9.9 生日問題

工作紙

題目

STEP II 2007，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

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

提示

請先自行嘗試解題。如果遇到困難，請逐一閱讀提示，每讀完一個便回到題目再試。標有停一停，試一試的提示是關鍵步驟。每個提示之後均留有空位供你演算。

> **想一想**　為甚麼計算沒有兩位賓客同一天生日的概率會比較容易？請賓客排成一行：第二位避開第一位生日的機會是多少？第三位同時避開前兩位的機會又是多少？

階段 1　一個有用的近似

涵蓋：近似式

> **想一想**　寫出 e^{-x} 的級數（series），並代入 $x = r/n$。當 r 遠小於 n 時，哪些項可以略去？

階段 2 生日相同的概率

涵蓋：至少兩位賓客生日相同

停一停，試一試 把全部 k 個生日各不相同的概率（probability）寫成分數 $\frac{365-r}{365}$ 的乘積。把每個因子用指數函數（exponential）近似，再把各指數相加。

階段 3 二十三

涵蓋：機會過半所需的最少人數

想一想 令你求得的概率至少為 $\frac{1}{2}$，然後取對數（logarithm）。利用 $\ln 2 \approx 0.693$，並試驗答案附近的整數。

階段 4　與主人同一天生日

涵蓋：與主人生日相同

> **想一想**　對每位賓客而言，生日與主人不同的機會是多少？這裏各位賓客是否互相獨立（independent）？像之前一樣作近似並求解。

背後的數學

平方根定律

若把 365 天換成 n 個等可能的日子，同樣的論證給出

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

所以只要 k 達到 $\sqrt{n}$ 左右，而不是 n，就很可能出現相同的情況。更準確地說，概率在 $k \approx 1.18\sqrt{n}$ 時達到 $\frac{1}{2}$，而在首次出現重複之前你需要遇到的人數，期望值約為 $\sqrt{\pi n/2}$（當 $n = 365$ 時，約為 24.6）。這條*平方根定律*（square-root law）正是巧合如此常見的原因：在任何一大批事件之中，總很可能有某一對在某方面吻合，儘管任何一個指定的吻合都不大可能發生。

生日攻擊

電腦會計算檔案的*雜湊值*（hash），以檢查檔案有否被竄改：那是一個指紋，可能的取值有（比方說）$n = 2^{128}$ 個。偽造者若想得到兩份雜湊值相同的文件（一份無害的合約和一份欺詐的合約），並不需要試 2^{128} 份文件。根據生日問題的論證，為每份文件各自產生大約 $\sqrt{2^{128}} = 2^{64}$ 個變體，就很可能得到一對相同的雜湊值。這種*生日攻擊*（birthday attack）正是現代雜湊函數採用 256 位元或以上的原因：有一半的位元因平方根定律而白白損失。類似的計算亦決定了隨機識別碼必須有多長，才能令兩部電腦永遠不會碰巧選中同一個。

誰最先提出？

生日問題通常歸功於馮·米塞斯（Richard von Mises），他在 1939 年討論過這個問題，不過它在此之前已經流傳。據說 Harold Davenport 亦更早知道這個問題，只是覺得它太簡單而沒有發表。按 Warren Weaver 的說法，他在第二次世界大戰期間喜歡在軍官的晚宴上示範這個問題：他打賭在場二十二人之中有兩人同一天生日，結果輸了，直至一位女侍應透露她與其中一位軍官同一天生日。

延伸練習

1 **首次重複。**試做 STEP II 2014，第 13 題（step.maths.org/questions/14-s2-q13）。這道題研究隨機數產生器（random number generator）首次出現重複的值，其實是換了個面貌的生日問題，並把有關概率化成兩個出人意表的恆等式（identities）。

2 **三人相同。**大約需要多少人，才會令某三人同一天生日的機會大於一半？（提示：同一天生日的三人組數目的期望值是 $\binom{k}{3}/365^2$。令它約等於 $\ln 2$。）真正的答案是 88。

3 **差之毫釐。**需要多少人，才會令兩人的生日在同一天或相鄰兩天的機會達到一半？

4 **不均勻的生日。**現實中的生日並非均勻（uniform）分佈。證明若各日的概率為 $p_1, \ldots, p_{365}$，則兩個人同一天生日的機會是 $\sum p_i^2$，並且當所有 p_i 相等時這個值最小。（所以不均勻只會令相同的情況更容易出現。）

延伸閱讀

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley），第 II 章。生日問題及其各種近似。
- **Persi Diaconis & Frederick Mosteller,「Methods for studying coincidences」**，*Journal of the American Statistical Association* 84 (1989)。解釋巧合為何常見；非常易讀。
- **David Hand, *The Improbability Principle***（Bantam）。平方根定律，以及不大可能的事情為何時刻都在發生的其他原因。

完整的引導式解答見第 9.9 節：

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

每星期一道 STEP 題目，完整講解：The STEP Up，以電郵免費發送，訂閱網址：keithmcnulty.substack.com