Rachunek Prawdopodobieństwa MAP1181
Wydział Matematyki, Matematyka Stosowana

Paradoks Simpsona - przykłady

Opracowanie: Zofia Wróblewska

Paradoks Simpsona [1] pokazuje jak trudne jest wnioskowanie oparte na prawdopodobieństwie warunkowym. Dla zilustrowania tego paradoksu przeanalizujemy następujące przykłady...

1. **Umieralność na gruźlicę**
   Powiedzmy, że wylosowano ze spisu zmarłych w latach powojennych (np. 1945-1950) dane osoby z miasta M_1 lub miasta M_2. Jeżeli przyjmiemy następujące oznaczenia dla zdarzeń A, B i C: A=„osoba z miasta M_1 umarła na gruźlicę", B=„osoba z miasta M_2 umarła na gruźlicę", C=„osoba jest kobietą", to według Simpsona może zdarzyć się, że ... *W mieście M_1 jest większa umieralność chorych na gruźlicę niż w mieście M_2, a mimo to większa umieralność chorych kobiet, a także większa umieraność chorych mężczyzn, na gruźlicę, jest w mieście M_2.* [2] Powyższy efekt możemy zapisać za pomocą wzorów w następujący sposób:

$$P(A) > P(B)$$

i jednocześnie

$$P(A|C) < P(B|C) \text{ oraz } P(A|C^C) < P(B|C^C).$$

Zatem jeśli do pary cech (A i B) dołączy trzecia (C), to powstaje układ, który jest bardziej skomplikowany, niżby się to z pozoru wydawało.
Poniżej, w Tabelach 1 i 2, postaramy się tak dobrać dane dla obu miast, aby wystąpił paradoks Simpsona.

Tabela 1. Dane dla miasta M_1

	Osoba umarła na gruźlicę		Procentowe
	TAK	**NIE**	**TAK**
Kobieta	20	7000	0,28%
Mężczyzna	260	8000	3,15%
Łącznie	280	15000	1,83%

Tabela 2. Dane dla miasta M_2

	Osoba umarła na gruźlicę		Procentowe
	TAK	**NIE**	**TAK**
Kobieta	60	13000	0,46%
Mężczyzna	110	2000	5,21%
Łącznie	170	15000	1,12%

Rzeczywiście umieralność na gruźlicę w mieście M_1 ($P(A)$) jest o 0,71% większa niż w mieście M_2 ($P(B)$). Jeżeli uwzględnimy płeć jako zmienną grupującą to umieralność chorych kobiet w mieście M_2 ($P(B|C)$) jest o 0,18% większa niż w mieście M_1 ($P(A|C)$) oraz umieralność chorych mężczyzn w mieście M_2 ($P(B|C^C)$) jest o 2,06% większa niż w mieście M_1 ($P(A|C^C)$).
Paradoks ten ostrzega nas, że w rozważaniu relacji pradopodobieństw zdarzeń (tu $P(A)$ i $P(B)$) nie wystarczy udowodnić, że dana relacja zachodzi dla wszystkich przypadków pewnej zmiennej grupującej (tu taką zmienną jest płeć: C i C^C). Ostateczna konkluzja, jak widać, i tak może być inna.
Jak można wyjaśnić paradoks Simpsona? Zauważmy, że w pierwszym mieście umarło więcej mężczyzn, a wśród mężczyzn umieralność na gruźlicę była większa. To podwyższa umieralność na gruźlicę w mieście M_1. Ponieważ

[1] *Nie chodzi bynajmniej o Barta Simpsona, bohatera popularnej kreskówki, lecz o brytyjskiego statystyka o tymże nazwisku, który w 1951 roku opisał ten statystyczny efekt.*
[2] *Przykład częściowo pochodzi z książki: J.Jakubowski i R.Sztencel,* **Wstęp do teorii prawdopodobieństwa**, *SCRIPT, Warszawa 2001.*

w drugim mieście umarło mniej mężczyzn, więc możemy zadbać o to, aby procentowa umieralność na gruźlicę, wśród mężczyzn, była tu wyższa niż w pierwszym mieście, przy łącznej umieralności na gruźlicę niższej.

2. **Kara śmierci**

W Przykładzie 2 rozważymy wpływ rasy ofiary i rasy sprawcy na orzeczenie kary śmierci. W Tabeli 3 zebrano dane 674 morderstw popełnionych na Florydzie między 1976 a 1987. [3] Zamiast dwóch tabel 2x2, jak w Przykładzie 1, można rozważyć jedną tabelę wyższego wymiaru 2x2x2.

Tabela 3. Kara śmierci a rasa sprawcy i ofiary

Rasa **Ofiary**	Rasa **Sprawcy**	Kara śmierci **TAK**	**NIE**	Procentowe **TAK**
Biała	**Biała**	53	414	11.3%
	Czarna	11	37	22.9%
Czarna	**Biała**	0	16	0.0%
	Czarna	4	139	2.8%
Łącznie	**Biała**	53	430	11.0%
	Czarna	15	176	7.9%

Przyjmijmy oznaczenia. Niech A=„orzeczono karę śmierci", B=„sprawca jest Biały"oraz C=„ofiara jest Biała". Podobnie jak w poprzednio obliczamy odpowiednie prawdopodobieństwa:

$P(A|B) = \frac{53}{483} \approx 0,110$ oraz $P(A|B^C) = \frac{15}{191} \approx 0,079$ czyli $P(A|B) > P(A|B^C)$;
natomiast
$P(A|B \cap C) = \frac{53}{467} \approx 0,113$ oraz $P(A|B^C \cap C) = \frac{11}{48} \approx 0,229$
czyli $P(A|B \cap C) < P(A|B^C \cap C)$;
$P(A|B \cap C^C) = \frac{0}{16} = 0$ oraz $P(A|B^C \cap C^C) = \frac{4}{143} \approx 0,028$
czyli $P(A|B \cap C^C) < P(A|B^C \cap C^C)$.

Jeżeli ofiarą jest biała osoba to kara śmierci jest orzeczona o 11,6% częściej w przypadku czarnego sprawcy niż w przypadku białego sprawcy. Jeżeli ofiarą jest czarna osoba to kara śmierci jest orzeczona o 2,8% częściej w przypadku czarnego sprawcy niż w przypadku białego sprawcy. Natomiast nie uwzględniając rasy ofiary konkluzja jest odwrotna: kara śmierci jest orzeczona o 3,1% częściej w przypadku białego sprawcy niż w przypadku czarnego sprawcy.

Ten pozornie niemożliwy efekt niespodziewanie pojawia się w naukach społecznych i statystyce związanej z medycyną, kiedy zmienna ważona, która różni się od wartości określonej indywidualnie dla poszczególnych grup, jest używana do oceny połączonych grup. [4]

[3] *Dane pochodzą z: M. L. Radelet i G. E. Pierce,* **Florida Law Review**, *vol. 43, 1-34, 1991.*
[4] https:/pl.wikipedia.org/wiki/Paradoks.Simpsona