# Prawdopodobieństwo, informacja i zasady nieoznaczoności

*Iwo BIAŁYNICKI-BIRULA*

*Informacja* to jedno z kluczowych pojęć obecnej epoki. Celem tego artykułu jest przedstawienie *zasady nieoznaczoności* – jednej z podstawowych zasad teorii kwantowej – w takiej formie, w której pojęcie informacji będzie odgrywało podstawową rolę. Tradycyjna postać tej zasady, pochodząca od jej odkrywcy Wernera Heisenberga, dana jest wzorem

$$\Delta x \Delta p \geq h. \tag{1}$$

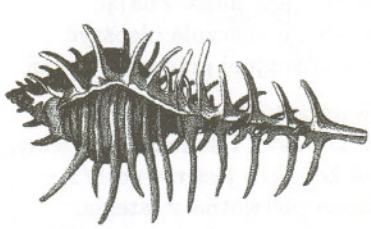
Iloczyn niepewności położenia i niepewności pędu nie może być mniejszy od stałej Plancka. Heisenberg wyraził tę zasadę w następujący sposób: „Im dokładniej określone jest położenie, tym gorzej znany jest pęd i na odwrót". Wzór przedstawiający zasadę nieoznaczoności stał się znakiem firmowym teorii kwantowej, podobnie jak równanie Einsteina $E = mc^2$ stało się znakiem firmowym teorii względności. W tym artykule pokażę, że bardziej trafne sformułowanie tej zasady można uzyskać, przedstawiając niepewność przy użyciu pojęcia informacji. Przy tej okazji będziemy mogli lepiej poznać istotę informacji i nauczymy się także precyzyjnie mierzyć jej ilość.

Rozpoczniemy od spostrzeżenia, iż niepewność i informacja to dwie strony tego samego medalu. Uzyskując informację, likwidujemy niepewność i na odwrót – tracąc informację, powiększamy niepewność. Zdobywanie informacji można traktować jako proces podobny do napełniania wodą zbiornika. Napływająca woda (informacja) wypełnia pustą przestrzeń zbiornika (niepewność). Objętość zbiornika można mierzyć ilością wody, którą można w nim zmieścić. Możemy zatem mierzyć niepewność i informację tą samą miarą – posługując się tą samą jednostką. Jednostką informacji jest *bit*. Oczywiście jest to jednostka bardzo mała i dlatego posługujemy się zazwyczaj jednostką 8 razy większą – *bajtem*, która występuje na ogół w dużych porcjach w postaci kilobajtów, megabajtów, gigabajtów i terabajtów. Skąd wzięły się te jednostki i jak wiążą się one z pomiarem ilości informacji? Za datę narodzin współczesnej teorii informacji przyjmujemy rok 1948, w którym amerykański matematyk i inżynier Claude Shannon sformułował matematyczną teorię łączności. W teorii tej podstawową rolę odgrywa sławny wzór Shannona

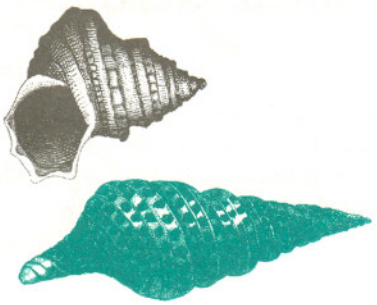
$$H = -\sum p_i \log_2 p_i, \tag{2}$$

w którym liczby $p_i$ oznaczają prawdopodobieństwa wystąpienia zdarzeń, o których więcej powiem w dalszym ciągu. We wzorze Shannona $H$ oznacza informację wyrażoną w bitach. Ponieważ prawdopodobieństwo jest zawsze liczbą mniejszą od jedności, informacja jest wielkością nieujemną, logarytm dwójkowy liczby mniejszej od 1 jest ujemny.

