rozmowy, wszystkich rozmów musiało być zatem co najmniej $2n-4$. Ponieważ mieliśmy do czynienia z ciągiem optymalnym, nierówność $\mathtt{X} \geqslant 2n-4$ została udowodniona.

Przedstawiony problem zaczął krążyć wśród matematyków na początku lat 70. XX wieku i bardzo szybko doczekał się rozwiązania (np. [1]). Powyższy dowód zaczerpnięty został z pracy [2], w której w pełni przeanalizowano sytuację bliższą współczesnym udogodnieniom technologicznym, mianowicie plotkarze mogą wymieniać się informacjami podczas konferencji, w których może uczestniczyć co najwyżej K osób. Możliwych uogólnień i związanych z tematem pytań jest zresztą bardzo wiele i do dziś pojawiają się publikacje (np. [3]), których źródła można doszukać się w naszej wdzięcznej zagadce. Plotka głosi, że nie jest to ostatni raz, kiedy pojawia się ona w *Delcie*.

Literatura:

[1] Robert Tijdeman, *On a telephone problem*, Nieuw Archief voor Wiskunde (3), XIX, (1971), 188–192.
[2] Daniel Kleitman, James Shearer, *Further gossip problems,* Discrete Mathematics 30.2 (1980): 151-156.
[3] Krzysztof Apt, Eryk Kopczyński, Dominik Wojtczak, *On the Computational Complexity of Gossip Protocols,* IJCAI (2017), 765–771.

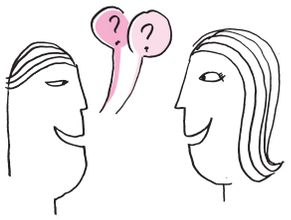
O sztuce zadawania pytań

*Damian NIWIŃSKI**

* Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

– *No to do jutra! Czy spotkamy się znów o dziesiątej rano?*
– *Niestety, Jasiu, rano nie mogę, mam próbę orkiestry...*
– *Aa... Nie wiedziałem, Małgosiu, że grasz w orkiestrze! Na jakim instrumencie?*
– *Zgadnij!* – *Małgosia uśmiechnęła się.* – *Myślę, że wystarczą ci trzy pytania. Jesteś przecież matematykiem.*

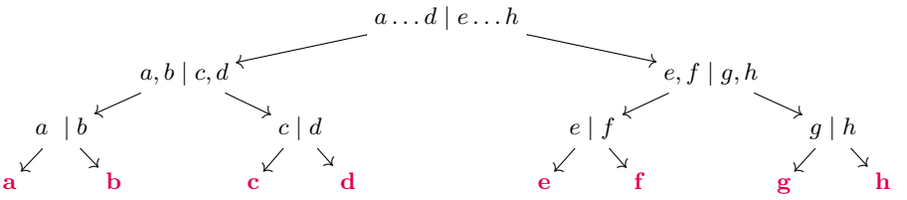
– *No cóż... współczesna orkiestra liczy kilkadziesiąt różnych instrumentów. Ale, choć znamy się już od tygodnia, jeszcze nie widziałem cię z futerałem. Chyba więc twój instrument nie jest łatwo przenośny. Może kontrabas, może fortepian, może perkusja...* – *Jaś zamyślił się.*



– *Czy Twój instrument ma struny?*
– *Nie!*
– *A klawiaturę?*
– *Tak!*
– *A zatem są to organy... muszę koniecznie cię usłyszeć!*
– *Zapraszam na koncert, za tydzień gramy Symfonię Organową Saint–Saënsa!*

