

选票定理与卡塔兰数

万精油

许多趣味数学题目都以概率的形式出现。有些是直接问某种问题出现的概率，有些是问在各种情况下，采取哪一种策略最有利，归根到底还是要算概率。当然，要使问题真正有趣，所求的概率不能太简单，大多数情况下都要包含无穷序列，或者有包含任意数 N 的公式，甚或递归等等，用不同的手段求出各种情况下的组合。上一期的题目就属于这种类型。我们小时候学会了用手指头数数，长大了需要数的数手指头数不过来了，我们就学会了进位，还学会了归类。当需要归类的东西变得很复杂的时候，就需要一些技巧。麻省理工学院有一门课就叫“数数的艺术”（The Art of Counting）。我们这篇文章就是想通过一个趣味题目介绍一下与技巧数数有关的一个有趣定理以及一个常用手段。

上期一共有两个题目，一个是买电影票，一个是洗碗。看似日常生活中的琐碎小事，实际上他们都与一个完全不琐碎，甚至可以说很堂皇的东西有关，那就是“选票定理”。在解答我们上期问题以前我们先来介绍一下这个“选票定理”。



选票定理全称叫伯特兰德（Bertrand）选票定理。其实早在 1878 年惠特沃斯就得到了这个定理，但后来（1887 年）又被伯特兰德重新发现，并被冠以他的名字。选票定理说，如果一个选举有 p 个人投 A 的票， q 个人投 B 的票， $p > q$ ，那么，在整个投票过程中 A 的票累计数一直严格多于 B 的票累计数的概率是

$$\frac{p-q}{p+q} \quad (1)$$

比如， $p = 3$ ， $q = 2$ ；也就是说总共有 5 个人投票，3 人投给 A，2 人投给 B。那么投票过程可能有如下 10 种可能。AAABB，AABAB，ABAAB，BAAAB，AABBA，

ABABA, BAABA, ABBAA, BABAA, BBAAA。这 10 种次序中只有第一与第二种从一开始 A 票累计数就一直严格大于 B 票累计数, 所以, 只有 1/5 的可能。上面的公式 $(3-2)/(3+2) = 1/5$ 得到验证。

$p = 3, q = 2$ 的情况我们可以把所有情况都列出来数。对于任意 p, q 就不可能都列出来, 必须要有更一般的方法来数。比如把某种情况归类, 或者把某种情况对应到另一个已知情况。选票定理已经有一百多年的历史了。贝尔特兰德当初用的是递归方法。后来被公认最简单的方法是安德烈 (Désiré André) 发明的反射方法。我们下面就来介绍这个方法。

一个投票次序如果 A 的选票累计数从头到尾都一直严格领先于 B 的选票累计数, 我们就把它叫做有利次序, 否则叫无利次序。显然, 凡是以 B 票开始的次序都是无利次序。在以 A 票开始的次序中, 也有一些是无利次序。怎么来数这些无利次序呢? 我们知道有利次序永远不能有 A 票数与 B 票数相等的时候。换句话说, 任何无利次序都必须要在某一个时间出现两个选票数相等的时候。我们把第一次 A, B 票数相等的点叫 E 点。现在把从第一票到 E 点的所有票都反转, 也就是说 A 票换成 B 票, B 票换成 A 票。这个次序就变成了一个以 B 票开始的次序。这样, 我们建立了一个以 A 开始的无利次序到以 B 开始的次序的对应。反过来, 因为 A 的总票数大于 B 的总票数, 所以, 以 B 票开始的次序必须出现 A 票数与 B 票数相等的时候。通过上面的反转变换, 我们可以把任意以 B 开始的次序换成一个以 A 开始的无利次序。也就是说, 我们在以 A 开始的无利次序与所有以 B 开始的次序中建立了一个 1-1 对应。所以, 所有无利次序等于两倍于以 B 开始的所有次序。那么, 以 B 开始的次序有多少呢? 由于 A 有 p 票, B 有 q 票, 以 B 票开始的次序占总次序数的 $q/(p+q)$ 。所以, 有利次序的概率是 $1 - 2 * q / (p + q) = (p - q) / (p + q)$ 。证毕。

