Prawdopodobieństwo warunkowe

Jacek JAKUBOWSKI, Rafał SZTENCEL

W tym artykule pokażemy trzy przykłady, z których widać, że prawdopodobieństwo warunkowe ma, mimo prostej definicji, zaskakujące własności. Zaczniemy od przykładu, który można spotkać w prawie każdym podręczniku, następnie podamy przykład związany z grami losowymi, a zakończymy przykładem z medycyny.

Przykład 1

Losujemy jedną rodzinę spośród rodzin z dwojgiem dzieci. Obliczyć prawdopodobieństwo tego, że wybierzemy rodzinę z dwoma chłopcami, jeśli wiemy, że w tej rodzinie:
a) starsze dziecko jest chłopcem,
b) jest co najmniej jeden chłopiec.

Rozwiązanie. Przyjmiemy, że zbiór zdarzeń elementarnych Ω składa się z czterech jednakowo prawdopodobnych par:

$$\{(c,c),(c,d),(d,c),(d,d)\},$$

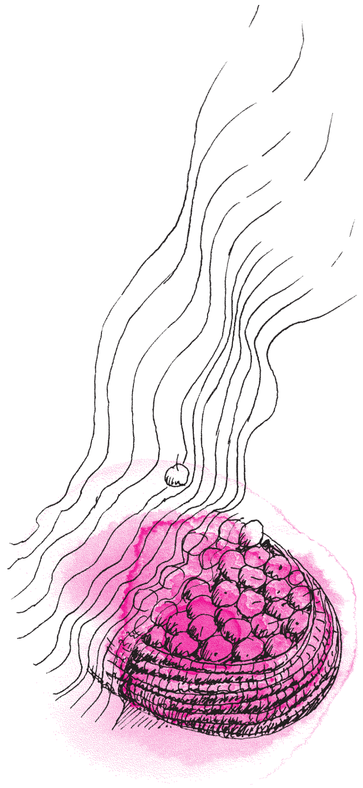
gdzie pierwszy element pary oznacza młodsze dziecko, a drugi – starsze.

W punkcie a) spodziewamy się odpowiedzi 1/2 i rzeczywiście

$$P(\{(c,c)\}|\{(c,c),(d,c)\}) = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

Natomiast odpowiedź na pytanie b) może być niespodzianką:

$$P(\{(c,c)\}|\{(c,c),(d,c),(c,d)\}) = \frac{1}{3}.$$



Drugi przykład dotyczy oceny szans w grze wywołującej wiele emocji zarówno wśród uczestników, jak i telewidzów w USA w latach 90. ubiegłego wieku. Jeszcze więcej emocji wzbudziło rozwiązanie.

Przykład 2

Po dojściu do ostatniego etapu teleturnieju uczestnik ma wybrać jedne drzwi spośród trojga. Wie, że za jednymi jest samochód, a za pozostałymi żywe gęsi. Dokonuje wyboru. Dla ustalenia uwagi przyjmijmy, że grający wybrał drzwi nr 1. Wtedy prowadzący (który wie, gdzie co jest) otwiera jedne z pozostałych, a mianowicie te, za którymi jest gęś. Niech to będą drzwi nr 2. Teraz uczestnik może pozostać przy swoim wyborze lub zmienić go, czyli wybrać drzwi nr 3. Co powinien zrobić?

Rozwiązanie. Wynikiem tego doświadczenia jest sposób rozmieszczenia obiektów i, co nie mniej ważne, podpowiedź prowadzącego. Zatem

$$\Omega = \{(AGG,2),(AGG,3),(GAG,3),(GGA,2)\},$$

przy czym pierwsza trójka wskazuje, gdzie jest auto (A), a gdzie gęś (G), a liczba mówi, które drzwi wskazał prowadzący. W sytuacji (AGG) prowadzący może wybrać drzwi, które otwiera. Jego strategię opisuje prawdopodobieństwo otwarcia drzwi nr 2, czyli liczba $p \in [0,1]$.

Wtedy

$$P(\{(AGG,2)\}) = p/3 = 1 - P(\{(AGG,3)\}).$$

W pozostałych przypadkach prowadzący nie ma wyboru, tzn.

$$P(\{(GAG,3)\}) = 1/3 = P(\{(GGA,2)\}).$$

Wobec tego, gdy zdarzenie W oznacza wygraną przy zmianie decyzji, D_2 – wskazanie przez prowadzącego drzwi nr 2, to

$$P(W|D_2) = \frac{P(W\cap D_2)}{P(D_2)} = \frac{P(\{(GGA,2)\})}{P(\{(GGA,2),(AGG,2)\})} = \frac{1/3}{p/3+1/3} = \frac{1}{1+p}.$$

Ponieważ $\frac{1}{1+p} \geqslant 1/2$ dla $p\in[0,1]$, więc zawsze lepiej zmienić wybór, bowiem szansa wygranej, gdy gracz pozostaje przy pierwotnym wyborze, jest równa 1/3.

Rozwiązanie zadania F 605.
Warunek równowagi daje:

$$\rho_1 g\frac{m}{\rho_1'} - mg = \rho_2 g\frac{m}{\rho_2'} - mg,$$

skąd otrzymujemy

$$\frac{\rho_1'}{\rho_2'} = \frac{\rho_1}{\rho_2}.$$

Trzeci przykład dotyczy testów diagnostycznych. Rozwiązanie wymaga zastosowania wzoru Bayesa, w którym z pozoru proste rachunki dają nieoczekiwane wyniki.