* * *

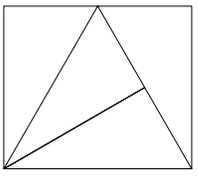
Jaś poradził sobie w dwóch pytaniach, choć zapewne miał trochę szczęścia. Zadawanie pytań tak, by wyciągnąć maksimum wiedzy, będzie tematem naszych rozważań. Będziemy zwykle dopuszczać tylko dwie odpowiedzi: tak lub nie. Większość pytań da się sprowadzić do tej postaci, czasem przez zastąpienie jednego pytania serią, jak to było z pytaniem o instrument w powyższej rozmowie. Spójrzmy na rzecz abstrakcyjnie. Chcemy obmyślić taką strategię zadawania pytań, by jak najszybciej dojść do celu. Gdy poszukiwany obiekt pochodzi ze zbioru o n elementach, o których niewiele wiemy, to rozsądną strategią jest podzielenie naszego zbioru na dwie części o tej samej liczności (być może z dokładnością do jednego elementu) i zapytanie, czy nasz obiekt znajduje się w pierwszej części (jeśli nie – jest w drugiej). Dalej postępujemy w ten sam sposób, aż nasz zbiór stanie się jednoelementowy. Proces ten można przedstawić jako drzewo, gdzie węzłami są pytania, a przejścia w lewo lub w prawo zależą od otrzymanych odpowiedzi. Na przykład, gdy poszukiwany obiekt jest jedną z 8 liter a, b, c, d, e, f, g, h, nasza strategia może wyglądać tak:



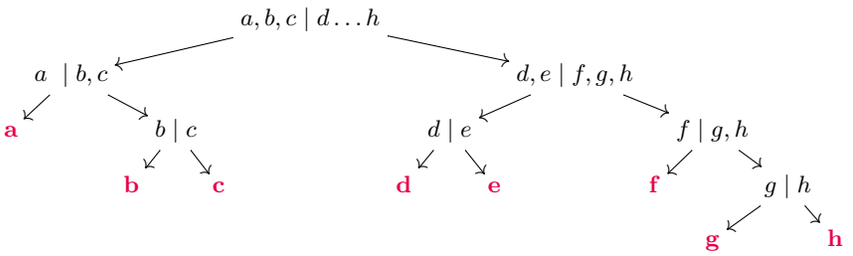
Kiedy otrzymamy odpowiedzi: tak, nie, tak, wiemy, że poszukiwanym obiektem jest c. W każdym przypadku zadamy 3 pytania, czyli binarny logarytm z $n = 8$.

Ogólnie, gdy $2^k < n \leqslant 2^{k+1}$, wtedy postępując w podobny sposób zadamy k lub $k+1 = \lceil \log_2 n \rceil$ pytań, co można łatwo sprawdzić przez indukcję po n.

Rozwiązanie zadania M 1729.
Nie. Rozważmy prostokąt o bokach 2 i $\sqrt{3}$ podzielony w następujący sposób:



Gdyby jednak Jaś użył tej strategii dla instrumentów w orkiestrze symfonicznej, nie osiągnąłby szybko celu. Pozostając przy przykładzie z literami – spróbujmy innych strategii. Możemy stawiać pytania mniej symetrycznie, na przykład



Wtedy dla litery a zadamy tylko 2 pytania. Jednak średnia liczba pytań jest równa

$$(2 + 5 \cdot 3 + 2 \cdot 4) \cdot \tfrac{1}{8} = 3\tfrac{1}{8},$$

a więc gorzej; podobnie będzie dla każdej innej niesymetrycznej strategii.

Dla dowolnego n strategia symetryczna zawsze gwarantuje najlepszą średnią, ale dla n niebędących potęgami 2 strategia niesymetryczna może czasem być równie dobra.

Co innego, gdy o naszych obiektach już coś wiemy i niektórych spodziewamy się bardziej niż innych. Tak bywa w rzeczywistym świecie, a także w popularnej grze w 20 pytań, o której w Δ^{07}_{88} pisał Wojciech Guzicki. Ujmując rzecz matematycznie, poszukiwany obiekt jest wartością pewnej *zmiennej losowej* X, przyjmującej wartości w skończonym zbiorze $\mathcal{X}$, wartość x z *prawdopodobieństwem* $\Pr(X = x)$, w skrócie $\Pr(x)$. Gdy rozkład jest jednostajny, wartość oczekiwana liczby pytań jest średnią arytmetyczną. Jeśli jednak na przykład

$$\Pr(a) = \tfrac{1}{4}, \quad \Pr(b) = \Pr(c) = \Pr(d) = \Pr(e) = \Pr(f) = \tfrac{1}{8}, \quad \Pr(g) = \Pr(h) = \tfrac{1}{16},$$

to *oczekiwana* liczba pytań przy naszej drugiej strategii jest równa

$$2 \cdot \tfrac{1}{4} + 5 \cdot 3 \cdot \tfrac{1}{8} + 2 \cdot 4 \cdot \tfrac{1}{16} = 2\tfrac{7}{8},$$

a więc lepiej niż dla strategii symetrycznej!