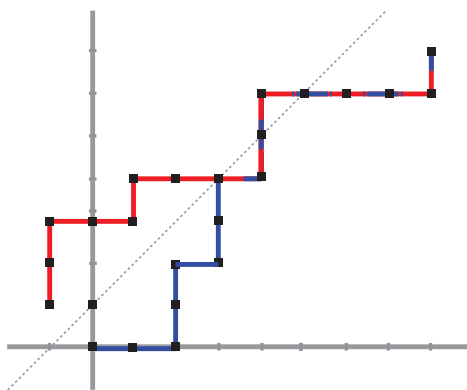
选票定理有一个变种, 不要求严格大于, 也就是说允许有 A 累计票数与 B 累计票数相等的时候。由于放宽了条件, 这个变种应用更广, 我们后面会提到。那么, 这个变种的概率又该如何算呢? 前面提到过的反射法几乎可以完全照搬过来。唯一需要改变的是, 这次的反射点是第一次 B 累计票超过 A 累计票的时候, 而不是它们相等的时候。我们把这个问题换一种几何表达法, 这个反射法就可以通过图形有效地表现出来。

假设我们从原点出发。每得一张 A 票我们就向右行一个单位, 每得一张 B 票我们就向上行一个单位。这样, 每一个选票次序都是一条从 $(0, 0)$ 到 (p, q) 向右或向上的折线。这样的折线总共有 $\binom{p+q}{q}$ 条, 在 $p+q$ 步中选出向上的 q 步, 其余的 p 步都是向右的。这些折线中有利的折线必须永远在对角线以下 (包括对角线)。不利的折线必须要碰到对角线上面那条与它平行的线, 也就是过 $(0, 1)$ 的 45 度线 (图中的黑线)。我们把第一次碰到这条线的点作为反转点。把这个点以前的点对黑线做镜像反射, 那么 A 票变 B 票, B 票变 A 票。原来从 $(0, 0)$ 开始的折线变成了从 $(-1, 1)$ 开始的折线。与上面类似, 这个对应是 1-1 对应。我们知道从 $(-1, 1)$ 到 (p, q) 的向右向上折线一共有 $\binom{p+q}{q-1}$ 条, 那么, 有利的折线数就是

$$\binom{p+q}{q} - \binom{p+q}{q-1} = \binom{p+q}{q} \frac{p+1-q}{p+1}.$$

除去总条数 $\binom{p+q}{q}$ ，我们得到有利折线的概率是

$$\frac{p+1-q}{p+1}. \quad (2)$$



其实，这个变种也可以直接从原始问题的结果得到。原始题目要求严格大于，那么，第一票必须是 A 票。第一票以后，剩下的就是满足 $(p-1, q)$ 的变种问题。通过代数操作，我们可以从原始题目的结果 (1) 得到上面的结果 (2)。

这个问题更进一步的变种是当 $p = q$ 的特殊情况。从 $(0, 0)$ 到 (p, q) 的向右向上的折线变成了从 $(0, 0)$ 到 (n, n) 。如果只是算概率，那么这个特殊情况也没有什么特别的结果，只需把 $p = q = n$ 代入上面的 (2)，我们可以得到有利折线的概率是 $1/(n+1)$ 。对这种特殊情形，人们更关心的是有利折线的数目。

从 $(0, 0)$ 到 (n, n) 的向右向上折线总数是 $\binom{2n}{n}$ 。从上面的推导我们知道，这些折线中有利折线占总数的 $1/(n+1)$ 。所以有利折线的总数是

$$C_n = \frac{1}{n+1} \binom{2n}{n} = \frac{(2n)!}{(n+1)!n!} = \prod_{k=2}^n \frac{n+k}{k}, \quad n \geq 0.$$

这个数列叫卡特兰数 (Catalan Numbers)，因为比利时数学家卡特兰 (Eugène Charles Catalan) 对这个数列做过很深的研究，所以冠上了他的名字。单从定义公式不能明显看出卡特兰数是整数。如果换一种形式表达

$$C_n = \binom{2n}{n} - \binom{2n}{n+1} = \frac{1}{n+1} \binom{2n}{n}, \quad n \geq 0,$$

