Rachunek prawdopodo-bieństwa i statystyka

Wykład 2

Prawdopodobieństwo geometryczne

Przykład: Przestrzeń zdarzeń elementarnych określona jest przez zestaw punktów (x, y) na płaszczyźnie i wypełnia wnętrze kwadratu $[0 \le x \le 1; 0 \le y \le 1]$. Znajdź p-stwo, że dowolny punkt wewnątrz kwadratu znajduje się w odległości mniejszej niż r od wybranego rogu kwadratu.

Szukane p-two wyraża się stosunkiem pola powierzchni części wspólnej okręgu o promieniu r i kwadratu do pola powierzchni kwadratu (równego 1):

$$r < 1 \qquad P = \int_0^r \sqrt{r^2 - x^2}\,dx = \frac{1}{2}\left(x\sqrt{r^2 - x^2} + r^2 \arcsin\frac{x}{r} \right)\Bigg|_0^r = \frac{\pi r^2}{4}$$

$$r > 1 \qquad P = \int_0^r \sqrt{r^2 - x^2}\,dx - 2\int_1^r \sqrt{r^2 - x^2}\,dx = \frac{\pi r^2}{4} - 2\left(\frac{\pi r^2}{4} - \frac{1}{2}\sqrt{r^2 - 1} - \frac{1}{2} r^2 \arcsin\frac{1}{r} \right) =$$

$$= \sqrt{r^2 - 1} + r^2 \arcsin\frac{1}{r} - \frac{\pi r^2}{4}$$

Widać, że dla $r = \sqrt{2}$, p-two osiąga wartość równą 1.

Wzór Stirlinga

Funkcja gamma Eulera zdefiniowana jest przez:

$$\Gamma(z+1) = \int_0^\infty x^z e^{-x}\,dx$$

Ze związku tego wynika, że: $\Gamma(z+1) = z\Gamma(z)$ a stąd dla wartości całkowitych $z = n$ mamy: $\Gamma(n+1) = n!$

Postać całkowa dopuszcza uogólnienie funkcji silnia na dowolne liczby rzeczy-wiste, różne od ujemnych całkowitych. W szczególności (dla n=1, 2, ...):

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \qquad \Gamma\left(n+\frac{1}{2}\right) = \frac{1\cdot 3\cdot 5\cdot \ldots \cdot (2n-1)}{2^n}\sqrt{\pi}$$

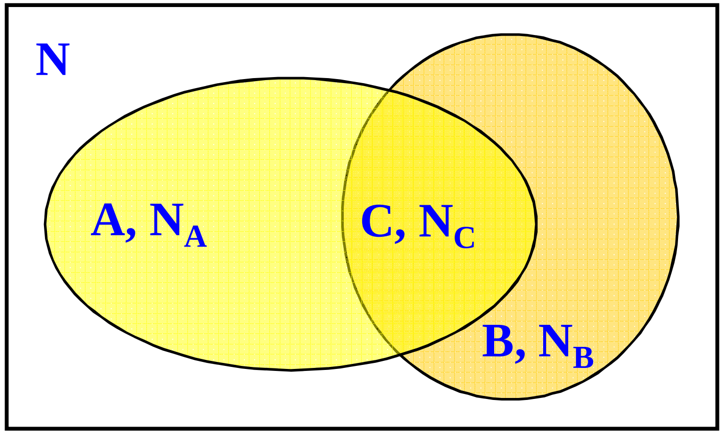
Wzór Stirlinga pozwala na przybliżone obliczanie funkcji silnia:

$$\Gamma(x+1) \equiv x! \cong \sqrt{2\pi}\, x^{x+\frac{1}{2}} \exp\left(-x + \frac{1}{12x} - \frac{1}{360x^2} + \ldots\right)$$

Przykład: 0.922 dla 1! , 1.92 dla 2! , 118.02 dla 5!=120, 0.08% błąd dla 100!

