解答例

STEP II 2007、第 13 問

確率・統計 ・ 誕生日問題

> **STEP II 2007、問 13**
>
> 先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：
>
> step.maths.org/questions/07-s2-q13

この問題は、*Beyond STEP* の第 9.9 節「**誕生日問題**」で詳しく解説されている：https://beyondstep.keithmcnulty.org/ja/ch09/birthday.html

鍵となる考え方

どの 2 人の客も誕生日が一致しない確率（probability）は $\frac{365-r}{365}$ の形の因子の積であり、積は扱いにくい。最初の小問では各因子を指数関数（exponential）に置き換える。指数関数の積は和の指数関数であり、その和は正確に計算できる。主催者の場合は、各客についての事象が独立（independent）なので、余事象（complement）の確率は 1 つのべき乗となり、同じ近似（approximation）が使える。

解答

近似式

任意の実数 x に対して

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

$x = r/n$ とおく。r は n よりはるかに小さいので x は小さく、x^2 以降の項は x に比べて無視できる。よって

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

となり、示すべき式が得られた。

少なくとも 2 人の客の誕生日が一致する場合

k 人の客のうち少なくとも 2 人の誕生日が一致するという事象を A とする。すべての k 人の誕生日が異なる確率 $\mathrm{P}(A')$ を求めるほうが易しい。客を 1 人ずつ考える。1 人目はどの誕生日でもよい。最初の r 人の誕生日が r 個の異なる日であるとき、$(r+1)$ 人目の客はその r 日を避けなければならず、その確率は $\frac{365-r}{365}$ である。誕生日は互いに独立で、365 日にわたって一様（uniform）に分布しているからである。これらを掛け合わせて

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

それほど大きくないパーティーでは、ここに現れるどの r も 365 よりはるかに小さいので、最初の小問で $n = 365$ とすれば

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

（$r = 0$ の因子はちょうど $1 = \mathrm{e}^0$ なので、これを含めても何も変わらない。）したがって

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

となり、求める結果が得られる。

> **考えてみよう** 多くの因子を一度に近似するのは心配ではないか。誤差が積み重なるかもしれない。

各因子の誤差は $(r/365)^2$ 程度であり、一方で指数そのものは $k^2/730$ 程度である。必要となる k の値（20 前後）では、$r/365$ の最大値は約 0.06 であり、指数における相対誤差（relative error）の合計はわずか数パーセントにすぎない。$k = 23$ のとき、近似値は 0.500、真の値は 0.507 である。

五分五分になる最小のパーティー

必要なのは

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

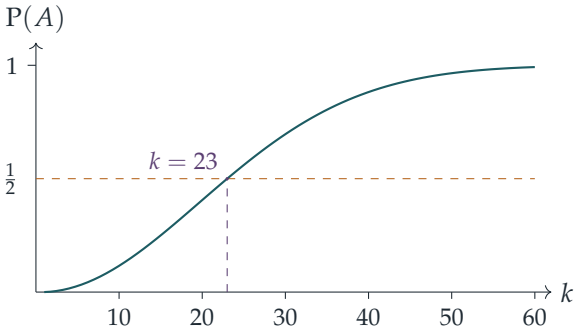
である。ここでは対数（logarithm）をとり（対数は増加関数なので不等号の向きは変わらない）、-1 を掛けた（これにより向きが逆になる）。$\ln 2 \approx \frac{253}{365}$ を用いると、条件は

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

となる。$k \geq 1$ において $k(k-1)$ は k とともに増加し、

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

であるから、最小値は $k = 23$ である。



主催者と誕生日が一致する場合

今度は客の人数を m とする（主催者は客に含めない）。少なくとも 1 人の客が主催者と同じ誕生日であるという事象を B とする。主催者の誕生日が何日であっても、各客は独立に確率 $\frac{364}{365}$ でその日を外すので

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

となる。ここでは最初の小問を $n = 365$、$r = 1$ として用いた。今回は m の大きさに関係なく、$r/n = 1/365$ は本当に小さい。

必要なのは $\mathrm{P}(B) \geq \frac{1}{2}$、すなわち $\mathrm{e}^{-m/365} \leq \frac{1}{2}$ である。先ほどとまったく同様に、これは

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

となる。よって客は少なくとも 253 人必要である。

考えてみよう なぜ答えは 23 よりこれほど大きいのか。

最初の問題では、どの 2 人の客の組でも一致しうるので、23 人の客からは $\binom{23}{2} = 253$ 組ができる。2 番目の問題では、客と主催者の m 組だけが数に入る。一致の機会を同じ数だけ得るには $m = 253$ が必要であり、まさにこれが両方の答えに 253 が現れる理由である。

検算

- 正確な積を計算すると、両方の答えが確かめられる。誕生日が一致する真の確率は、客が 22 人のとき 0.4757、23 人のとき 0.5073 である。$1-(364/365)^m$ の真の値は $m=252$ のとき $\frac{1}{2}$ より小さく、$m=253$ のとき 0.5005 である。
- $\frac{253}{365}=0.69315\ldots$、$\ln 2=0.69315\ldots$ であり、小数第 5 位まで一致するので、与えられた近似はどちらの答えにも影響しない。
- 23 人のパーティーを 200 000 回シミュレーションしたところ、50.7% の割合で誕生日の一致が起きた。

よくある間違い

- **特定の 2 人の客の誕生日が一致する確率を計算してしまう、あるいは各組ごとに $\frac{1}{365}$ を 1 回ずつ足してしまう。**重なりのある事象の確率は単純に足し合わせることはできない。余事象を考えればこの問題を避けられる。
- 最初の積のすべての因子に $\frac{364}{365}$ を使ってしまう。$(r+1)$ 人目の客は 1 日だけでなく r 日を避けなければならない。
- 最後の小問で主催者を客の人数に含めてしまう。主催者は別の人物なので、必要な客は 253 人である。

STEP のその先へ

ここでの発見的な考え方（heuristic）は、N 個の箱に入れた k 個の対象の間で一致が起こりやすくなるのは、k がおよそ $\sqrt{2N\ln 2}\approx 1.18\sqrt{N}$ に達したときだ、というものである。これは暗号理論（cryptography）における「誕生日攻撃（birthday attack）」の基礎となっている。誕生日攻撃は、N ビットのハッシュ（hash）の衝突（collision）を、2^N 回ではなく約 $2^{N/2}$ 回の試行で見つける。この近似の背後にはポアソン極限（Poisson limit）がある。一致する組の個数はおおよそ平均 $\binom{k}{2}/365$ のポアソン分布に従い、$\mathrm{P}(\text{一致なし})\approx \mathrm{e}^{-k(k-1)/730}$ となる。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）では、誕生日問題とそのポアソン近似が詳しく論じられている。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。