詳細解答

STEP II 2007，第 13 題

概率與統計 • 生日問題

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

本題在 *Beyond STEP* 第 9.9 節生日問題中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch09/birthday.html

關鍵思路

沒有任何兩位賓客生日相同的概率（probability），是若干個形如 $\frac{365-r}{365}$ 的因子的乘積，而乘積並不好處理。第一部分把每個因子化成指數函數（exponential），而指數函數的乘積等於以和為指數的指數函數，這個和可以準確地求出。至於主人的問題，各位賓客的事件互相獨立（independent），所以互補事件（complement）的概率只是單獨一個冪，同一個近似方法照樣適用。

解答

近似式

對任意實數 x，

$$\mathrm{e}^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots.$$

設 $x = r/n$。由於 r 遠小於 n，x 很小，x^2 及其後各項與 x 相比可以忽略不計。所以

$$\mathrm{e}^{-r/n} \approx 1 - \frac{r}{n} = \frac{n-r}{n},$$

這正是所求。

最少兩位賓客生日相同

設 A 為 k 位賓客中最少有兩位生日相同的事件。求 $\mathrm{P}(A')$，即 k 個生日全部不同的概率，會比較容易。逐一考慮各位賓客。第一位的生日可以是任何一天。已知首 r 位賓客有 r 個不同的生日，第 $(r+1)$ 位賓客必須避開這 r 天，其概率為 $\frac{365-r}{365}$，因為各人的生日互相獨立，並且在 365 天上均勻分佈（uniform）。把這些概率相乘，

$$\mathrm{P}(A') = \prod_{r=0}^{k-1} \frac{365-r}{365}.$$

若派對的人數不多，這裏每個 r 都遠小於 365，所以由第一部分，取 $n = 365$，

$$\begin{aligned}\mathrm{P}(A') &\approx \prod_{r=0}^{k-1} \mathrm{e}^{-r/365} = \exp\Big(-\frac{1}{365}\sum_{r=0}^{k-1} r\Big) \\ &= \exp\Big(-\frac{1}{365}\cdot\frac{k(k-1)}{2}\Big) = \mathrm{e}^{-k(k-1)/730}.\end{aligned}$$

（$r = 0$ 的因子恰好是 $1 = \mathrm{e}^0$，所以把它包括在內並無影響。）因此

$$\mathrm{P}(A) = 1 - \mathrm{P}(A') \approx 1 - \mathrm{e}^{-k(k-1)/730},$$

這就是所求的結果。

> **想一想** 一次過對這麼多個因子取近似值，難道不令人擔心嗎？誤差可能會累積起來。

每個因子的誤差屬 $(r/365)^2$ 的數量級，而指數本身則屬 $k^2/730$ 的數量級。對於我們所需的 k 值（20 左右），最大的 $r/365$ 約為 0.06，指數的總相對誤差只有百分之幾。當 $k = 23$ 時，近似式給出 0.500，而真確值為 0.507。

機會過半所需的最少人數

我們需要

$$1 - \mathrm{e}^{-k(k-1)/730} \geq \tfrac{1}{2} \iff \mathrm{e}^{-k(k-1)/730} \leq \tfrac{1}{2} \iff \frac{k(k-1)}{730} \geq \ln 2,$$

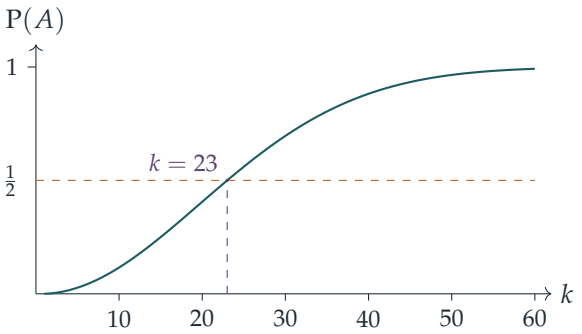
其中先取對數（對數函數是遞增的，所以不等號方向不變），再乘以 -1（不等號方向因而倒轉）。利用 $\ln 2 \approx \frac{253}{365}$，條件變成

$$k(k-1) \geq 730 \times \frac{253}{365} = 2 \times 253 = 506.$$

由於當 $k \geq 1$ 時 $k(k-1)$ 隨 k 遞增，而且

$$22 \times 21 = 462 < 506, \qquad 23 \times 22 = 506,$$

所以最小的值是 $k = 23$。



與主人生日相同

現在設 m 為賓客人數（主人不計算在內）。設 B 為最少有一位賓客與主人生日相同的事件。無論主人的生日是哪一天，每位賓客的生日不是這一天的概率都是 $\frac{364}{365}$，而且各人互相獨立，所以

$$\mathrm{P}(B') = \left(\frac{364}{365}\right)^m \approx \left(\mathrm{e}^{-1/365}\right)^m = \mathrm{e}^{-m/365},$$

這裏用了第一部分的結果，取 $n = 365$ 和 $r = 1$。這一次，無論 m 有多大，$r/n = 1/365$ 都確實很小。

我們需要 $\mathrm{P}(B) \geq \frac{1}{2}$，即 $\mathrm{e}^{-m/365} \leq \frac{1}{2}$。與之前完全一樣，這等價於

$$\frac{m}{365} \geq \ln 2 \approx \frac{253}{365} \iff m \geq 253.$$

所以最少要有 253 位賓客。

? **想一想** 為甚麼答案比 23 大這麼多？

在第一個問題中，任何一對賓客都可以生日相同，而 23 位賓客可組成 $\binom{23}{2} = 253$ 對。在第二個問題中，只有 m 對「賓客與主人」的配對才算數。要有同樣多的配對機會，便需要 $m = 253$，這正是 253 在兩個答案中都出現的原因。

驗算

- 準確計算乘積，可以確認兩個答案。有人生日相同的真確概率，22 位賓客時為 0.4757，23 位時為 0.5073。$1 - (364/365)^m$ 的真確值在 $m = 252$ 時小於 $\frac{1}{2}$，在 $m = 253$ 時為 0.5005。
- $\frac{253}{365} = 0.69315\ldots$，而 $\ln 2 = 0.69315\ldots$，兩者準確至小數點後五位都相同，所以題目給出的近似值不影響任何一個答案。
- 模擬 200 000 場各有 23 人的派對，結果有 50.7% 的派對出現生日相同的情況。

常見錯誤

- 計算了兩位指定賓客生日相同的概率，或者為每一對賓客各加一次 $\frac{1}{365}$。互相重疊的事件，其概率不能直接相加；改為考慮互補事件便可避開這個問題。
- 在第一個乘積中每個因子都用 $\frac{364}{365}$。第 $(r+1)$ 位賓客必須避開 r 天，而不只是一天。
- 在最後一部分把主人也算進賓客人數內。主人是另外一個人，所以需要 253 位賓客。

STEP 之外

這裏的經驗法則（heuristic）是：把 k 件物件放進 N 個盒子，一旦 k 達到大約 $\sqrt{2N\ln 2} \approx 1.18\sqrt{N}$，便很可能出現重複。密碼學（cryptography）中的「生日攻擊」（birthday attack）正是以此為基礎：要在一個 N 位元的雜湊（hash）中找出碰撞（collision），只需大約 $2^{N/2}$ 次嘗試，而不是 2^N 次。這個近似式的背後是泊松極限（Poisson limit）：生日相同的配對數目近似服從平均值為 $\binom{k}{2}/365$ 的泊松分佈，而 $\mathrm{P}(\text{沒有配對相同}) \approx \mathrm{e}^{-k(k-1)/730}$。Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門教材）詳細討論了生日問題及其泊松近似。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。