轻还是重。

这个问题是一道流传已久的智力题。网络上也有很多讲解，还有泛化到N个球的情况下的严格证明。也有零星的一些地方提到从信息论的角度来看待最优解法。本来我一直认为这道题目除了试错之外没有其它高妙的思路了，只能一个个方法试，并尽量从结果中寻找信息，然后看看哪种方案最少。

然而，实际上它的确有其它的思路，一个更本质的思路，而且根本用不着信息论这么拗口的知识。

我们先回顾一下猜数字游戏。为了保证任何情况下以最少次数猜中，我们的策略是每次都排除恰好一半的可能性。类比到[称球](http://www.nowamagic.net/librarys/veda/tag/%E7%A7%B0%E7%90%83)问题上：

* 坏球可能是12个球中的任意一个，这就是12种可能性。
* 而其中每种可能性下坏球可能轻也可能重。于是“坏球是哪个球，是轻是重”这个问题的答案就有12×2=24种可能性。

现在我们用天平来称球，就等同于对这24种可能性发问，由于天平的输出结果有三种“平衡、左倾、右倾”，这就相当于我们的问题有三个答案，即可以将所有的可能性切成三份，根据猜数字游戏的启发，我们应当尽量让这三个分支概率[均等](http://www.nowamagic.net/librarys/veda/tag/%E5%9D%87%E7%AD%89)，即平均切分所有的可能性为三等份。如此一来的话一次称量就可以将答案的可能性缩减为原来的1/3，三次就能缩减为1/27。而总共才有24种可能性，所以理论上是完全可以3次称出来的。

如何称的指导原则有了，构造一个称的策略就不是什么太困难的事情了。

首先不妨解释一下为什么最直观的称法不是最优的——6、6称：在6、6称的时候，天平平衡的可能性是0。刚才说了，最优策略应该使得天平三种状态的概率均等，这样才能三等分答案的所有可能性。

为了更清楚的看待这个问题，我们不妨假设有6个球，来考虑一下3、3称和2、2称的区别：

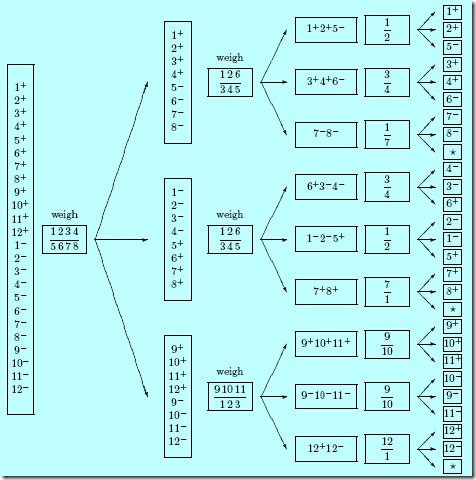
在未称之前，一共有12种可能性：1轻、1重、2轻、2重、…、6轻、6重。

3、3称：现在将1、2、3号放在左边，4、5、6放在右边3、3称了之后，不失一般性假设天平左倾，那么小球的可能性就变成了原来的一半（6种）：1重、2重、3重、4轻、5轻、6轻。即这种称法能排除一半可能性。

2、2称：现在再来看2、2称法，即1、2放左边，3、4放右边，剩下的5、6不称，放一边。假设结果是天平平衡，那么可能性剩下——4种：5重、5轻、6重、6轻。假设天平左倾，可能性也剩下4种：1重、2重、3轻、4轻。右倾和左倾的情况类似。总之，这种称法，不管天平结果如何，情况都被我们缩小到了原来的三分之一！我们充分利用了“天平的结果状态可能有三种”这个条件来三等分所有可能性，而不是二等分。

说到这里，剩下的事情就实在很简单了：第二步称法，只要记着这样一个指导思想——你选择的称法必须使得当天平平衡的时候答案剩下的可能性和天平左倾（右倾）的时候答案剩下的可能性一样多。实际上，这等同于你得选择一种称法，使得天平输出三种结果的概率是均等的，因为天平输出某个结果的概率就等同于所有支持这个结果（左倾、右倾、平衡）的答案可能性的和，并且答案的每个可能性都是等概率的。

MacKay在他的书《Information Theory: Inference and Learning Algorithms》（[作者开放免费电子书](http://users.aims.ac.za/~mackay/itila/book.html)）里面4.1节专门讲了这个称球问题，还画了一张不错的图，我就照抄了：



图中“1+”是指“1号小球为重”这一可能性。一开始一共有24种可能性。

4、4称了之后不管哪种情况（分支），剩下来的可能性总是4种。这是一个完美的三分。

然后对每个分支构造第二次称法，这里你只要稍加演算就可以发现，分支1上的第二次称法，即“1、2、6对3、4、5”这种称法，天平输出三种结果的可能性是均等的（严格来说是几乎均等）。这就是为什么这个称法能够在最坏的情况下也能表现最好的原因，没有哪个分支是它的弱点，它必然能将情况缩小到原来的1/3。