Prawdopodobieństwo warunkowe

<u>Przykład:</u> Wydajemy książkę. Po złożeniu przez zecera i wydrukowaniu tzw. szczotki oddajemy po jednej kopii dwóm korektorom, których zadaniem jest sprawdzenie tekstu.



N_A – liczba błędów znaleziona przez korektora A

N_B – liczba błędów znaleziona przez korektora B

N_C – liczba błędów znaleziona jednocześnie przez obu korektorów.

$$P(A) = \frac{N_A}{N}$$

$$P(B) = \frac{N_B}{N}$$

$$P(C) = P(A \cap B) = \frac{N_C}{N} = \begin{cases} \dfrac{N_C}{N_A}\dfrac{N_A}{N} = \dfrac{N_C}{N_A} P(A) \equiv P(B|A)P(A) \\ \dfrac{N_C}{N_B}\dfrac{N_B}{N} = \dfrac{N_C}{N_B} P(B) \equiv P(A|B)P(B) \end{cases}$$

P-twem warunkowym $P(A|B)$ zdarzenia A pod warunkiem B nazywamy p-two zajścia zdarzenia A obliczone przy założeniu, że zaszło zdarzenie B (przy czym zakładamy, że $P(B)>0$).

Prawdopodobieństwo warunkowe

Definicja: Prawdopodobieństwem warunkowym nazywamy liczbę określoną przez ($P(B) > 0$):

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

Twierdzenie: P-two iloczynu n zdarzeń można przedstawić w postaci:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1 | A_2 \cap A_3 \cap \ldots \cap A_n)\, P(A_2 | A_3 \cap A_4 \cap \ldots \cap A_n) \ldots P(A_n)$$

Przykład: Załóżmy, że w populacji złożonej z N osób jest N_K kobiet, N_D osób obu płci obciążonych wadą wzroku zwaną daltonizmem oraz N_{KD} kobiet daltonistek.

P-two, że losowo wybrana osoba jest kobietą wynosi: $P(K) = \frac{N_K}{N}$

P-two, że losowo wybrana osoba jest daltonistą wynosi: $P(D) = \frac{N_D}{N}$

P-two, że losowo wybrana osoba jest i kobietą i daltonistką: $P(K \cap D) = \frac{N_{KD}}{N}$

P-two, że losowo wybrana kobieta nie rozróżnia kolorów:

$$\frac{N_{KD}}{N_K} = \frac{N_{KD}/N}{N_K/N} = \frac{P(D \cap K)}{P(K)} = P(D|K)$$

P-two, że losowo wybrana osoba, która nie rozróżnia kolorów jest kobietą:

$$\frac{N_{KD}}{N_D} = \frac{N_{KD}/N}{N_D/N} = \frac{P(D \cap K)}{P(D)} = P(K|D)$$

Prawdopodobieństwo warunkowe

Przykład: Jakie jest p-two, że dni urodzin dwóch losowo wybranych osób są różne (B_2)?

$$P(B_2) = 1 - \frac{1}{365}$$

Jakie jest p-stwo , że dni urodzin trzech losowo wybranych osób są różne (B_3)?

A_3 – „dzień urodzin trzeciej osoby jest różny od dni urodzin dwóch pierwszych osób”

$$P(B_3) = P(A_3 \cap B_2) = P(A_3 | B_2) P(B_2)$$

$$P(A_3 | B_2) = 1 - \frac{2}{365} \quad \Rightarrow \quad P(B_3) = P(A_3 | B_2) P(B_2) = \left(1 - \frac{2}{365}\right)\left(1 - \frac{1}{365}\right) = 0.9918$$

W przypadku n osób mamy:

$$P(B_n) = P(A_n \cap B_{n-1}) = P(A_n | B_{n-1}) P(B_{n-1}) = P(A_n | B_{n-1}) P(A_{n-1} | B_{n-2}) P(B_{n-2}) =$$

$$= P(A_n | B_{n-1}) \cdot \ldots \cdot P(A_3 | B_2) P(B_2) = \left(1 - \frac{n-1}{365}\right)\left(1 - \frac{n-2}{365}\right) \cdot \ldots \cdot \left(1 - \frac{2}{365}\right) \cdot \left(1 - \frac{1}{365}\right)$$