W celu uzasadnienia swojego wzoru, Shannon posłużył się związkiem między niepewnością i informacją. Wyjaśnimy poniżej znaczenie wzoru Shannona, posługując się znaną wszystkim grą w 20 pytań. Najpierw jednak omówimy kilka podstawowych własności informacji i sposobu jej zapisu. Na to, by informacja uzyskała konkretną treść, założymy, że jest ona zakodowana w ciągu znaków o ustalonej długości. W ogólnym przypadku nie musi to być tekst słowny, może to być przekaz muzyczny albo graficzny. Będziemy zakładać, że ilość informacji zawarta w takim ciągu znaków jest proporcjonalna do jego długości; dwa razy grubsza książka zawiera dwa razy więcej informacji. Posługując się terminologią fizyczną, można powiedzieć, że informacja jest wielkością ekstensywną, proporcjonalną do objętości nośnika, tak jak energia, masa, czy entropia. W dalszych rozważaniach przyjmiemy najprostszy możliwy sposób zapisu, korzystający z alfabetu składającego się jedynie z dwóch znaków: 0 i 1. Jest to zapis bardzo naturalny dla komputera, który w swej istocie rozumie tylko taki alfabet binarny. W języku komputera słowem jest ciąg zer i jedynek o długości $N$. Liczba $N$ mierzy ową objętość nośnika. Informacja zawarta w słowie jest zatem proporcjonalna do $N$. Umówimy się, że współczynnik proporcjonalności w tym związku jest równy jedności, to znaczy, że



$$Informacja\_H = Długość\_Słowa. \tag{3}$$



Łatwo obliczyć, że istnieje $2^N$ różnych słów binarnych o długości $N$. Między długością słowa i liczbą możliwych słów zachodzi więc związek

(4) $$Długość\_Słowa = \log_2(Liczba\_Słów).$$

Umowa nasza oznacza zatem, że, kładąc we wzorze (3) współczynnik równy jedności, przyjęliśmy jako jednostkę pomiaru informacji jeden bit; liczba liter w binarnym słowie równa się liczbie bitów. Jeżeli wszystkie słowa są równie prawdopodobne, to prawdopodobieństwo $p$ wystąpienia danego słowa wynosi

(5) $$p = 1/Liczba\_Słów.$$

Otrzymujemy zatem wynik, iż informacja zawarta w słowie, dla którego prawdopodobieństwo wystąpienia wynosi $p$, jest równa

(6) $$H(p) = -\log_2 p.$$

Zilustruję teraz ten wynik na przykładzie gry w 20 pytań. Wykażę, że określona w powyższy sposób miara informacji jest po prostu równa liczbie pytań, które są potrzebne do odgadnięcia słowa. Rozważania te ograniczę na razie tylko do tej uproszczonej sytuacji, gdy wszystkie słowa są równoprawdopodobne. Dla ułatwienia obliczeń ponumeruję wszystkie słowa o długości $N$ kolejnymi liczbami naturalnymi od 1 do $2^N$ i zastosuję następujący sposób opisu procesu zgadywania. Mamy przed sobą $2^N$ zakrytych komórek. W jednej z nich znajduje się SKARB. Miejsce ukrycia skarbu jest informacją, którą chcemy posiąść. Znalezienie tego skarbu jest oczywiście tym samym, z matematycznego punktu widzenia, co odgadnięcie słowa. Dla zlokalizowania skarbu jako pierwsze zadajemy następujące pytanie:

*Czy skarb znajduje się w lewej połowie komórek?*

Jeżeli odpowiedź na pierwsze pytanie brzmi TAK, to drugie pytanie będzie brzmiało:

*Czy skarb znajduje się w pierwszej ćwiartce komórek?*

Jeżeli odpowiedź na pierwsze pytanie brzmi NIE, to drugie pytanie będzie brzmiało:

*Czy skarb znajduje się w trzeciej ćwiartce komórek?*


**Rys. 1**

Kontynuując taką strategię, odgadniemy z pewnością po dokładnie $N$ pytaniach miejsce ukrycia skarbu. Niepewność co do położenia skarbu, istniejąca na początku, zostaje całkowicie zlikwidowana przy użyciu $N$ pytań. Odpowiedź na każde pytanie daje nam 1 bit informacji, zmniejszając także niepewność o 1 bit. Liczba pytań potrzebna do uzyskania pełnej informacji jest równa niepewności, z jaką przystępujemy do gry. Jest ona jednocześnie równa ilości informacji uzyskanej po odgadnięciu miejsca ukrycia,

$$Informacja\_H = Liczba\_Pytań.$$

Proces zgadywania został przedstawiony na rysunku 1 w postaci „drzewa pytań" w prostym przypadku, gdy $N = 2$.

