9.9 生日问题

工作纸

题目

STEP II 2007，第 13 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/07-s2-q13

提示

请先自行尝试解题。如果遇到困难，请逐一阅读提示，每读完一个便回到题目再试。标有*停一停*，*试一试*的提示是关键步骤。每个提示之后均留有空位供你演算。

> **想一想**　为甚么计算*没有*两位宾客同一天生日的概率会比较容易？请宾客排成一行：第二位避开第一位生日的机会是多少？第三位同时避开前两位的机会又是多少？

阶段 1　一个有用的近似

涵盖：近似式

> **想一想**　写出 e^{-x} 的级数（series），并代入 $x = r/n$。当 r 远小于 n 时，哪些项可以略去？

阶段 2 生日相同的概率

涵盖：至少两位宾客生日相同

停一停，试一试 把全部 k 个生日各不相同的概率（probability）写成分数 $\frac{365-r}{365}$ 的乘积。把每个因子用指数函数（exponential）近似，再把各指数相加。

阶段 3 二十三

涵盖：机会过半所需的最少人数

想一想 令你求得的概率至少为 $\frac{1}{2}$，然后取对数（logarithm）。利用 $\ln 2 \approx 0.693$，并试验答案附近的整数。

阶段 4　与主人同一天生日

涵盖：与主人生日相同

> **想一想**　对每位宾客而言，生日与主人不同的机会是多少？这里各位宾客是否互相独立（independent）？像之前一样作近似并求解。

背后的数学

平方根定律

若把 365 天换成 n 个等可能的日子，同样的论证给出

$$\mathrm{P}(\text{有人相同，人数为}k) \approx 1 - e^{-k^2/(2n)},$$

所以只要 k 达到 $\sqrt{n}$ 左右，而不是 n，就很可能出现相同的情况。更准确地说，概率在 $k \approx 1.18\sqrt{n}$ 时达到 $\frac{1}{2}$，而在首次出现重复之前你需要遇到的人数，期望值约为 $\sqrt{\pi n/2}$（当 $n = 365$ 时，约为 24.6）。这条*平方根定律*（square-root law）正是巧合如此常见的原因：在任何一大批事件之中，总很可能有某一对在某方面吻合，尽管任何一个指定的吻合都不大可能发生。

生日攻击

电脑会计算文件的*杂凑值*（hash），以检查文件有否被窜改：那是一个指纹，可能的取值有（比方说）$n = 2^{128}$ 个。伪造者若想得到两份杂凑值相同的文件（一份无害的合约和一份欺诈的合约），并不需要试 2^{128} 份文件。根据生日问题的论证，为每份文件各自产生大约 $\sqrt{2^{128}} = 2^{64}$ 个变体，就很可能得到一对相同的杂凑值。这种*生日攻击*（birthday attack）正是现代杂凑函数采用 256 位元或以上的原因：有一半的位元因平方根定律而白白损失。类似的计算亦决定了随机识别码必须有多长，才能令两部电脑永远不会碰巧选中同一个。

谁最先提出?

生日问题通常归功于冯·米塞斯（Richard von Mises），他在 1939 年讨论过这个问题，不过它在此之前已经流传。据说 Harold Davenport 亦更早知道这个问题，只是觉得它太简单而没有发表。按 Warren Weaver 的说法，他在第二次世界大战期间喜欢在军官的晚宴上示范这个问题：他打赌在场二十二人之中有两人同一天生日，结果输了，直至一位女侍应透露她与其中一位军官同一天生日。

延伸练习

1 **首次重复。**试做 STEP II 2014，第 13 题（step.maths.org/questions/14-s2-q13）。这道题研究随机数产生器（random number generator）首次出现重复的值，其实是换了个面貌的生日问题，并把有关概率化成两个出人意表的恒等式（identities）。

2 **三人相同。**大约需要多少人，才会令某三人同一天生日的机会大于一半?（提示：同一天生日的三人组数目的期望值是 $\binom{k}{3}/365^2$。令它约等于 $\ln 2$。）真正的答案是 88。

3 **差之毫厘。**需要多少人，才会令两人的生日在同一天或相邻两天的机会达到一半?

4 **不均匀的生日。**现实中的生日并非均匀（uniform）分布。证明若各日的概率为 $p_1, \ldots, p_{365}$，则两个人同一天生日的机会是 $\sum p_i^2$，并且当所有 p_i 相等时这个值最小。（所以不均匀只会令相同的情况更容易出现。）

延伸阅读

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley），第 II 章。生日问题及其各种近似。
- **Persi Diaconis & Frederick Mosteller, "Methods for studying coincidences"**，*Journal of the American Statistical Association* 84 (1989)。解释巧合为何常见；非常易读。
- **David Hand, *The Improbability Principle***（Bantam）。平方根定律，以及不大可能的事情为何时刻都在发生的其他原因。

完整的引导式解答见第 9.9 节：

https://beyondstep.keithmcnulty.org/zh-hans/ch09/birthday.html

每星期一道 STEP 题目，完整讲解：The STEP Up，以电子邮件免费发送，订阅网址：keithmcnulty.substack.com