Matematyka dla biologów — Zajęcia nr 11.

Rachunek prawdopodobieństwa

Aksjomaty rachunku prawdopodobieństwa (przypomnienie)

Niech Ω, zbiór mający skończenie wiele elementów, będzie przestrzenią zdarzeń elementarnych. Prawdopodobieństwo zdarzenia A oznaczamy przez $P(A)$.

1. Dla dowolnego zdarzenia $A \subset \Omega\,,\ 0 \leqslant P(A) \leqslant 1\,.$
2. $P(\Omega) = 1$
3. Dla każdej pary rozłącznych zdarzeń A i B

$$P(A \cup B) = P(A) + P(B)\,.$$

Drugi aksjomat mówi, że zajście któregoś ze zdarzeń spośród wszystkich możliwych jest pewne tzn. że zbiór Ω uwzględnia wszystkie możliwe zdarzenia, które mogą zajść w danych okolicznościach.

Trzeba podkreślić, że prawdopodobieństwo jest funkcją określoną na zbiorze zdarzeń, czyli podzbiorów zbioru Ω i z tego powodu, bez wcześniejszego ustalenia konwencji notacyjnej, zapis $P(\omega)$ dla określenia prawdopodobieństwa zdarzenia elementarnego $\omega \in \Omega$ jest niepoprawny. Powinno zapisać się $P(\{\omega\})$

W danym zbiorze zdarzeń elementarnych prawdopodobieństwo można wprowadzić na wiele różnych sposobów byle spełnione były aksjomaty. Najprostsza sytuacja jest wtedy, gdy zdarzenia elementarne są jednakowo prawdopodobne. Rozpatrzmy n-elementową przestrzeń zdarzeń elementarnych

$$\Omega = \{\omega_1\,,\omega_2\,,\ldots\,,\omega_n\}$$

Wtedy dla każdego $i \in \{1\,,\ldots n\}$ mamy

$$P(\{\omega_i\}) = \frac{1}{n}\,.$$

Konsekwentnie jeśli jakieś zdarzenie $A \subset \Omega$ ma m elementów to

$$P(A) = \frac{m}{n}\,.$$

W dwuelementowym zbiorze zdarzeń elementarnych $\Omega_{0R} \overset{ozn.}{=} \{0\,,R\}$ możemy zadać prawdopodobieństwo P_1 przyjmując, że zdarzenia elementarne są jednakowo prawdopodobne

$$P_1(\{0\}) = \frac{1}{2} = P_1(\{R\})$$

lub przyjąć, że

$$P_2(\{0\}) = 0.499\,, \quad P_2(\{R\}) = 0.501\,.$$

Przestrzeń probabilistyczna $(\Omega_{0R}\,, P_1)$ jest **modelem probabilistycznym** dla doświadczenia losowego polegającego na jednokrotnym rzucie monetą idealnie symetryczną.

Czy to jest dobry model przekonać można się tylko wykonując doświadczenia w postaci długiej serii powtórzeń rzutu monetą w tych samych warunkach. Matematyczna teoria prawdopodobieństwa nie zajmuje się w zasadzie tym czy dany model probabilistyczny dobrze opisuje przebieg konkretnego doświadczenia. Dostarcza jedynie ogólnej teorii dającej podstawy do prowadzenia takich badań i daje także podstawy do badań statystycznych. Jednym z celów statystki jest oszacowanie na podstawie danych empirycznych prawdopodobieństw różnych zjawisk. Tak otrzymane prawdopodobieństwa konkretyzują model probabilistyczny. Po przeprowadzeniu dostatecznie wielu prób może się okazać, że przestrzeń probabilistyczna (Ω_{0R}, P_2) lepiej opisuje rzeczywisty przebieg doświadczenia, a więc moneta nie jest idealnie symetryczna.

Podstawy do wyciągnięcia tego typu wniosków daje słynne twierdzenie Bernoulliego (Jacob Bernoulli (1654-1705) zwane **prawem wielkich liczb.** Wynika z niego, że częstość względna wyrzucenia orła w ciągu niezależnych rzutów monetą dąży, przy liczbie prób dążącej do nieskończoności, do prawdopodobieństwa jego wyrzucenia w jednej próbie. W tym miejscu aksjomatycznie wprowadzone pojęcie prawdopodobieństwa ”spotyka się” ze swoim pierwowzorem - pojęciem częstości.