Rzeczywistość jest na ogół jednak bardziej złożona i rzadko do uzyskania informacji możemy skorzystać z tak banalnej strategii. Problem doboru odpowiedniej strategii pojawia się dlatego, że na ogół prawdopodobieństwa występowania różnych konfiguracji nie są jednakowe. Wytłumaczę to na przykładzie zwykłego języka. Gdy, grając w szubienicę, mamy do odgadnięcia słowo zawierające zestaw liter KU_A, to nie wiemy, czy ma to być KULA, KUMA, KUNA, KUPA czy KURA. Jeżeli, natomiast, spotkamy się z zestawem ŚW_T, to wiemy, że szukane słowo to ŚWIT. Dzieje się tak dlatego, iż kombinacja liter KU_A występuje często, zestaw zaś ŚW_T występuje rzadko, prawdopodobieństwo jego wystąpienia jest małe i informacja niesiona przez ten zestaw liter, zgodnie ze wzorem (6), jest duża. Znając prawdopodobieństwa wystąpienia różnych słów, możemy znacznie ulepszyć strategię ich odgadywania. Tak właśnie robimy w tradycyjnej grze w 20 pytań.

Zilustruję tę nową strategię znowu na przykładzie poszukiwania skarbu. Przypuśćmy, że jest on ukryty w jednej z czterech komórek i dodatkowo wiemy, że prawdopodobieństwo znalezienia go w pierwszej komórce jest równe 1/2, w drugiej 1/4, w trzeciej i czwartej zaś po 1/8. Nasza pierwotna strategia, polegająca na przepoławianiu zbioru komórek, pozwoli na znalezienie skarbu

**Rozwiązanie zadania M 922.**
Załóżmy, że istnieje punkt $S$ wewnątrz czworokąta $A_1A_2A_3A_4$, który nie należy do żadnego koła. Wtedy jednak każdy z kątów $A_iSA_{i+1}$ byłby ostry i suma tych kątów byłaby mniejsza niż 360°. Sprzeczność.

zawsze po dwóch pytaniach. Nie jest to jednak strategia optymalna, gdyż nie wykorzystujemy w niej informacji o prawdopodobieństwach. Skuteczniejsza jest strategia oparta na następujących pytaniach:

*Czy skarb znajduje się w pierwszej komórce?*

Jeżeli odpowiedź na pierwsze pytanie brzmi TAK, to znamy jego miejsce. Jeżeli odpowiedź na pierwsze pytanie brzmi NIE, to drugie pytanie będzie brzmiało:

*Czy skarb znajduje się w drugiej komórce?*

Jeżeli odpowiedź na drugie pytanie brzmi TAK, to znowu znamy jego miejsce. Jeżeli odpowiedź na drugie pytanie brzmi NIE, to trzecie pytanie będzie brzmiało:

*Czy skarb znajduje się w trzeciej komórce?*

Rys. 2

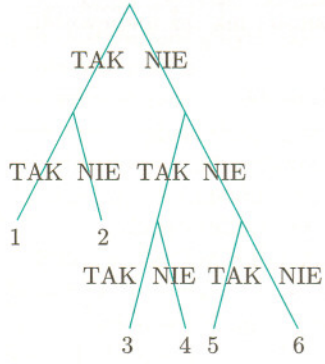
Po trzecim pytaniu znamy już na pewno położenie skarbu. Drzewo pytań w tym przypadku przedstawione jest na rysunku 2. Na pierwszy rzut oka może się wydawać, że nowa strategia jest gorsza, bo wymaga trzech, zamiast standardowych dwóch pytań dla czterech komórek. Jeżeli jednak skarb jest w pierwszej komórce, to znajdujemy go już po pierwszym pytaniu. Ten właśnie zysk przeważa nad stratą. Można to potwierdzić, obliczając średnią liczbę pytań, według znanego przepisu