Przykład 3

Lekarze używają dwóch wskaźników jakości testu. Czułość testu jest zdefiniowana jako odsetek chorych, u których test daje wynik dodatni. Swoistość testu to odsetek zdrowych, u których test daje wynik ujemny.

i) Test na rzadką i groźną chorobę, którą dotknięta jest średnio jedna osoba na tysiąc, daje fałszywy wynik dodatni u 10% zdrowych (u chorego daje zawsze wynik dodatni). Jaka jest szansa, że osoba, u której test dał wynik dodatni, jest faktycznie chora? Zakładamy, że nic nie wiemy o innych możliwych objawach u badanej osoby.

Niech zdarzenie A oznacza wynik dodatni, H_1 – chorego, H_2 – zdrowego. Ze wzoru Bayesa obliczamy

$$P(H_1|A) = \frac{P(A|H_1)P(H_1)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2)} =$$

$$= \frac{1 \cdot \frac{1}{1000}}{1 \cdot \frac{1}{1000} + \frac{1}{10} \cdot \frac{999}{1000}} = \frac{10}{1009} \approx 0{,}0099.$$

Jest to około 1%.

Test omawiany w tym przykładzie ma czułość 100% i swoistość 90%. Gdy swoistość testu wzrośnie do 95%, to obliczane prawdopodobieństwo wzrasta do 2%. Mimo tak doskonałych wskaźników wynik dodatni w tym teście znaczy niewiele, i co najwyżej sugeruje konieczność dalszych badań.

ii) Czułość i swoistość testu są prawdopodobieństwami warunkowymi, co widać bezpośrednio z definicji. Zbadamy teraz jakość pewnego testu, powszechnie stosowanego w celu stwierdzenia, czy pacjent ma chorobę wieńcową. Jest to próba wysiłkowa, której czułość wynosi 65%, swoistość zaś 85% (dane za: Robert L. Bratton, Sprawdziany testowe z medycyny rodzinnej. Wiedza Medyczna, Centrum Medyczne Kształcenia Podyplomowego, Warszawa 2001).

Załóżmy, że jest 10% chorych na chorobę wieńcową. Obliczyć

a) prawdopodobieństwo tego, że próba wysiłkowa doprowadzi do prawidłowej diagnozy;

b) prawdopodobieństwo tego, że pacjent z wynikiem dodatnim jest chory;

c) prawdopodobieństwo tego, że pacjent z wynikiem ujemnym jest zdrowy.

Rozwiązanie. Oznaczmy zdarzenia następująco:
C – pacjent chory,
Z – pacjent zdrowy,
D – dodatni wynik testu,
U – ujemny wynik testu.

Z warunków zadania wynika, że
$P(C) = 0{,}1$,
$P(Z) = 0{,}9$,
$P(D|C) = 0{,}65$ (jest to czułość testu),
$P(U|Z) = 0{,}85$ (swoistość testu).

a) Prawdopodobieństwo prawidłowej diagnozy jest równe

$$P((D \cap C) \cup (U \cap Z)) = P(D \cap C) + P(U \cap Z) =$$

$$= P(D|C)P(C) + P(U|Z)P(Z) =$$

$$= 0{,}65 \cdot 0{,}1 + 0{,}85 \cdot 0{,}9 = 0{,}83.$$

Otrzymaliśmy średnią ważoną obu wskaźników; 83% jest całkiem przyzwoitą wiarygodnością.

b) Prawdopodobieństwo tego, że osoba z dodatnim wynikiem jest chora

$$P(C|D) = \frac{P(D|C)P(C)}{P(D|C)P(C) + P(D|Z)P(Z)} =$$

$$= \frac{P(D|C)P(C)}{P(D|C)P(C) + (1 - P(U|Z))P(Z)} =$$

$$= \frac{0{,}65 \cdot 0{,}1}{0{,}65 \cdot 0{,}1 + 0{,}15 \cdot 0{,}9} = \frac{65}{65 + 135} = 0{,}325.$$

Tak nieduże prawdopodobieństwo nie powinno jednak uspokajać osoby z wynikiem dodatnim. Nie może ona uważać się za losowo wybraną z całej populacji, bo jeśli już zgłosiła się do lekarza, to pewnie miała jakieś powody.

Zresztą w przypadku choroby wieńcowej decydujące znaczenie dla oceny szans chorego ma płeć, której w ogóle nie wzięliśmy pod uwagę!

c) Prawdopodobieństwo tego, że osoba z ujemnym wynikiem jest zdrowa, wynosi

$$P(Z|U) = \frac{P(U|Z)P(Z)}{P(U|Z)P(Z) + P(U|C)P(C)} =$$

$$= \frac{P(U|Z)P(Z)}{P(U|Z)P(Z) + (1 - P(D|C))P(C)} =$$

$$= \frac{0{,}85 \cdot 0{,}9}{0{,}85 \cdot 0{,}9 + 0{,}35 \cdot 0{,}1} = \frac{765}{765 + 35} \approx 0{,}956.$$

I tu rzeczywiście test się sprawdził.

Jeśli Czytelnik jest zaskoczony wynikami, niech narysuje zbiór zdarzeń elementarnych Ω w postaci prostokąta i niech zadba o to, by pola figur, reprezentujących zdarzenia, były zgodne z danymi. Wtedy okaże się, że wszystkie odpowiedzi są oczywiste. Cóż, jesteśmy wzrokowcami, a rachunkami zajmujemy się (jako gatunek) zaledwie od paru, może parunastu tysięcy lat.