Paradoks dni urodzin i pokrewne, czyli o pewnych zagadnieniach związanych z rozmieszczeniem kul w komórkach

*Tomasz NIKODEM**

Motywacją do napisania artykułu o powyższym tytule było następujące zadanie, znane w literaturze jako paradoks dni urodzin: jak liczna powinna być grupa osób, aby z prawdopodobieństwem nie mniejszym od $1/2$ znalazły się w niej co najmniej dwie osoby obchodzące urodziny tego samego dnia. Zakłada się, że rozkład dni urodzin jest rozkładem jednostajnym na zbiorze $\{1, 2, \ldots, 365\}$, co mało odbiega od rzeczywistości, natomiast lata przestępne pomijamy.

Jaki związek mogą mieć kule i komórki z ludźmi i ich datami urodzin? Otóż, układ ponumerowanych od 1 do 365 komórek można interpretować następująco: pierwsza komórka to pierwszy stycznia, druga – drugi stycznia, ostatnia – Sylwester. Fakt, że w szufladzie o numerze 31 jest kula, interpretujemy następująco: w grupie osób jest osoba urodzona 31 stycznia. Komórki są rozróżnialne, bowiem każda z nich ma inny numer. Podobne założenie dotyczy kul: każda z nich odpowiada dacie urodzenia innej osoby, naturalne zaś jest założenie, że osoby są rozróżnialne.

Obliczymy prawdopodobieństwo tego, że w k-osobowej grupie znajdą się przynajmniej dwie osoby urodzone tego samego dnia. Chwilę później wyznaczymy minimalne k, dla którego obliczone prawdopodobieństwo jest nie mniejsze niż $1/2$. Odpowiedź, którą niebawem uzyskamy, jest zaskakująco niska: wystarczą zaledwie 23 osoby, by taka sytuacja miała miejsce.

Za Ianem Stewartem podaję, że na pytanie o minimalną liczbę osób, dla której z prawdopodobieństwem co najmniej $1/2$ znajdziemy dwie osoby urodzone tego samego dnia, które „stawiano studentom kierunków uniwersyteckich [w USA], *średnia* z odpowiedzi wynosiła 385". Natomiast już dla 366 osób z zasady szufladkowej Dirichleta wynika, iż z prawdopodobieństwem 1 znajdą się dwie osoby obchodzące urodziny tego samego dnia. Wyniki ankiety dobrze uzasadniają tytuł: paradoks dni urodzin.

Przystąpmy do rozwiązania zadania sformułowanego na początku artykułu. Zdarzeniami elementarnymi są ciągi długości k o wyrazach w zbiorze $\{1, 2, \ldots, 365\}$. Wszystkich takich ciągów jest 365^k. Oznaczmy przez p_k prawdopodobieństwo zdarzenia przeciwnego do rozpatrywanego, czyli zdarzenia polegającego na tym, że k-elementowy ciąg będzie różnowartościowy. Wówczas:

$$(1)\quad p_k = \frac{365 \cdot 364 \cdot \ldots \cdot (365-k+1)}{365^k} = \left(1-\frac{1}{365}\right) \cdot \ldots \cdot \left(1-\frac{k-1}{365}\right) \quad \text{dla } k \leqslant 365$$

oraz

$$p_k = 0 \quad \text{dla } k > 365.$$

Teraz naszym zadaniem jest znalezienie minimalnego k, dla którego $p_k \leqslant 1/2$. W dalszym ciągu będziemy korzystać z oczywistej obserwacji, iż dla interesujących nas k, czyli $k \leqslant 365$, ciąg (p_k) jest malejący. Dla $k > 365$ omawiany ciąg jest stały.

Wykorzystując nierówność $1 + x \leqslant e^x$ (prawdziwą dla dowolnej liczby rzeczywistej x), można szacować prawą stronę wzoru (1):

$$(2)\quad p_k \leqslant e^{-\frac{1}{365}} \cdot \ldots \cdot e^{-\frac{k-1}{365}} = e^{-\frac{1+2+\ldots+(k-1)}{365}} = e^{-\frac{k(k-1)}{730}}.$$

Aby zachodziło $p_k \leqslant 1/2$, wystarczy wyznaczyć minimalne k (naturalne), dla którego zachodzi

$$(3)\quad k^2 - k - 730 \cdot \ln 2 \geqslant 0.$$

Pierwszym dodatnim rozwiązaniem nierówności (3) jest

$$(4)\quad \frac{1+\sqrt{1+4\cdot 730 \cdot \ln 2}}{2} = 22{,}99994315\ldots,$$

czyli rozwiązanie naszego zadania to $k = 23$. Rezultat jest zgodny z wynikiem przytoczonym na początku artykułu; niemniej, aby to dokładnie sprawdzić, warto wykonać rachunki przy użyciu komputera. Otrzymujemy: $p_{23} = 0{,}493\ldots$ oraz $p_{22} = 0{,}524\ldots$, czyli pomimo szacowania wykonanego w (2) otrzymujemy prawidłowy wynik.

