Metody probabilistyczne
Rozwiązania zadań

1. Prawdopodobieństwo klasyczne

03.10.2017

Zadanie 1. *Ile słów można utworzyć ze słowa BARBARA zmieniając kolejność liter?*

Odpowiedź: Wpierw rozważmy prostsze słowo BABA. Możliwych ułożeń liter jest tyle, ile czteroelementowych permutacji, czyli $4! = 24$. Niestety, z powodu powtarzających się liter, wiele z tych permutacji prowadzi do tych samych słów:

permutacja	słowo	permutacja	słowo	permutacja	słowo	permutacja	słowo
(1,2,3,4)	BABA	(2,3,1,4)	ABBA	(4,3,2,1)	ABAB	(3,2,4,1)	BAAB
(1,4,3,2)	BABA	(2,1,3,4)	ABBA	(2,3,4,1)	ABAB	(1,2,4,3)	BAAB
(3,4,1,2)	BABA	(3,1,2,4)	BBAA	(2,1,4,3)	ABAB	(4,2,1,3)	AABB
(3,2,1,4)	BABA	(3,1,4,2)	BBAA	(4,1,2,3)	ABAB	(2,4,1,3)	AABB
(4,1,3,2)	ABBA	(1,3,4,2)	BBAA	(1,4,2,3)	BAAB	(2,4,3,1)	AABB
(4,3,1,2)	ABBA	(1,3,2,4)	BBAA	(3,4,2,1)	BAAB	(4,2,3,1)	AABB

Weźmy przykładowo słowo ABBA. Przestawienie między sobą liter A, lub liter B, prowadzi do *tego samego słowa* ABBA. Litery A można między sobą przestawić na $2! = 2$ sposoby, podobnie litery B. Tym samym $2 \cdot 2 = 4$ permutacje (związane z przestawieniem między sobą *tych samych* liter) prowadzą do tego samego słowa.

Łącznie jest więc $\frac{4!}{4} = 6$ różnych słów, które można utworzyć ze słowa BABA.

Teraz rozważmy słowo BARBARA. Mamy 3 litery A, 2 litery B oraz 2 litery R, łącznie 7 liter. Podobnie jak poprzednio, przestawianie między sobą tych samych liter nie zmienia słowa. Litery A można przestawić na 3! sposobów, litery B oraz R – na 2! sposobów. Więc $3! \cdot 2! \cdot 2!$ przestawień prowadzi do tego samego słowa.

Łącznie mamy więc $\frac{7!}{3!2!2!} = 210$ różnych słów, które można utworzyć ze słowa BARBARA. Nie będziemy ich tutaj wypisywać.

Zadanie 2. *Jaka jest szansa trafienia „szóstki" w totolotka? (wybieramy 6 z 49 liczb, maszyna również losuje 6 z 49 liczb i musimy trafić wszystkie). Jaka jest szansa trafienia „piątki"? „czwórki"? „trójki"?*

Odpowiedź: Liczba wszystkich możliwych wyników to liczba 6-elementowych podzbiorów 49-elementowego zbioru, czyli $|\Omega| = \binom{49}{6}$. Tylko jeden z tych podzbiorów daje główną wygraną, stąd prawdopodobieństwo zdarzenia A_6 („szóstka w totolotka") wynosi:

$$P(A_6) = \frac{1}{\binom{49}{6}} = \frac{6!43!}{49!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{44 \cdot 45 \cdot 46 \cdot 47 \cdot 48 \cdot 49} = \frac{1}{13\,983\,816}.$$

Niewiele. Aby mieć istotne szanse na wygraną, trzeba by grać miliony razy (stąd kupowanie losów na totolotka zwane jest czasem „podatkiem od marzeń").

Trafienie „piątki". Dla przejrzystości załóżmy, że wylosowane liczby to $1, 2, 3, 4, 5, 6$. Musimy trafić dokładnie 5 z tych 6 liczb. Jest sześć takich „piątek", a każda z tych „piątek" będzie miała szóstą liczbę z zakresu $7 - 49$ (43 możliwości). Łącznie jest więc $6 \times 43 = 258$ 6-elementowych podzbiorów, które dają wygraną typu „piątka". Szukane prawdopodobieństwo wynosi więc:

$$P(A_5) = \frac{256}{\binom{49}{6}} \simeq \frac{1}{54\,201}.$$

Trafienie „czwórki”. Musimy trafić dokładnie 4 z 6 wylosowanych liczb, co można zrealizować na $\binom{6}{4}$ sposobów. Pozostałe dwie liczby muszą pochodzić *spoza* wylosowanej szóstki, co daje $\binom{49-6}{2}$ sposobów. Mamy więc:

$$P(A_4) \;=\; \frac{\binom{6}{4}\cdot\binom{43}{2}}{\binom{49}{6}} \;\simeq\; \frac{1}{1\,032}.$$

