9.9 誕生日問題

ワークシート

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

ヒント

まずは自力で問題に取り組んでみよう。行き詰まったら、ヒントを**一つずつ**読み、そのたびに問題に戻ること。**手を止めて、やってみよう**と記したヒントが重要な段階である。各ヒントの後には計算用の余白がある。

> **考えてみよう**　どの2人の客も誕生日が一致しない確率の方が計算しやすいのはなぜか。客を一列に並べてみよ。2人目が1人目の誕生日を避ける確率はいくらか。3人目がその2人の誕生日をともに避ける確率はどうか。

段階1　有用な近似

扱う内容：近似式

> **考えてみよう**　e^{-x} の級数（series）展開を書き出し、$x = r/n$ とおく。r が n よりはるかに小さいとき、どの項を無視できるか。

段階 2 誕生日が一致する確率

扱う内容：少なくとも 2 人の客の誕生日が一致すること

手を止めて、やってみよう k 人の誕生日がすべて異なる確率（probability）を、分数 $\frac{365-r}{365}$ の積として表せ。各因数を指数関数（exponential）で近似し、指数を足し合わせよ。

段階 3 23 人

扱う内容：五分五分の確率になる最小の人数

考えてみよう 求めた確率を $\frac{1}{2}$ 以上とおき、対数（logarithm）をとる。$\ln 2 \approx 0.693$ を用い、答えに近い整数を試してみよ。

段階 4　主催者と誕生日が同じ

扱う内容：主催者と誕生日が一致すること

> **考えてみよう**　各客について、主催者と誕生日が一致しない確率はいくらか。ここでは客どうしは独立（independent）か。前と同様に近似して解け。

背後にある数学

平方根則

365 の代わりに、等確率な n 個の日があるとすると、同じ議論から

$$\mathrm{P}(\text{一致がある、人数}k) \approx 1 - e^{-k^2/(2n)},$$

が得られる。よって一致が起こりやすくなるのは、k が n ではなく $\sqrt{n}$ 程度になったときである。より正確には、確率が $\frac{1}{2}$ に達するのは $k \approx 1.18\sqrt{n}$ のときであり、最初の重複が起こるまでに出会う必要のある人数の期待値は約 $\sqrt{\pi n/2}$ である（$n = 365$ なら約 24.6）。この**平方根則**（square-root law）こそが、偶然の一致がこれほどありふれている理由である。大量の出来事の集まりの中では、特定の一致はどれも起こりにくいにもかかわらず、何らかの組が何らかの形で一致する可能性は高いのである。

誕生日攻撃

コンピュータは、ファイルが改ざんされていないことを**ハッシュ**（hash）を計算して確認する。ハッシュとは、たとえば $n = 2^{128}$ 通りの値をとる指紋のようなものである。同じハッシュをもつ 2 つの文書（無害な契約書と不正な契約書）を作りたい偽造者は、2^{128} 個の文書を試す必要はない。誕生日の議論により、それぞれの変種を約 $\sqrt{2^{128}} = 2^{64}$ 個ずつ生成すれば、おそらく一致が見つかる。この**誕生日攻撃**（birthday attack）があるために、現代のハッシュ関数は 256 ビット以上を用いる。平方根則によってビット数の半分が失われるからである。同様の計算により、2 台のコンピュータが偶然同じものを選ぶことが決してないようにするには、ランダムな識別子をどれだ

け長くすべきかも決められる。

最初に問うたのは誰か

誕生日問題（birthday problem）は通常、1939年にこれを論じたリヒャルト・フォン・ミーゼス（Richard von Mises）に帰せられるが、それ以前から知られていた。Harold Davenport もそれ以前にこれを知っていたが、あまりに簡単に思えたので発表しなかったと言われている。Warren Weaver の語るところによれば、彼は第二次世界大戦中、陸軍将校たちとの夕食会でこれを実演するのを好んだ。その場にいた22人のうち2人の誕生日が一致するほうに賭けて負けたが、そのときウェイトレスが、自分の誕生日が将校の1人と同じであることを明かしたという。

次の一歩

1. **最初の重複。** STEP II 2014、問13（step.maths.org/questions/14-s2-q13）に挑戦せよ。乱数生成器から最初に重複して現れる値を調べる問題で、形を変えた誕生日問題である。その確率から2つの意外な恒等式が導かれる。

2. **3人が一致。** 誰か3人の誕生日が一致する確率が半分を超えるには、おおよそ何人必要か。（ヒント：誕生日が一致する3人組の個数の期待値は $\binom{k}{3}/365^2$ である。これを約 $\ln 2$ とおけ。）真の答えは88である。

3. **ニアミス。** 2人の誕生日が同じ日または隣り合う日である確率が半分になるには、何人必要か。

4. **偏りのある誕生日。** 現実の誕生日は一様ではない。各日の確率が $p_1, \ldots, p_{365}$ であるとき、2人の誕生日が一致する確率は $\sum p_i^2$ であり、これはすべての p_i が等しいときに最小となることを示せ。（よって一様でないことは、一致をより起こりやすくするだけである。）

さらに学ぶために

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley）、第II章。誕生日問題とその近似。
- **Persi Diaconis & Frederick Mosteller, "Methods for studying coincidences"**, *Journal of the American Statistical Association* 84 (1989)。偶然の一致がありふれている理由。非常に読みやすい。
- **David Hand, *The Improbability Principle***（Bantam）。平方根則をはじめ、起こりそう

にないことが絶えず起こる理由を解説する。

詳しい解説付きの解答は第 9.9 節にある：

https://beyondstep.keithmcnulty.org/ja/ch09/birthday.html

毎週 1 問の STEP の問題を丁寧に解き明かす The STEP Up。メールで無料配信中：keithmcnulty.substack.com