那么，很明显，所有卡特兰数都是整数。对应于 $n = 0$ 到 10，它们的值是：1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796。

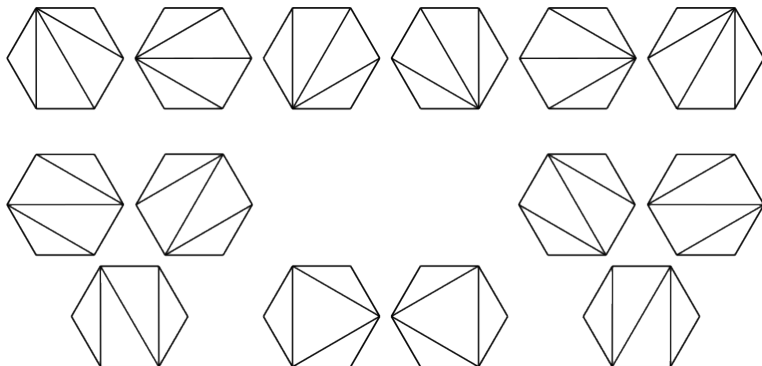
这些数在许多不同场合下重复出现。英文维基上列出了许多与这个问题等价的问题，有十好几条。其中我觉得最有意思的两个是：

1. 括弧问题。 C_n 对应于 $n+1$ 个运算可以被括弧包起来的不同包法的总数。比如 $abcd$ ，就可以有如下 5 种括弧包法：

((ab)c)d (a(bc))d (ab)(cd) a((bc)d) a(b(cd))

这个问题与选票问题的特例等价是很显然的。任何时候反括弧的累计数都不能大于正括弧的累计数（否则语法不通）。这个 5 正好对应于卡塔兰数 $n=3$ 的值。

2. 凸多边形三角剖分问题。 C_n 对应于 $n+2$ 边凸多边形不同的三角剖分法总数。比如， $n=4$ 时，我们有如下 14 种三角剖分法



这个问题与选票问题的特例等价不是那么明显。但如果我们知道卡塔兰数满足的一个递归公式

$$C_0 = 1 \text{ 且 } C_{n+1} = \sum_{i=0}^n C_i C_{n-i}, \quad n \geq 0,$$

那么，用数学归纳法证明起来就很简单了。

当然，还有两个对应就是我们上期的题目，排队买票问题以及猜拳洗碗问题。在解答上期题目以前，我们先讲一个与卡塔兰数有关的历史趣事。

前面已经说过这个数列在很多场合出现，不少人都研究过它。卡塔兰在研究汉诺塔以及前面提到过的括弧问题时对这个数列有比较深入的研究，后来有人在文章中把这个数列放到了卡塔兰名下，于是就叫开了。但人们后来又发现，欧拉在研究凸多边形三角剖分问题的时候就已经推出了卡塔兰数，那时候卡塔兰还没出生呢。更有趣的是，又有人发现，中国清朝数学家明安图比欧拉还要早推出这个数列以及一些相关公式。我们就来讲一下这个明安图与卡塔兰数的故事。

18 世纪早期，法国传教士杜德美（Pierre Jartoux）到中国，带来了三个级数展开式：其中一个是正弦展开式，一个是正矢展开式。这两个级数用现在的眼光看其实就是麦克劳林级数的特例，但那时不是很清楚这一点。最有意思的是第三个，一个牛顿的关于 π 的展开式

$$\pi = 3 \left(1 + \frac{1^2}{4 \cdot 3!} + \frac{1^2 \cdot 3^2}{4^2 \cdot 5!} + \frac{1^2 \cdot 3^2 \cdot 5^2}{4^3 \cdot 7!} + \dots \right)$$

这个式子给出的结果很精确，但这个传教士没有带这些展开式的证明到中国。明安图知道这些老外金针不度，要留一手，于是决定自己来证明这些式子。

他后来真的搞出来了，而且还搞出其它一些三角函数的级数，开创了我国无穷级数的先河。他的成果后来统一收集在他的著作《割圆密率捷法》中。里面有关于卡塔