Jaka idea kryje się za naszą nową strategią? Otóż zamiast dzielić zbiór na części o równej (lub prawie równej) liczności, dzielimy go teraz na części o możliwie bliskich prawdopodobieństwach. W naszym przykładzie są to nawet równości: $\Pr(X \in \{a, b, c\}) = \Pr(X \in \{d, e, f, g, h\})$, $\Pr(X \in \{d, e\}) = \Pr(X \in \{f, g, h\})$, i podobnie jest w każdym węźle drzewa.

Zapamiętajmy: pytanie jest najpomocniejsze wtedy, gdy obie dopuszczalne odpowiedzi są jednakowo prawdopodobne.

Wracając do przykładu: czy wartość $2\frac{7}{8}$ jest dla zadanego rozkładu optymalna?

Zauważmy, że prawdopodobieństwa są tu szczególnej postaci: $\left(\frac{1}{2}\right)^{\ell}$, gdzie ℓ jest liczbą naturalną. Natomiast liczba pytań prowadząca do obiektu o takim prawdopodobieństwie jest przy naszej strategii właśnie równa ℓ. Wartość oczekiwaną liczby pytań wyraziliśmy więc formułą:

$$H(X) = \sum_{x \in \mathcal{X}} \Pr(x) \cdot \log_2 \frac{1}{\Pr(x)} \tag{1}$$

(do oznaczenia H(X) powrócimy za chwilę). Widzimy, że im bardziej prawdopodobny jest dany obiekt, tym szybciej zostanie odgadnięty, co wydaje się rozsądne. Istotnie, dla prawdopodobieństw będących całkowitymi potęgami $\frac{1}{2}$ formuła (1) gwarantuje optymalność, co można sprawdzić przez indukcję po n. Co jednak z przypadkiem ogólnym? Problemem może być $\Pr(x) = 0$, co załatwiamy konwencją, że $0 \cdot \log_2 \frac{1}{0} = 0$, gdyż $\lim_{y \to 0} y \log_2 \frac{1}{y} = 0$. Wtedy formuła (1) ma zawsze sens, ale liczby $\log_2 \Pr(x)$ mogą nie być całkowite – i wtedy nie są liczbami pytań w żadnej strategii.

Otóż metodami analitycznymi można wykazać, że oczekiwana liczba pytań $S(X)$ spełnia przy każdej strategii nierówność

$$H(X) \leqslant S(X), \tag{2}$$

a przy optymalnej odległość między tymi wielkościami jest nie większa niż 1. Zależność tę odkrył Claude Shannon, który w swojej pionierskiej pracy

Rozwiązanie zadania M 1730.
Na początku zauważmy, że w każdym ruchu liczba liczb na tablicy zmniejsza się o 1. Ponadto suma liczb się nie zmienia, gdyż

$$a + b + c + (a + b + c) = \\ = (a + b) + (b + c) + (c + a).$$

Równość

$$(a + b)^2 + (b + c)^2 + (c + a)^2 = \\ = a^2 + b^2 + c^2 + (a + b + c)^2$$

pokazuje, że suma kwadratów również nie ulega zmianie. Wobec tego jeśli $a_1, a_2, \ldots, a_n$ oznaczają liczby, które są w pewnym momencie wypisane na tablicy, to

$$\sum_{i=1}^{n} a_i = \frac{1}{2} \cdot 2022 \cdot 2023,$$

$$\sum_{i=1}^{n} a_i^2 = \frac{1}{6} \cdot 2022 \cdot 2023 \cdot 4045.$$

Z nierówności Cauchy'ego-Schwarza mamy

$$n \sum_{i=1}^{n} a_i^2 \geqslant \Big(\sum_{i=1}^{n} a_i\Big)^2,$$

więc

$$n \geqslant \frac{(\frac{1}{2} \cdot 2022 \cdot 2023)^2}{\frac{1}{6} \cdot 2022 \cdot 2023 \cdot 4045} \approx 1516{,}9.$$

Oznacza to, że proces musi skończyć się po mniej niż $2022 - 1516 = 506$ sekundach, a to mniej niż 9 minut.

z 1948 roku *A Mathematical Theory of Communication* wprowadził funkcję H i nadał jej nazwę *entropii*, podobno za radą Johna von Neumanna.