Wybór przestrzeni probabilistycznej jako modelu zjawiska losowego

Przykład: Jakie jest prawdopodobieństwo, że w wybranej na chybił trafił (czyli losowo) rodzinie dwudzietnej jest dwóch chłopców.

Przedstawimy dwa rozumowania

Błędne rozumowanie (błędny wybór przestrzeni zdarzeń)

Możliwe są trzy rodzaje rodzin z dwojgiem dzieci:

dwóch chłopców, dwie dziewczynki oraz mieszane.

Przestrzeń zdarzeń elementarnych jest zatem trójelementowa. Przyjmując, że wszystkie zdarzenia elementarne są jednakowo prawdopodobne dostajemy odpowiedź $\frac{1}{3}$.

Jest to w pierwszej chwili racjonalne, ale nie odpowiada rzeczywistości. Rodziny z parą dzieci różnej płci spotyka się częściej niż pozostałe.

Wskazówka: Porównajmy tę sytuację do dwukrotnego rzutu monetą symetryczną.

Rozumowanie poprawne.

Zbudujmy inny model probabilistyczny, który uwzględnia kolejność urodzeń. Mamy zatem

$$\Omega = \{(D\,,C)\,,(D\,,D)\,,(C\,,C)\,,(C\,,D)\}$$

Teraz przestrzeń zdarzeń elementarnych jest czteroelementowa. Jeśli przyjąć zgodnie z wiedzą medyczną i zdrowym rozsądkiem, że wszystkie zdarzenia elementarne są tu jednakowo prawdopodobne dostajemy odpowiedź $\frac{1}{4}$. Ten wynik odpowiada z dobrą dokładnością temu co obserwuje się w rzeczywistości. Warto się zastanowić dlaczego przyjęliśmy, że zdarzenia elementarne są jednakowo prawdopodobne. Wrócimy do tej kwestii gdy powiemy co to jest niezależność zdarzeń. Zauważmy jeszcze, że zdarzenie odpowiadające spotkaniu rodzin z dwójką dzieci różnej płci $\{(D\,,C)\,,(C\,,D)\}$ ma dwa razy więcej elementów niż zdarzenia $\{(D\,,D)\}$ oraz $\{(C\,,C)\}$ a więc błędem było przyjęcie na początku, że każdy z trzech rodzajów rodzin jest jednakowo prawdopodobny.

Zastosowanie kombinatoryki w rachunku prawdopodobieństwa

Zadanie: Liczby naturalne od 1 do 5 są zapisane na pięciu kartkach (na jednej kartce tylko jedna liczba). Losujemy trzy kartki. Jakie jest prawdopodobieństwo P tego, że suma wylosowanych liczb jest liczbą parzystą?

Odpowiedź: Ponieważ kolejność składników w sumie nie wpływa na wynik, zbiór zdarzeń elementarnych ma tyle elementów ile jest kombinacji 3-elementowych ze zbioru 5-elementowego, czyli C_5^3. Wsród liczb zapisanych na kartkach są trzy nieparzyste i dwie parzyste zatem zdarzeniu o które chodzi sprzyja wylosowanie dwóch liczb nieparzystych i jednej parzystej. Łatwo zauważyć, że

$$P = \frac{C_3^2 \cdot C_2^1}{C_5^3} = \frac{6}{10} = \frac{3}{5}\,.$$

Niezależność zdarzeń

Nech Ω będzie przestrzenią zdarzeń elementarnych a P prawdopodobieństwem.

Definicja

Zdarzenia $A, B \subset \Omega$ określamy jako niezależne w.t.w. gdy

$$P(A \cap B) = P(A) \cdot P(B).$$

Zauważmy, że jeśli zdarzenia A i B są rozłączne i każde z nich zachodzi z dodatnim prawdopodobieństwem to nie są one niezależne.

Sprawdzić, że zdarzenia

A -"wyciągnięcie losowo asa z talii 52 kart" ,

B- "wyciągnięcie karty kier"

są niezależne.

Prawdopodobieństwo warunkowe

Definicja

Przyjmijmy, że $B \subset \Omega$ jest takim zdarzeniem, że $P(B) > 0$. Prawdopodobieństwem warunkowym zajścia zdarzenia A pod warunkiem, że zaszło B nazywamy iloraz

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

Jeśli ustalimy zdarzenie $B \subset \Omega$ to $P(A|B)$ określa także prawdopodobieństwo na zdarzeniach $A \subset \Omega$. By się o tym przekonać wystarczy sprawdzić czy $P(A|B)$ spełnia wszystkie aksjomaty. Tak określone prawdopodobieństwo różni się od prawdopodobieństwa $P(A)$ bo wyraża dodatkową wiedzę o zajściu zdarzenia B.