Trafienie „trójki”. Analogicznie do poprzedniego przypadku:

$$P(A_3) \;=\; \frac{\binom{6}{3}\cdot\binom{43}{3}}{\binom{49}{6}} \;\simeq\; \frac{1}{57}.$$

Zadanie 3. *Paradoks urodzin: Jaka jest szansa, że w grupie 23 osób są przynajmniej dwie osoby mające urodziny tego samego dnia? (dla uproszczenia załóż, że nikt nie urodził się 29 lutego!)*

Odpowiedź: Ponieważ wykluczamy datę 29 lutego, jest 365 możliwości dnia urodzin. Mamy grupę 23 osób, więc jest $|\Omega| = 365^{23}$ możliwych wyników w tym doświadczeniu. Ile z nich prowadzi do „kolizji”, tj. do sytuacji, w której przynajmniej dwie osoby mają urodziny tego samego dnia? Oznaczmy to zdarzenie jako A. Łatwiej będzie nam wyznaczyć prawdopodobieństwo zdarzenia A' („nie ma osób o tej samej dacie urodzenia”). Ile zdarzeń elementarnych (wyników) prowadzi do zdarzenia A'? Pierwsza z osób może mieć urodziny dowolnego dnia (365 możliwości), druga z osób może mieć urodziny dowolnego dnia poza dniem urodzin 1. osoby (364 możliwości), trzecia z osób może mieć urodziny dowolnego dnia poza dniem urodzin 1. lub 2. osoby (363 możliwości), itd. Mamy więc:

$$|A'| = 365 \cdot 364 \cdot \ldots \cdot (365-22),$$

a stąd:

$$P(A') \;=\; \frac{365\cdot 364\cdot \ldots 343}{365^{23}} \;\simeq\; 0.4927. \tag{1}$$

Otrzymujemy więc:

$$P(A) \;=\; 1 - P(A') \;=\; 0.5073.$$

Prawdopodobieństwo wynosi ponad 50% i jest zaskakująco duże (stąd „paradoks”).

Dodatek: Ograniczenie na prawdopodobieństwo kolizji w paradoksie urodzin. Rozważmy teraz n dni i m osób, a sytuację w której co najmniej 2 osoby mają urodziny tego samego dnia nazwijmy *kolizją*. Dla uproszczenia obliczeń założymy, że $m \leq n/2$, tzn liczba osób nie przekracza połowy możliwych dni (w zadaniu powyżej było to oczywiście spełnione. Do dalszej analizy potrzebujemy dwóch ograniczeń związanych z logarytmami:

Lemat 1. *Dla dowolnego $x > -1$, $\ln(1+x) \leq x$.*

Dowód. Rozważmy funkcję $f(x) = \ln(1+x) - x$. Aby udowodnić lemat, musimy po prostu pokazać, że $f(x) \leq 0$ (dla $x > -1$). Poszukajmy ekstremów tej funkcji. Obliczamy pochodną:

$$f'(x) = \frac{1}{1+x} - 1,$$

a następnie przyrównujemy ją do zera:

$$f'(x) = 0 \quad\iff\quad \frac{1}{1+x} - 1 = 0 \quad\iff\quad x = 0.$$

Mamy więc ekstremum w $x = 0$. Aby sprawdzić czy to minimum, czy może maksimum, liczymy drugą pochodną:

$$f''(x) = -\frac{1}{(1+x)^2}.$$

Ponieważ $f''(0) = -1 < 0$, mamy maksimum w 0. Wynosi ono $f(0) = 0$. Skoro jest to maksimum, to $f(x) \leq 0$ w całej swojej dziedzinie (czyli dla $x > -1$). $\square$

Lemat 2. *Dla dowolnego* $0 \leq x \leq \frac{1}{2}$ *mamy* $\ln(1-x) \geq -2x$

Dowód. Ponieważ $\ln(a) = -\ln\left(\frac{1}{a}\right)$:

$$\ln(1-x) = -\ln\left(\frac{1}{1-x}\right) = -\ln\left(\frac{1-x\ +x}{1-x}\right) = -\ln\left(1+\frac{x}{1-x}\right).$$