Anegdota podana za: Jimmy Soni & Rob Goodman, *A Mind at Play. How Claude Shannon Invented the Information Age*, Simon & Schuster, 2017.

„Po pierwsze, jest to solidny termin w fizyce – miał powiedzieć wybitny matematyk młodszemu koledze – a co ważniejsze, nikt do końca nie wie, co to jest, co da panu atut w dyskusji".

W istocie nierówność (2) odnosi się nie tylko do strategii zadawania pytań, ale także do kodowania wartości zmiennej losowej przez ciągi bitów o zmiennej długości. Na przykład nasze powyższe strategie wyznaczają takie kodowania:

Kodowanie to taka funkcja c przyporządkowująca literom (lub innym obiektom) ciągi bitów, że z ciągu $c(s_1)c(s_2)\dots c(s_m)$ można jednoznacznie odtworzyć ciąg $s_1 s_2 \dots s_m$.

a	b	c	d	e	f	g	h
000	001	010	011	100	101	110	111
00	010	011	100	101	110	1110	1111

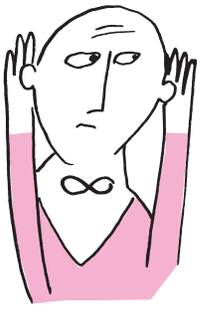
Celem jest tu osiągnięcie minimalnej średniej długości zapisu, co stanowi zagadnienie kompresji, o którym w obecnym numerze *Delty* pisze na stronie 15 Tomasz Kazana, a w Δ^{11}_{21} pisał Jarosław Duda.

Teoria informacji

Co właściwie wyraża funkcja H?

Shannon podjął pytanie, jakie w połowie XX wieku nabrało szczególnej wagi: jak skutecznie przesyłać informację pomimo zakłóceń na łączu. Nie chodzi tu o eliminację przekłamań, ale o komunikację pomimo nich: np. gdy w ciągu zer i jedynek przekłamaniu ulegnie każdy bit, oryginalną wiadomość odtworzymy bez trudu! Studia nad tym problemem doprowadziły Shannona do matematycznej definicji *informacji*, niezależnej od treści, jakich dana informacja dotyczy.

Dla ilustracji przypuśćmy, że rzucamy dwukrotnie kostką do gry i chcielibyśmy poznać wyniki obu rzutów, ale dane jest nam jedynie poznać ich sumę lub ich iloczyn. Która opcja da nam więcej informacji? Intuicja podpowiada, że to pytanie ma jakiś sens – teoria Shannona pozwala ściśle na nie odpowiedzieć.



Użyteczne jest pojęcie entropii zmiennej losowej Y (przyjmującej wartości w zbiorze $\mathcal{Y}$) *pod warunkiem* zadanym przez zmienną losową X. Otóż jeśli ustalimy wartość x przyjmowaną przez zmienną X (z niezerowym prawdopodobieństwem) i rozważymy prawdopodobieństwa warunkowe $\Pr(Y = y \mid X = x)$, w skrócie $\Pr(y \mid x)$, to możemy utworzyć wielkość analogiczną do formuły (1):

$$H(Y \mid x) = \sum_{y \in \mathcal{Y}} \Pr(y \mid x) \cdot \log_2 \frac{1}{\Pr(y \mid x)}. \tag{3}$$

Jeśli teraz uśrednimy tę wielkość po x, otrzymamy *entropię warunkową*:

$$H(Y \mid X) = \sum_{x \in \mathcal{X}} H(Y \mid x) \cdot \Pr(X = x). \tag{4}$$

Pamiętamy o naszej konwencji, że $0 \cdot \log \frac{1}{0} = 0$.