Prosty przykład

Rzucamy kością do gry. Rozpatrzmy zdarzenia

A- wyrzucono ”3”

B- wyrzucono nieparzystą liczbę oczek Wtedy

$$P(A) = \frac{1}{6} \quad \text{ale} \quad P(A|B) = \frac{1}{3}.$$

A więc ta dodatkowa informacja o wyniku rzutu zwiększa prawdopodobieństwo dwukrotnie!!

Przykład na zast. prawdopodobieństwa warunkowego

Przykład liczba chłopców w rodzinach dwudzietnych. Postawmy dwa pytania:

1. **Jakie jest prawdopodobieństwo zdarzenia, że w losowo wybranej rodzinie dwudzietnej jest dwóch chłopców pod warunkiem, że w tej rodzinie jest przynajmniej jeden chłopiec.** (stwierdzono na przykład, że w pokoju dziecinnym na podłodze rozstawiona jest kolejka elektryczna, więc jeden chłopiec jest:-)).
2. **Jakie jest prawdopodobieństwo zdarzenia, że w losowo wybranej rodzinie dwudzietnej jest dwóch chłopców.**

Błędne oraz poprawne rozumowanie

Ad.1.(**błędne rozumowanie**) Skoro wiemy, że w rodzinie jest przynajmniej jeden chłopiec to oczywiście drugi może być z prawdopodobieństwem $\frac{1}{2}$ bo są dwie możliwości chłopiec albo dziewczynka.

Poprawne rozumowanie jest następujące. Przestrzeń zdarzeń elementarnych odpowiada chronologii urodzeń, mianowicie

$$\Omega = \{(D\,, C)\,, (D\,, D)\,, (C\,, C)\,, (C\,, D)\}\,.$$

Przyjmijmy dodatkowo, zgodnie ze zdrowym rozsądkiem, że zdarzenia elementarne są jednakowo prawdopodobne. Niech A będzie zdarzeniem, że w rodzinie są dwaj chłopcy a B, że jest co najmniej jeden . Wtedy

$$A \cap B = \{(C\,, C)\} \quad \text{oraz} \quad B = \{(D\,, C)\,, (C\,, C)\,, (C\,, D)\}\,.$$

Skoro zdarzenia elementarne są jednakowo prawdopodobne, z definicji prawdopodobieństwa warunkowego otrzymujemy, że poszukiwane prawdopodobieństwo jest znacznie mniejsze niż przy błędnym rozumowaniu!

$$P(A|B) = \frac{1/4}{3/4} = \frac{1}{3}\,,$$

Ad.2 Łatwo obliczyć, że

$$P(\{C\,,C\}) = \frac{1}{4}\,.$$

a więc jest ono istotnie mniejsze od przypadku gdy posiadamy dodatkową informację o obecności choćby jednego chłopca wśród pary dzieci. W praktyce można ”grubo” oszacować, że przy tej dodatkowej informacji w około 33 przypadkach na 100 w rodzinie są dwaj chłopcy.

Prawdopodobieństwo całkowite

Bardzo ważne konsekwencje ma wzór, który praktycznie pozwala obliczać prawdopodobieństwo zdarzenia A jeśli mamy dodatkowo dwie informacje o prawdopodobieństwie tego, że A zaszło pod warunkiem zajścia zdarzenia B oraz, że zaszło A pod warunkiem, że zaszło zdarzenie przeciwne do B oznaczone przez $\bar{B}$. Skoro mianowicie

$$A = (A \cap B) \cup (A \cap \bar{B})$$

to korzystając z trzeciego aksjomatu, a potem z definicji prawdopodobieństwa warunkowego otrzymujemy

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})\,.$$

Ogólniej w ten sam sposób można udowodnić tzw. **twierdzenie o prawdopodobieństwie całkowitym** sformułowane poniżej.