Teraz wykorzystujemy lemat 1, jako x przyjmując $\frac{x}{1+x}$, czyli:

$$\ln\left(1+\frac{x}{1+x}\right) \leq \frac{x}{1+x}.$$

Wstawiając to do poprzedniego równania dostajemy:

$$\ln(1-x) = -\ln\left(1+\frac{x}{1-x}\right) \geq -\frac{x}{1-x},$$

Teraz wystarczy tylko pokazać, że $\frac{x}{1-x} \leq 2x$ dla $0 \leq x \leq \frac{1}{2}$, aby zakończyć dowód. Dla $x = 0$ ta nierówność jest oczywista, załóżmy więc, że $x > 0$. Ponieważ $1 - x \geq 1 - \frac{1}{2} \geq \frac{1}{2}$, mamy więc $\frac{1}{1-x} \leq 2$. Przemnażając obustronnie przez x, dostajemy $\frac{x}{1-x} \leq 2x$, co kończy dowód. □

Kontynuujmy obliczanie prawdopodobieństw kolizji. Uogólniając wzór (1) na przypadek dowolnego n i m otrzymujemy prawdopodobieństwo zdarzenie A' („brak kolizji") w postaci:

$$P(A') \;=\; \frac{n \cdot (n-1) \cdot \ldots \cdot (n-m+1)}{n^m} \overset{(*)}{=} \prod_{i=1}^{m-1} \frac{n-i}{n} \;=\; \prod_{i=1}^{m-1}\left(1-\frac{i}{n}\right),$$

gdzie równość $(*)$ bierze się stąd, że pierwszy czynnik (odpowiadający $i = 0$) znika, ponieważ $\frac{n}{n} = 1$. Logarytmując obustronnie (logarytmem naturalnym), iloczyn zamienia się na sumę, przez co dostajemy:

$$\ln P(A') \;=\; \sum_{i=1}^{m-1} \ln\left(1-\frac{i}{n}\right). \tag{2}$$

Teraz wykorzystamy lemat 1 z $x = -\frac{i}{n}$ w poszczególnych składnikach sumy po prawej stronie równania (2), otrzymując:

$$\ln P(A') \;\leq\; \sum_{i=1}^{m-1}\left(-\frac{i}{n}\right) \;=\; -\frac{1}{n}\sum_{i=1}^{m-1} i \;=\; \frac{1}{n}\frac{m(m-1)}{2},$$

gdzie w ostatniej równości po prostu policzyliśmy sumę ciągu arytmetycznego. Z drugiej strony, stosując do poszczególnych składników sumy w (2) lemat 2 z $x = \frac{i}{n}$, dostajemy:

$$\ln P(A') \;\geq\; 2\sum_{i=1}^{m-1}\left(-\frac{i}{n}\right) \;\geq\; \frac{2}{n}\frac{m(m-1)}{2},$$

stosując te same obliczenia, co poprzednio. Zauważmy, że założenie lematu 2 dotyczące x jest tu spełnione, ponieważ $0\frac{i}{n} \leq \frac{m}{n} \leq \frac{1}{2}$, zgodnie z naszym założeniem na n i m. Otrzymaliśmy więc:

$$-\frac{m(m-1)}{n} \;\leq\; \ln P(A') \;\leq\; -\frac{m(m-1)}{2n}.$$

Z definicji logarytmu naturalnego, jeśli $\ln a \leq b$ to $a \leq e^b$. Tym samym:

$$\exp\left(-\frac{m(m-1)}{n}\right) \;\leq\; P(A') \;\leq\; \exp\left(-\frac{m(m-1)}{2n}\right).$$

Co to oznacza? Aby prawdopodobieństwo kolizji $P(A')$ było znaczące, powiedzmy rzędu $\frac{1}{2}$, wtedy $\frac{m(m-1)}{n}$ musi być rzędu jedności:

$$\frac{m(m-1)}{n} = O(1) \qquad \Longleftrightarrow \qquad m = O(\sqrt{n}).$$

Oznacza to, że liczba osób rzędu pierwiastka liczby dni daje znaczące prawdopodobieństwo kolizji. Te obliczenia mają bezpośrednie zastosowanie do wyznaczania prawdopodobieństwa kolizji funkcji haszujących.