$$\begin{aligned} \mathit{Średnia\_Wartość} = \mathit{Prawdopodobieństwo\_1} * \mathit{Wartość\_1} + \\ + \mathit{Prawdopodobieństwo\_2} * \mathit{Wartość\_2} + \ldots \end{aligned}$$

Średnia liczba pytań, obliczona według tego wzoru, dla naszego zadania wynosi

$$\mathit{Średnia\_Liczba\_Pytań} = \frac{1}{2} * 1 + \frac{1}{4} * 2 + \frac{1}{8} * 3 + \frac{1}{8} * 3 = \frac{7}{4} < 2.$$

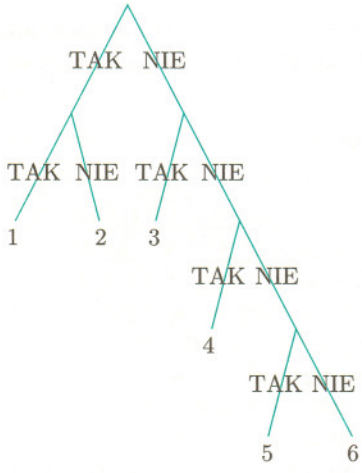
Uzyskaliśmy, dzięki nowej strategii, wynik lepszy od poprzedniego *średnio* o 1/4 pytania. Można prosto wyjaśnić, na czym polegała nasza nowa strategia. Pytania dobieraliśmy w ten sposób, żeby prawdopodobieństwa uzyskania odpowiedzi twierdzącej i przeczącej były takie same. W podanym przykładzie udało się nam to osiągnąć. Przed sformułowaniem ogólnych wniosków rozważymy jeszcze jeden przykład. Tym razem skarb ukryty jest w jednej z sześciu komórek z następującymi prawdopodobieństwami:



Rys. 3

$$p_1 = \frac{1}{3}, \quad p_2 = \frac{1}{5}, \quad p_3 = \frac{1}{5}, \quad p_4 = \frac{2}{15}, \quad p_5 = \frac{1}{15}, \quad p_6 = \frac{1}{15}.$$

Tutaj wybór optymalnej strategii nie jest już oczywisty. W ogólnym przypadku możemy jedynie dążyć do tego, by równy podział prawdopodobieństw uzyskać w jak najlepszym przybliżeniu. Jedna z rozsądnych strategii zdefiniowana jest przez drzewo pytań przedstawione na rysunku 3. Średnia liczba pytań wynosi w tym przypadku 37/15. Okazuje się, że nie jest to jeszcze strategia optymalna. Lepszy wynik dostajemy, zadając pytania według przepisu przedstawionego na rysunku 4. Mimo iż wydłużyliśmy listę pytań w niektórych przypadkach aż do czterech, średnia liczba pytań wynosi tylko $36/15 = 2{,}4$.

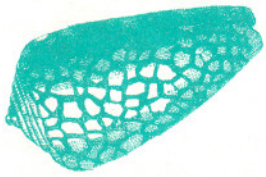


Rys. 4

Ponieważ liczba pytań określała nam ilość informacji, to *średnią* liczbę pytań można utożsamić ze *średnią* informacją. W ten sposób doszliśmy do uzasadnienia wzoru Shannona. Występująca w tym wzorze suma jest po prostu średnią informacją, obliczoną w ogólnym przypadku, gdy znane są wszystkie prawdopodobieństwa znalezienia skarbu w $i$-tej komórce. Można dowieść, że niepewność jest największa, gdy wszystkie prawdobodobieństwa są równe i wynosi wtedy $\log_2 N$. Natomiast w przypadku, gdy wiemy, w której komórce znajduje się skarb, niepewność powinna być równa zeru. I rzeczywiście, jeżeli wszystkie prawdopodobieństwa $p_i$, poza jednym z nich, są równe zeru, to $H$ jest równe zeru, bo logarytm liczby 1 jest równy zeru.