Mówimy, że zdarzenia $B_1, B_2, \ldots B_n \subset \Omega$ tworzą **zupełny układ zdarzeń** jeśli

- dla każdego i mamy $P(B_i) > 0$,
- są parami rozłączne

$$B_i \cap B_j = \emptyset \quad \text{dla} \quad i \neq j$$

- jedno z nich na pewno zajdzie

$$B_1 \cup B_2 \cup \ldots B_n = \Omega \, .$$

Twierdzenie

Niech $B_1, B_2, \ldots B_n$ będzie zupełnym układem zdarzeń w Ω . Wówczas dla dowolnego zdarzenia $A \subset \Omega$ zachodzi

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots P(A|B_n)P(B_n) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \, .$$

Przykład

Rozważmy rodziny trójdzietne i obliczmy prawdopodobieństwo zdarzenia A że, losowo wybrane dziecko z rodziny trójdzietnej jest dziewczynką, która ma starszego brata. Przestrzeń zdarzeń elementarnych jest teraz 2^3-elementowym zbiorem trójek uporządkowanych reprezentujących kolejne dzieci w rodzinie. Rozpatrzmy układ zupełny zdarzeń

1. B_1- pierwsza dziewczynka w rodzinie jest pierwszym dzieckiem,
2. B_2- pierwsza dziewczynka w rodzinie jest drugim dzieckiem,
3. B_3- pierwsza dziewczynka w rodzinie jest trzecim dzieckiem,
4. B_4- w rodzinie są sami chłopcy .

(te zdarzenia są rozłączne proszę to sprawdzić wypisując elementy tych zbiorów!)

Przyjmując, że zdarzenia elementarne są jednakowo prawdopodobne mamy

$$P(B_1) = \frac{1}{2}\,, P(B_2) = \frac{1}{4}\,, P(B_3) = \frac{1}{8}\,, P(B_4) = \frac{1}{8}\,.$$

Łatwo sprawdzić, że $P(A|B_1) = \frac{1/8}{1/2} = \frac{1}{4}\,, P(A|B_2) = 1\,, P(A|B_3) = 1$ i korzystając ze wzoru na prawdopodobieństwo całkowite dostajemy

$$P(A) = \frac{1}{4}\cdot\frac{1}{2} + 1\cdot\frac{1}{4} + 1\cdot\frac{1}{8} + 0\cdot\frac{1}{8} = \frac{1}{2}\,.$$

Można było oczywiście wyliczyć to prawdopodobieństwo bezpośrednio znajdując cztery zdarzenia elementarne sprzyjające zdarzeniu A.

Wywiad ze zrandomizowaną (ulosowioną) odpowiedzią.

Chodzi tu o sposób uzyskania w ramach badań socjologicznych, danych dotyczących szczegółów życia intymnego w jakiejś dostatecznie dużej populacji ludzkiej w taki sposób aby udzielający odpowiedzi mógł całkowicie zataić odpowiedź na postawione "trudne" pytanie dotyczące np. kontaktów pozamałżeńskich. Osobie, której dane personalne są nieznane zadaje się pytanie uzależniając odpowiedź od rezultatu doświadczenia losowego o znanym prawdopodobieństwie zajścia np. przed udzieleniem odpowiedzi TAK lub NIE badana osoba rzuca monetą. Jeśli wyrzuci orła zobowiązana jest uczciwie odpowiedzieć na pytanie intymne, jeśli wyrzuci reszkę osoba udziela odpowiedzi na "niewinne"pytanie pomocnicze typu czy urodziła się w pierwszej połowie roku. Po uzyskaniu odpowiedzi TAK badający nie wie czy jest to odpowiedź na pytanie intymne czy pomocnicze. W ten sposób z jednej strony zagwarantowana jest anonimowość w każdym poszczególnym przypadku a z drugiej dane te wystarczają na uzyskanie poszukiwanej informacji dotyczącej całej populacji.

Aby to wyjaśnić rozpatrzmy zdarzenia: $''O''$- wyrzucenie orła i udzielenie odpowiedzi na pytanie intymne, $''R''$- wyrzucenie reszki i udzielenie odpowiedzi na pytanie pomocnicze, "T"-udzielenie odpowiedzi TAK . Stosując wzór na prawdopodobieństwo całkowite i oznaczając przez P prawdopodobieństwo zdarzenia otrzymujemy, że

$$P(''T'') = P(''T''|''O'')P(''O'') + P(''T''|''R'')P(''R'')$$

Chcemy obliczyć $P(''T''|''O'')$. Oczywiście możemy przyjąć, że $P(''O'') = P(''R'') = \frac{1}{2}$ oraz, że prawdopodobieństwo spotkania osoby urodzonej w pierwszej połowie roku czyli w naszym przypadku P("T"|"R") równe jest też $\frac{1}{2}$. Dzieląc liczbę wszystkich odpowiedzi TAK przez liczbę wszystkich badanych uzyskujemy częstość, która daje oszacowanie wielkości $P(''T'')$. Ostatecznie otrzymujemy

$$P(''T''|''O'') = \frac{P(''T'') - 1/4}{1/2}\,.$$

co pozwala stwierdzić jaka jest frakcja osób w badanej grupie przejawiająca skłonność o którą chodzi w pytaniu intymnym.