Przeanalizujmy, jak się mają przybliżenia do poprawnej odpowiedzi, gdy zwiększymy bądź zmniejszymy liczbę szuflad (możliwości dni urodzin). Taka sytuacja będzie możliwa, gdy przeniesiemy się na chwilę, na przykład, na Marsa, gdzie jeden pełny obieg wokół Słońca trwa 687 ziemskich dni, a co za tym idzie, na Marsie mamy 687 różnych dat urodzin. Stawiamy pytanie analogiczne jak na Ziemi: jak liczna powinna być grupa Marsjan, by z prawdopodobieństwem co najmniej $1/2$ znalazło się wśród nich przynajmniej dwóch obchodzących urodziny tego samego dnia? Zakładamy, że rozkład narodzin na Marsie jest jednostajny na $\{1, 2, \ldots, 687\}$, choć właściwie nie mamy do tego żadnych podstaw.

Ponieważ liczby 365 oraz 687 tak naprawdę nie są bardziej interesujące niż pozostałe, przeanalizujmy powyższy przykład w pełnej ogólności. Mamy n możliwych dni urodzin. Oznaczmy przez p_k^n prawdopodobieństwo, że k-elementowy ciąg o wyrazach w zbiorze $\{1, 2, \ldots, n\}$ jest różnowartościowy. Aby znaleźć wzór na p_k^n, wystarczy we wzorze (1) każdorazowo w miejsce 365 wpisać n. Wykonując szacowanie analogiczne do (2), otrzymujemy, że warunkiem wystarczającym, by zachodziło $p_k^n \leqslant 1/2$, jest spełnienie nierówności $k^2 - k - 2 \cdot n \cdot \ln 2 \geqslant 0$, której pierwszym dodatnim rozwiązaniem jest

$$\frac{1+\sqrt{1+4\cdot 2\cdot n \cdot \ln 2}}{2} = \frac{1}{2} + \frac{\sqrt{1+4\cdot 2 \cdot n \cdot \ln 2}}{2}.$$

*student, Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Aby wzór stał się przyjemniejszy (tracąc na dokładności, lecz co najwyżej o jeden), decydujemy się zapisać go w przybliżonej postaci: $k_0^n = \sqrt{2 \cdot \ln 2} \cdot \sqrt{n}$ – jako rozwiązanie wystarczy brać minimalną liczbę całkowitą nie mniejszą od k_0^n. W ogólnym przypadku odpowiedź jest rzędu $\sqrt{n}$ razy stała $\sqrt{2 \cdot \ln 2} = 1{,}177410\ldots$ Żeby zobaczyć, jak dobrze przybliżone wyniki uzyskane dzięki wyznaczeniu k_0^n oddają faktyczne, prawidłowe wyniki uzyskane numerycznie dla pewnych przypadkowych, ale konkretnych n, warto spojrzeć na trzecią oraz czwartą kolumnę tabeli na dole strony.

Przystąpimy teraz do analizy nieco innego zagadnienia. Przypuśćmy, że dane jest 365 szuflad, co odpowiada 365 różnym datom urodzin. Założmy dalej, że mamy dostatecznie dużo kul, które będziemy umieszczać kolejno w losowo wybranej (zgodnie z rozkładem jednostajnym na $\{1, 2, \ldots, 365\}$) szufladzie. Wyróżnijmy jedną z szuflad przed rozpoczęciem doświadczenia, powiedzmy o numerze jeden, czyli pierwszy stycznia. Standardowo kula w szufladzie oznacza, że w grupie jest osoba urodzona danego dnia. Zagadnienie wygląda następująco: jak liczna powinna być grupa osób, aby z prawdopodobieństwem co najmniej 1/2 znalazła się wśród nich co najmniej jedna, która obchodzi urodziny właśnie pierwszego stycznia? Wybrana data nie jest żadnym magicznym dniem – chodzi jedynie o to, że została ona ustalona przed wykonaniem doświadczenia. Powyższe zagadnienie można zinterpretować następująco: jak liczna powinna być grupa osób, aby z prawdopodobieństwem co najmniej 1/2 znalazła się wśród nich co najmniej jedna osoba urodzona tego samego dnia co ja?