Wzór Shannona jest ogromnie ważny w teorii informacji, ponieważ związane jest z nim fundamentalne twierdzenie matematyczne, zwane twierdzeniem o bezszumowym kodowaniu. W naszej interpretacji, polegającej na zadawaniu pytań, głosi ono, że

NIE ISTNIEJE STRATEGIA, KTÓRA ŚREDNIO DAJE MNIEJSZĄ LICZBĘ PYTAŃ, NIŻ OKREŚLA TO WZÓR SHANNONA.



Z twierdzenia tego wynika, że nie można już dalej poprawić strategii zgadywania w naszym drugim przykładzie i uzyskać wyniku lepszego niż 36/15. Obliczona ze wzoru Shannona wartość informacji wynosi bowiem w tym przypadku:

$$H = \frac{\log_2(3)}{3} + 2 * \frac{\log_2(5)}{5} + \frac{2\log_2(15/2)}{15} + 2 * \frac{\log_2(15)}{15} = 2{,}3656.$$

Gdyby istniała strategia, która daje średnią liczbę pytań 35/15 (z ogólnego wzoru na średnią widać, że w tym przypadku średnie są zawsze wielokrotnościami 1/15), to byłoby to sprzeczne z twierdzeniem o bezszumowym kodowaniu, gdyż $35/15 = 2{,}3333 < 2{,}3656$.

Pora teraz na zastosowanie wprowadzonych pojęć do fizyki kwantowej. Dokonując pomiaru nad obiektem fizycznym, postępujemy podobnie, jak przy poszukiwaniu skarbu: dążymy do uzyskania informacji, czyli do usunięcia niepewności. Dla uproszczenia, rozważmy cząstkę poruszającą się tylko w jednym wymiarze, wzdłuż osi $x$. Podzielmy tę oś na jednakowe ponumerowane komórki o długości $\mathrm{d}x$. Rozmiar komórki możemy uważać za miarę dokładności pomiaru. Im mniejsza komórka, tym dokładniejszy jest pomiar. Oprócz pomiarów położenia będziemy także dokonywali pomiarów pędu cząstki $p_x$ w kierunku osi $x$. Oś pędu też podzielimy na jednakowe, ponumerowane komórki o długości $\mathrm{d}p$, charakteryzujące dokładność pomiaru pędu. Teoria kwantów pozwala na obliczenie wszystkich prawdopodobieństw znalezienia cząstki w $i$-tej komórce na osi $x$ i wykrycie jej pędu w $j$-tej komórce na osi $p_x$. Na podstawie tych prawdopodobieństw możemy obliczyć niepewność zgodnie ze wzorem Shannona. Otrzymujemy dwa takie wzory: jeden na niepewność położenia $H_x$, a drugi na niepewność pędu $H_p$. Nie ma żadnego ograniczenia na wartości tych dwóch wielkości rozważanych z osobna. Każda z tych dwóch miar niepewności może być równa dowolnej rzeczywistej liczbie nieujemnej. Komórek na osi rzeczywistej jest nieskończenie wiele i niepewność może być dowolnie duża. Jeżeli jednak wiemy dokładnie, w której komórce znajduje się cząstka, to niepewność jej położenia redukuje się do zera. Takie same własności ma też niepewność pędu,