Zauważmy, że gdy Y zależy funkcyjnie od X, tzn. dla każdego x takiego, że $\Pr(x) > 0$, istnieje y takie, że $\Pr(y \mid x) = 1$, wtedy $\Pr(y \mid x) \cdot \log_2 \frac{1}{\Pr(y|x)}$ jest zawsze zerem i w konsekwencji $H(Y \mid X) = 0$. Kiedy natomiast X i Y są niezależne (czyli $\Pr(y \mid x) = \Pr(y)$), wtedy $H(Y \mid X) = H(Y)$. Otóż Shannon zdefiniował wzajemną informację między zmiennymi X i Y jako różnicę:

$$I(X;Y) = H(Y) - H(Y \mid X). \tag{5}$$

Można udowodnić, że wielkość ta jest zawsze nieujemna, a także że $I(X;Y) = I(Y;X)$. Z własności entropii warunkowej wynika, że dla X i Y niezależnych wzajemna informacja jest zerem, a maksymalną wartość $\min(H(X), H(Y))$ osiąga przy zależności funkcyjnej między nimi. W pewnym sensie $I(X;Y)$ jest miarą zależności między X i Y. Czytelnik ma prawo uważać tę definicję informacji za nieco abstrakcyjną, dlatego zilustrujemy ją jeszcze jednym przykładem, powracając do tytułowego problemu: sztuki zadawania pytań.

Rozwiązanie zadania M 1731.
Niech $b = (n^2+1)a + n$. Wówczas

$$b^2 + 1 \equiv a^2 + 1 \equiv 0 \pmod{n}$$

oraz

$$b^2 + 1 \equiv n^2 + 1 \equiv 0 \pmod{n^2+1}.$$

Ponieważ n i n^2+1 są względnie pierwsze, więc $n(n^2+1)$ dzieli b.

Hobbit ma przed sobą n ponumerowanych skrytek. W jednej z nich znajduje się pierścień. Dla ułatwienia przyszłych rachunków zakładamy, że n jest podzielne przez 6. Hobbit chce znaleźć pierścień, ale nie ma dość czasu, by przejrzeć wszystkie skrytki. Z pomocą śpieszą dwa trolle, z których pierwszy gotów jest wskazać, w której połowie skrytek znajduje się pierścień: czy o numerach w przedziale $[1, \frac{n}{2}]$, czy w $[\frac{n}{2}+1, n]$. Drugi troll gotów jest wskazać, w której części trzeciej, czyli w którym z przedziałów $[1, \frac{n}{3}]$, $[\frac{n}{3}+1, \frac{2n}{3}]$ lub $[\frac{2n}{3}+1, n]$ znajduje się pierścień. Problem w tym, że pierwszy troll co trzeci raz kłamie, a drugi kłamie co drugi raz. Hobbit może zadać tylko jedno pytanie. Którego trolla bardziej warto pytać?

Jak widzimy, drugi troll daje dokładniejszą informację, ale niestety częściej kłamie. Czytelnik może zastanowić się przez chwilę, jak sam podszedłby do tego zagadnienia. Możemy zauważyć, że gdyby hobbit przeszukiwał skrytki po kolei, a pierścień znajdował się – pechowo – w ostatniej skrytce, to przy wskazówce pierwszego trolla oczekiwany czas byłby $\frac{2}{3} \cdot \frac{n}{2} + \frac{1}{3} \cdot n = \frac{2}{3} \cdot n$, a przy wskazówce drugiego $\frac{1}{2} \cdot (\frac{n}{3} + n)$, a więc tyle samo. Gdyby natomiast losował ze wskazanego obszaru, to w pierwszym przypadku szansa sukcesu byłaby $\frac{2}{3} \cdot \frac{2}{n} = \frac{4}{3n}$, a w drugim $\frac{1}{2} \cdot \frac{3}{n} = \frac{3}{2n}$, a więc lepiej. Ale możliwych jest wiele innych strategii...

Otóż nasz hobbit postanowił obliczyć, która odpowiedź da mu więcej informacji. Położenie pierścienia opisuje zmienna losowa R o wartościach $1, 2, \ldots, n$, odpowiedź pierwszego trolla zmienna T_1 o wartościach $1, 2$ (która połówka), a drugiego trolla zmienna T_2 o wartościach $1, 2, 3$. Nie wiedząc nic więcej, hobbit założył, że rozkład zmiennej R jest jednostajny, a pierwszy troll decyduje się na kłamstwo z prawdopodobieństwem $\frac{1}{3}$ niezależnie od wartości R, czyli

$$\Pr\left(T_1 = \left\lceil \frac{2r}{n} \right\rceil \;\middle|\; R = r\right) = \frac{2}{3},$$

ponieważ prawdziwą podpowiedzią jest $T_1 = \lceil \frac{2R}{n} \rceil$. Wynika stąd, że