Twierdzenie Bayesa

Z definicji prawdopodobieństwa i wzoru na prawdopodobieństwo całkowite wynika łatwo słynne **Twierdzenie Bayesa** które stosuje się bardzo często w różnych działach biologii i w medycynie. Przy oznaczeniach jak poprzednio, z definicji prawdopodobieństwa warunkowego wynika, że

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{P(A)}.$$

A stąd i wzoru na prawdopodobieństwo całkowite wynika twierdzenie Bayesa

Twierdzenie

Jeżeli $B_1, B_2, \ldots, B_i, \ldots, B_n$ jest zupełnym układem zdarzeń w zbiorze zdarzeń elementarnych Ω to dla dowolnego zdarzenia $A \subset \Omega$

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}. \tag{1}$$

Wspomaganie decyzji lekarskich.

Przypuśćmy, że w populacji ludzkiej można wyodrębnić rozłączne warstwy np. rasy , grupy zawodowe lub grupy ryzyka ze względu na jakiś czynnik chorobowy. Zakładając tę ostatnią sytuację przyjmijmy, że prawdopodobieństwo wybrania losowo osoby z i-tej grupy wynosi $P(B_i)$ i można je określić w przybliżeniu metodami statystycznymi jako częstość występowania. Niech A będzie zdarzeniem, że u przypadkowej osoby np. w trakcie rutynowych badań wykryto symptomy groźnej choroby i z jakiś powodów bezpośrednie stwierdzenie do której grupy należy dana osoba nie jest możliwe lub jest długotrwałe i kosztowne. Wiadomo jakie są prawdopodobieństwa warunkowe $P(A|B_i)$ czyli w praktyce jak często dana choroba występuje w i-tej grupie ryzyka. Chcemy zatem wiedzieć, jakie jest prawdopodobieństwo, że w danym przypadku po stwierdzeniu symptomów choroby dana osoba pochodzi z danej i-tej grupy. Taka informacja może być kluczowa jeśli od stwierdzenia przynależności do danej grupy zależy wybór terapii. Tu właśnie z pomocą przychodzi twierdzenie Bayesa.

Spośród wszystkich wartości $P(B_k|A)$ dla $k = 1, \ldots, n$ największa z nich wskazuje najbardziej prawdopodobną grupę ryzyka u której występuje dana choroba. Ta informacja pozwala podjąć lekarzowi racjonalną decyzję o wyborze terapii.

1.Zadania do zrobienia na zajęciach

Niech (Ω, P) będzie przestrzenią probabilistyczną, a $A, B \subset \Omega$ pewnymi zdarzeniami.

1. Wykazać, że

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

2. Wykazać, że jeśli zdarzenie A jest niezależne od B to A jest też niezależne od zdarzenia przeciwnego $\bar{B} = \Omega \setminus B$
3. Wykazać, że jeśli zdarzenia A i B są niezależne to

$$P(A) = P(A|B)$$

2.Zadania do zrobienia na zajęciach

Pewnego typu mikroskopy, niczym się zewnętrznie nie różniące, są produkowane przez fabryki A,B,C. Stwierdzono, że 35% mikroskopów na rynku pochodzi z firmy A i podobnie 35% z firmy B. Pozostałe pochodzą z firmy C.

Wiadomo, że odpowiednio $1,5\%, 1\%, 2\%$ mikroskopów produkowanych w firmach A, B,C jest wadliwych.

Natrafiono na wadliwy mikroskop. Korzystając z Tw. Bayesa, określić w której firmie został on najprawdopodobniej wykonany.

Wskaz. Niech $P(A) = 0,35$ to będzie prawdopodobieństwo napotkania mikroskopu zbudowanego w fabryce A, a D zdarzenie, że losowo wybrany mikroskop okazał się wadliwy. Wtedy

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{P(D \cap A)}{P(D)} = \frac{P(A)P(D|A)}{P(D)} = ...$$