$$0 \le H_x < \infty, \qquad 0 \le H_p < \infty.$$

Ograniczenia na niepewności położenia i pędu, wyrażające zasadę nieoznaczoności, pojawiają się przy jednoczesnym rozważaniu niepewności położenia i pędu. Mam tu na myśli taką sytuację, gdy dokonujemy pomiaru zarówno położenia i pędu na cząstce, która jest w tym samym stanie kwantowym, to znaczy została za każdym razem tak samo przygotowana do pomiaru. Może to, na przykład, oznaczać, że użyte do pomiaru cząstki zostały wytworzone w akceleratorze pracującym w trybie ciągłym. Kwantowa zasada nieoznaczoności, wyrażona przez wielkości $H_x$ i $H_p$, głosi, że suma dla każdego stanu kwantowego musi być większa niż pewna stała $C$, będąca funkcją iloczynu $\mathrm{d}x\mathrm{d}p$ wielkości komórek $\mathrm{d}x$ i $\mathrm{d}p$ charakteryzujących dokładność pomiaru,

$$H_x + H_p > C.$$

Niestety, nie znamy dokładnej zależności tej stałej od $\mathrm{d}x\mathrm{d}p$. Przed piętnastoma laty udało mi się jedynie wyprowadzić wzór na $C$, słuszny dla małych wartości $\mathrm{d}x\mathrm{d}p$, z którego wynika nierówność

$$H_x + H_p > \log_2(\mathrm{e}h/2\mathrm{d}x\mathrm{d}p).$$

We wzorze tym e oznacza podstawę logarytmów naturalnych, $h$ zaś jest stałą Plancka. Z nierówności tej widać, że gdy dokładność pomiaru $x$ i $p$ rośnie, czyli $\mathrm{d}x$ i $\mathrm{d}p$ maleją, to prawa strona nierówności staje się coraz większa, a zatem suma niepewności położenia i pędu rośnie. Jeżeli cząstka znajdzie się w jednej z komórek na osi $x$, to co prawda $H_x$ znika, ale nieoznaczoność w pędzie jest co najmniej równa stałej $C$. Nie można więc jednocześnie zlokalizować cząstki kwantowej w przestrzeni i zredukować niepewności pędu do zera. Nie można nigdy uzyskać jednocześnie pełnej informacji o położeniu cząstki i o jej pędzie. Posługując się pojęciem informacji, można więc nadać nową formę starej zasadzie nieoznaczoności odkrytej przed przeszło siedemdziesięcioma laty przez Wernera Heisenberga.

**Rozwiązanie zadania M 923.**
Niech $s$ będzie okręgiem o największym promieniu, $S$ – jego środkiem, $r$ – jego promieniem, $A_1A_2A_3$ – kolejnymi wierzchołkami leżącymi na $s$. Załóżmy, że wielokąt nie zawiera się w $s$. Wtedy istnieje wierzchołek $A_k$ leżący na zewnątrz $s$. Możemy bez straty ogólności założyć, że $A_k$ leży po tej samej stronie prostej $A_2S$ co $A_3$ i $k$ jest najmniejsze z możliwych. Łatwo zauważyć, że promień $R$ okręgu $s_1$ opisanego na trójkącie $A_2A_3A_k$ jest większy niż $r$, a wierzchołki $A_2, A_3, \ldots, A_k$ leżą w $s_1$. Jeśli $A_3A_k$ są kolejnymi wierzchołkami, to otrzymujemy sprzeczność. Załóżmy więc, że nie są kolejne. Rozważmy wszystkie okręgi opisane na trójkątach $A_3A_iA_k$, $3 < i < k$. Ich promienie są nie mniejsze niż $R$. Niech ten z nich ($s_2$), który ma najmniejszy promień, przechodzi przez $A_l$ $(3 < l < k)$. Jasne jest, że wszystkie wierzchołki $A_3, A_4, \ldots, A_k$ leżą w $s_2$. Jeśli $A_3A_lA_k$ są kolejnymi wierzchołkami, to otrzymaliśmy sprzeczność. Jeśli nie, to powtarzamy powyższą procedurę w odniesieniu do wierzchołków $A_3, \ldots, A_l$ itd. Po skończenie wielu krokach otrzymamy trzy kolejne wierzchołki, dla których okrąg przechodzący przez nie ma promień większy niż $r$. Sprzeczność.