

RAPORT 1

Aleksandra Niedziela

2023-04-28

Wzór Bayesa

Wzór Bayesa to twierdzenie, mówiące na temat prawdopodobieństwa warunkowego:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

gdzie A i B są zdarzeniami. W naszym przypadku będziemy badać prawdopodobieństwa związane z testowaniem osób na pewną chorobę. Mamy test, który u 92% osób chorych wskaże chorobę, natomiast u 94% zdrowych osób, test jej nie wskaże. Na tę chorobę choruje 10% populacji.

Zastanówmy się najpierw jakie jest prawdopodobieństwo tego, iż u losowo wybranej osoby test wykaże chorobę. Oznaczamy:

- B: test wykazał chorobę
- A: osoba jest chora
- A': osoba jest zdrowa

Policzmy teraz prawdopodobieństwo całkowite B:

$$P(B) = P(B|A)P(A) + P(B|A')P(A') = 0.92 \cdot 0.10 + 0.06 \cdot 0.90 = 14.6\%$$

Czyli z prawdopodobieństwem 14.6 % uzyskamy pozytywny wynik testu u losowo wybranej osoby z populacji.

Teraz stosując wzór Bayesa, policzmy prawdopodobieństwo, czy losowo osoba jest chora, jeżeli test wykazał u niej chorobę.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.92 \cdot 0.10}{0.92 \cdot 0.10 + 0.06 \cdot 0.90} = 63.01\%$$

Teraz przeprowadzimy symulację. Stwórzmy populację liczącą milion osób, w której 10% z nich jest chorych. Będziemy wybierać jedną osobę i poddawać ją testowi, zliczymy ile testów było pozytywnych, a następnie sprawdzimy jaki jest to procent wszystkich przeprowadzonych testów. Eksperyment będziemy powtarzać 100000

Otrzymaliśmy 14577 testów pozytywnych, stanowi to 14.58% wszystkich przeprowadzonych testów, co zgadza się z teoretycznymi wyliczeniami.

W drugiej symulacji najpierw przeprowadzimy 100000 testów, a następnie policzymy ile spośród pozytywnych testów, prawdziwie pokazało chorobę:

Otrzymaliśmy 14659 testów pozytywnych, w tym 9236 u osób chorych, co oznacza iż osoba, która otrzymała pozytywny test jest chora z prawdopodobieństwem 63.01%. Zgadza się to z wyliczeniami teoretycznymi.

Prawdopodobieństwo warunkowe

Państwo Kowalscy mają dwójkę dzieci. Rozpatrujemy cztery sytuacje:

1. Powiedziano ci, że co najmniej jedno z dzieci jest dziewczyną. Jakie jest prawdopodobieństwo, że mają dwie dziewczynki?
2. Teraz założmy, że powiedziano ci, iż starsze dziecko jest dziewczyną. Jakie jest prawdopodobieństwo, że mają dwie dziewczynki?
3. Tym razem powiedziano ci, że co najmniej jedno z dzieci jest dziewczyną urodzoną we wtorek. Jakie jest prawdopodobieństwo, że mają dwie dziewczynki?
4. Jakie jest prawdopodobieństwo, że mają dwie dziewczynki gdy wiesz, że co najmniej jedno z dzieci jest dziewczyną urodzoną we wtorek lub w piątek?

Zadałam te pytania kilku osobom, które nie są związane z matematyką. W poniższej tabelce znajdują się ich odpowiedzi:

Numer pytania	1	2	3	4
Masażystka	50%	50%	50%	50%
Księgowa	33.3%	50%	4.7%	9.5%
Artysta	50%	50%	50%	50%
Uczennica	25%	50%	50%	50%
Chemiczka	25%	50%	50%	50%
Przyszła lekarka	25%	50%	50%	50%
Kolega Artysty	50%	50%	50%	50%

Przeprowadźmy teraz symulacje. Tworzymy teraz 1000000 przykładowych rodzin za każdym razem zliczamy jaki procent pasuje do warunków postawionych w pytaniach:

1. Zliczamy ile jest rodzin w których co najmniej jedno z dzieci jest dziewczynką, a następnie ile jest rodzin, w których dwoje dzieci to dziewczynki. Otrzymujemy prawdopodobieństwo: 33.35%
2. Zliczamy ile jest rodzin w których starsze z dzieci jest dziewczynką, a następnie ile jest rodzin, w których dwoje dzieci to dziewczynki. Otrzymujemy prawdopodobieństwo: 50.02%
3. Zliczamy ile jest rodzin w których co najmniej jedno z dzieci jest dziewczynką urodzoną we wtorek, a następnie ile jest rodzin, w których dwoje dzieci to dziewczynki. Otrzymujemy prawdopodobieństwo: 48.17%
4. Zliczamy ile jest rodzin w których co najmniej jedno z dzieci jest dziewczynką urodzoną we wtorek lub w piątek, a następnie ile jest rodzin, w których dwoje dzieci to dziewczynki. Otrzymujemy prawdopodobieństwo: 48.12%

Wnioski

Na 1. pytanie poprawnie odpowiedziała jedna z badanych osób, na 2. natomiast wszystkie badane osoby. Pokazuje to, iż prawdopodobieństwo warunkowe może być problematyczne i nieintuicyjne. Na pytania 3. i 4. pierwszą odpowiedzią przeważnie było: “Ale co informacja o dniu urodzenia ma tutaj do rzeczy?”. Okazuje się, iż ta informacja faktycznie wpływa na prawdopodobieństwo.