详细解答

STEP II 2007，第 13 题

概率与统计 · 生日问题

> **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

本题在 *Beyond STEP* 第 9.9 节生日问题中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/birthday.html

关键思路

没有任何两位宾客生日相同的概率（probability），是若干个形如 $\frac{365-r}{365}$ 的因子的乘积，而乘积并不好处理。第一部分把每个因子化成指数函数（exponential），而指数函数的乘积等于以和为指数的指数函数，这个和可以准确地求出。至于主人的问题，各位宾客的事件互相独立（independent），所以互补事件（complement）的概率只是单独一个幂，同一个近似方法照样适用。

解答

近似式

对任意实数 x，

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

设 $x = r/n$。由于 r 远小于 n，x 很小，x^2 及其后各项与 x 相比可以忽略不计。所以

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

这正是所求。

最少两位宾客生日相同

设 A 为 k 位宾客中最少有两位生日相同的事件。求 $\mathrm{P}(A')$，即 k 个生日全部不同的概率，会比较容易。逐一考虑各位宾客。第一位的生日可以是任何一天。已知首 r 位宾客有 r 个不同的生日，第 $(r+1)$ 位宾客必须避开这 r 天，其概率为 $\frac{365-r}{365}$，因为各人的生日互相独立，并且在 365 天上均匀分布（uniform）。把这些概率相乘，

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

若派对的人数不多，这里每个 r 都远小于 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 $1 = \mathrm{e}^0$，所以把它包括在内并无影响。）因此

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

这就是所求的结果。

> ? **想一想** 一次过对这么多个因子取近似值，难道不令人担心吗？误差可能会累积起来。

每个因子的误差属 $(r/365)^2$ 的数量级，而指数本身则属 $k^2/730$ 的数量级。对于我们所需的 k 值（20 左右），最大的 $r/365$ 约为 0.06，指数的总相对误差只有百分之几。当 $k = 23$ 时，近似式给出 0.500，而真确值为 0.507。

机会过半所需的最少人数

我们需要

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

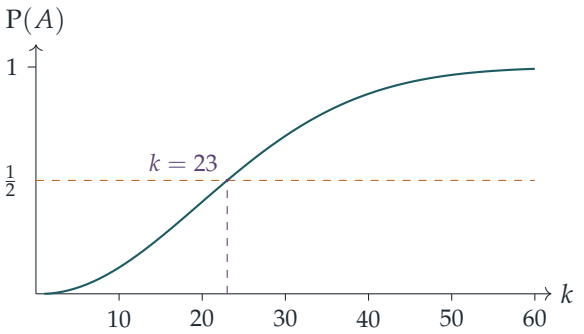
其中先取对数（对数函数是递增的，所以不等号方向不变），再乘以 -1（不等号方向因而倒转）。利用 $\ln 2 \approx \frac{253}{365}$，条件变成

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

由于当 $k \geq 1$ 时 $k(k-1)$ 随 k 递增，而且

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

所以最小的值是 $k = 23$。



与主人生日相同

现在设 m 为宾客人数（主人不计算在内）。设 B 为最少有一位宾客与主人生日相同的事件。无论主人的生日是哪一天，每位宾客的生日不是这一天的概率都是 $\frac{364}{365}$，而且各人互相独立，所以

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

这里用了第一部分的结果，取 $n = 365$ 和 $r = 1$。这一次，无论 m 有多大，$r/n = 1/365$ 都确实很小。

我们需要 $\mathrm{P}(B) \geq \frac{1}{2}$，即 $\mathrm{e}^{-m/365} \leq \frac{1}{2}$。与之前完全一样，这等价于

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

所以最少要有 253 位宾客。

> **想一想** 为甚么答案比 23 大这么多?

在第一个问题中，任何一对宾客都可以生日相同，而 23 位宾客可组成 $\binom{23}{2} = 253$ 对。在第二个问题中，只有 m 对“宾客与主人”的配对才算数。要有同样多的配对机会，便需要 $m = 253$，这正是 253 在两个答案中都出现的原因。

验算

- 准确计算乘积，可以确认两个答案。有人生日相同的真确概率，22 位宾客时为 0.4757，23 位时为 0.5073。$1 - (364/365)^m$ 的真确值在 $m = 252$ 时小于 $\frac{1}{2}$，在 $m = 253$ 时为 0.5005。
- $\frac{253}{365} = 0.69315\ldots$，而 $\ln 2 = 0.69315\ldots$，两者准确至小数点后五位都相同，所以题目给出的近似值不影响任何一个答案。
- 模拟 200 000 场各有 23 人的派对，结果有 50.7% 的派对出现生日相同的情况。

常见错误

- 计算了两位指定宾客生日相同的概率，或者为每一对宾客各加一次 $\frac{1}{365}$。互相重叠的事件，其概率不能直接相加；改为考虑互补事件便可避开这个问题。
- 在第一个乘积中每个因子都用 $\frac{364}{365}$。第 $(r+1)$ 位宾客必须避开 r 天，而不只是一天。
- 在最后一部分把主人也算进宾客人数内。主人是另外一个人，所以需要 253 位宾客。

STEP 之外

这里的经验法则（heuristic）是：把 k 件物件放进 N 个盒子，一旦 k 达到大约 $\sqrt{2N\ln 2} \approx 1.18\sqrt{N}$，便很可能出现重复。密码学（cryptography）中的“生日攻击”（birthday attack）正是以此为基础：要在一个 N 位元的杂凑（hash）中找出碰撞（collision），只需大约 $2^{N/2}$ 次尝试，而不是 2^N 次。这个近似式的背后是泊松极限（Poisson limit）：生日相同的配对数目近似服从平均值为 $\binom{k}{2}/365$ 的泊松分布，而 $\mathrm{P}(\text{没有配对相同}) \approx \mathrm{e}^{-k(k-1)/730}$。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门教材）详细讨论了生日问题及其泊松近似。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。