Przechodząc znów do rozważania zdarzenia przeciwnego, szukamy prawdopodobieństwa tego, że przy losowym rozmieszczeniu k kul w 365 komórkach otrzymamy rozmieszczenie, w którym wyróżniona komórka jest pusta. Innymi słowy, interesuje nas zdarzenie, że k kul zostanie umieszczonych w komórkach o numerach $2, 3, \ldots, 365$. Zdarzeniami elementarnymi są ciągi długości k o wyrazach w zbiorze $\{1, 2, \ldots, 365\}$. Rozmieszczenia sprzyjające omawianemu zdarzeniu przeciwnemu to ciągi długości k o wyrazach w zbiorze $\{2, 3, \ldots, 365\}$ – nie dopuszczamy możliwości, że jakakolwiek kula trafi do wyróżnionej szuflady numer jeden. Jest ich 364^k. Oznaczmy przez b_k szukane prawdopodobieństwo z treści zadania. Otrzymujemy:

$$b_k = 1 - \frac{364^k}{365^k} = 1 - \left(1 - \frac{1}{365}\right)^k. \tag{5}$$

W dalszym ciągu skorzystamy z tego, że ciąg (b_k) jest rosnący. Zwróćmy uwagę, że podczas gdy w grupie 366-osobowej z prawdopodobieństwem 1 znajdą się dwie osoby urodzone tego samego dnia, to dla dowolnie dużej grupy osób nie możemy mieć pewności, że znajdziemy w niej osobę urodzoną, na przykład, 1 stycznia.

Pytamy o to, kiedy b_k jest nie mniejsze niż 1/2. Wyznaczymy nierówność, którą ma spełniać k:

$$1 - \left(1 - \frac{1}{365}\right)^k \geqslant \frac{1}{2} \tag{6}$$

$$\Leftrightarrow \; k \geqslant -\frac{\ln 2}{\ln\left(1 - \frac{1}{365}\right)} = 252{,}651988\ldots,$$

czyli należy brać $k = 253$. Dokładne wyniki, uzyskane za pomocą komputera, są następujące: $b_{252} = 0{,}499105\ldots$ oraz $b_{253} = 0{,}500477\ldots$

W ogólnym przypadku zamiast rozważać 365 możliwości dni urodzin, możemy analogicznie obliczyć, kiedy b_k^n (czyli prawdopodobieństwo tego, że przy losowym rozmieszczeniu k kul w n komórkach wyróżniona przed rozpoczęciem doświadczenia komórka jest niepusta) jest nie mniejsze niż 1/2. Okazuje się, że wystarcza, by

$$k \geqslant -\frac{\ln 2}{\ln(1 - 1/n)}.$$

Aby zyskać na estetyce powyższego wzoru (kosztem jego dokładności), można szacować prawą stronę, wykorzystując nierówność $\ln(1 + x) \leqslant x$ dla $x = -1/n$. Otrzymujemy w ten sposób oszacowanie $k \geqslant n \ln 2$. Oczywiście, interesuje nas najmniejsze $k \in \mathbb{N}$, które spełnia zadaną nierówność. By przekonać się, jak dobry jest „bardziej estetyczny wzór”, warto spojrzeć na ostatnie dwie kolumny poniższej tabeli.

Na zakończenie odnotujmy, że podczas gdy odpowiedź na pytanie będące uogólnieniem paradoksu dni urodzin zachowywała się jak $\sqrt{n}$ ze stałą $1{,}177410\ldots$, odpowiedź na drugie pytanie rośnie liniowo wraz ze wzrostem n ze stałą $\ln 2 = 0{,}693147\ldots$

Literatura

Ian Stewart, *Co za traf!*, Świat Nauki 8/1998, ss. 74–75.

nazwa planety	n – czas obiegu wokół Słońca w dniach ziemskich (zaokrąglony do liczb całkowitych)	minimalne $k \in \mathbb{N}$ nie mniejsze od k_0^n (wynik przybliżony)	minimalne $k \in \mathbb{N}$, aby zachodziło $p_k^n \leqslant 1/2$ (wynik numeryczny)	minimalne $k \in \mathbb{N}$, aby zachodziła nierówność $k \geqslant \ln 2 \cdot n$ (wynik przybliżony)	minimalne $k \in \mathbb{N}$, aby zachodziła nierówność $k \geqslant -\frac{\ln 2}{\ln(1-1/n)}$ (wynik numeryczny)
Merkury	$n = 88$	12	12	61	61
Wenus	$n = 225$	18	18	156	156
Ziemia	$n = 365$	23	23	253	253
Mars	$n = 687$	31	32	477	476
Jowisz	$n = 4\,333$	78	78	3 004	3 004
Saturn	$n = 10\,756$	123	123	7 456	7 456
Uran	$n = 30\,708$	207	207	21 286	21 285
Neptun	$n = 60\,223$	289	290	41 744	41 744