$$\Pr(T_1 = 1) = \sum_{i=1}^{\frac{n}{2}} \tfrac{2}{3} \cdot \tfrac{1}{n} + \sum_{i=\frac{n}{2}+1}^{n} \tfrac{1}{3} \cdot \tfrac{1}{n} = \tfrac{1}{2},$$

czyli T_1 ma rozkład jednostajny. Poszukiwaną wartością jest

$$I(R; T_1) = H(T_1) - H(T_1 \mid R),$$

przy czym, jak łatwo sprawdzić,

$$H(T_1) = \tfrac{1}{2} \log_2 2 + \tfrac{1}{2} \log_2 2 = 1.$$

Żeby obliczyć $H(T_1 \mid R)$, przypomnijmy formułę (3). W naszym przypadku dla każdej wartości $R = r$ mamy

$$H(T_1 \mid r) = \tfrac{1}{3} \log_2 3 + \tfrac{2}{3} \log \tfrac{3}{2} = \log_2 3 - \tfrac{2}{3}$$

i w konsekwencji

$$I(R; T_1) = \tfrac{5}{3} - \log_2 3.$$

Przypadek drugiego trolla jest nieco bardziej skomplikowany. Jeśli nawet przyjmiemy, że troll decyduje się na kłamstwo z prawdopodobieństwem $\frac{1}{2}$ niezależnie od wartości $R = r$, to złe wartości są dwie i wybór jednej z nich w zależności od r może nawet coś hobbitowi podpowiadać. Nie wiedząc nic więcej, hobbit zakłada, że decydując się na kłamstwo troll rzuca (uczciwą!) monetą i w zależności od wyniku wskazuje jedną z dwóch części trzecich, w której pierścienia nie ma. Intuicja podpowiada, że jest to – z punktu widzenia hobbita – najgorszy przypadek; do tego szczegółu jeszcze wrócimy. Przy tym założeniu łatwo sprawdzić, że rozkład zmiennej T_2 jest również jednostajny i

$$H(T_2) = 3 \cdot \tfrac{1}{3} \cdot \log_2 3 = \log_2 3.$$

Natomiast dla każdej wartości $R = r$ wartości zmiennej T_2 przyjmowane są z prawdopodobieństwami warunkowymi, odpowiednio, $\frac{1}{2}$ (dobra), $\frac{1}{4}$ (zła), $\frac{1}{4}$ (zła). Dlatego, podobnie jak poprzednio, możemy obliczyć

$$H(T_2 \mid r) = \tfrac{1}{2} \log_2 2 + 2 \cdot \tfrac{1}{4} \log_2 4 = \tfrac{3}{2},$$

i taka jest też wartość $H(T_2 \mid R)$. W konsekwencji

$$I(R; T_2) = \log_2 3 - \tfrac{3}{2}.$$

Nierówność sprowadza się do $2 \log_2 3 - \frac{19}{6} > 0$, czyli $3^{12} > 2^{19}$.

Otóż okazuje się, że $I(R; T_2) > I(R; T_1)$, co możemy sprawdzić dość prostym rachunkiem, choć różnica jest mniejsza niż 0,01. Tak więc drugi troll ma – wprawdzie niewielką – przewagę. Wróćmy jeszcze do założenia o rzucie monetą. Czytelnik może sprawdzić, że przy każdym innym „sposobie kłamania” (ale utrzymując prawdopodobieństwo prawdy $\frac{1}{2}$) entropia warunkowa $H(T_2 \mid R)$ może się jedynie zmniejszyć; problem w tym, że zmniejszyć może się również $H(T_2)$, kiedy rozkład T_2 nie będzie jednostajny. Można jednak wykazać, że $\log_2 3 - \frac{3}{2}$ pozostaje ograniczeniem dolnym, czyli drugi troll jest nadal lepszy.

Daliśmy tu jedynie przedsmak teorii informacji, która powiązana jest z wieloma dziedzinami wiedzy: informatyką, fizyką, biologią, lingwistyką... Ale może się nam przydać, ilekroć jesteśmy gdzieś pomiędzy całkowitą pewnością a kompletną niewiedzą. Rozciąga się tu bowiem *continuum* możliwości, które warto brać pod uwagę – nie wyolbrzymiając, ale i nie pomniejszając ich znaczenia.