Uwaga: W rozwiązywaniu problemów ważne jest które zdarzenie jest warunkiem:

$$P(A \cap B) = P(A | B) P(B) \quad \text{albo} \quad P(A \cap B) = P(B | A) P(A)$$

Niezależność zdarzeń

Definicja: Zdarzenia A i B nazywamy statystycznie niezależnymi gdy spełniona jest relacja $P(A \cap B) = P(A) \cdot P(B)$.

Korzystając z definicji p-twa warunkowego, p-two iloczynu zdarzeń A i B możemy zapisać w postaci:

$P(A \cap B) = P(A) \cdot P(B|A)$ gdy $P(A) > 0$ oraz $P(A \cap B) = P(B) \cdot P(A|B)$ gdy $P(B) > 0$.

Jeśli $P(A)>0$ i $P(B)>0$ to warunkiem koniecznym i wystarczającym na to aby zdarzenia były niezależne jest aby $P(A|B) = P(A)$ lub $P(B|A) = P(B)$.

Przykład: Rozważmy dwa zdarzenia określone na zbiorze liczb całkowitych od 1 do 100:

A – liczba wylosowana z tego zbioru jest podzielna przez 3,

B – wylosowana liczba jest podzielna przez 7.

Czy zdarzenia A i B są statystycznie niezależne?

$P(A) = 33/100$, $P(B) = 14/100$, $P(A \cap B) = 4/100 \rightarrow P(A) \cdot P(B) \neq P(A \cap B)$

Gdy jednak rozszerzymy zbiór o liczby 101, 102, 103, 104 i 105 to wówczas mamy:

$P(A) = 35/105$, $P(B) = 15/105$, $P(A \cap B) = 5/105 \rightarrow P(A) \cdot P(B) = P(A \cap B)$

Niezależność zdarzeń

Twierdzenie: A jest niezależne od B $\Leftrightarrow$ $\bar{A}$ jest niezależne od B

D: $$P(\bar{A} \mid B) = 1 - P(A \mid B) = \| A \text{ niezależne od } B \| = 1 - P(A) = P(\bar{A})$$

$$P(A \mid B) = 1 - P(\bar{A} \mid B) = \| \bar{A} \text{ niezależne od } B \| = 1 - P(\bar{A}) = P(A)$$

Definicja: Mówimy, że zdarzenia $A_1, A_2, \dots, A_n$ są wzajemnie niezależne, jeśli p-two łącznego zajścia dowolnych $m \le n$ różnych zdarzeń spośród nich jest równe iloczynowi p-stw tych zdarzeń, czyli

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_m})$$

dla każdego $m \le n$ i każdego m-wyrazowego rosnącego ciągu $i_1, i_2, \dots, i_m$ liczb naturalnych $1 \le i_1, i_2, \dots, i_m \le n$.

Definicja: Zdarzenia $A_1, A_2, \dots, A_n$ są niezależne parami gdy każde dwa różne spośród nich są niezależne, czyli $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ dla $i \neq j$ (gdzie $i, j = 1, 2, \dots, n$).

Uwaga: Zdarzenia wzajemnie niezależne są niezależne parami.

Niezależność zdarzeń - przykłady

Przykład: W eksperymencie polegającym na dwukrotnym rzucie monetą określamy:

A – „orzeł w pierwszym rzucie”, B – „orzeł w drugim rzucie”, C – „to samo w obu rzutach”

$$P(A)=P(B)=\frac{1}{2} \quad \text{oraz} \quad P(C)=P(A\cap B)+P(\bar{A}\cap\bar{B})=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$

$$P(C|A)=P(B|A)=P(B)=\frac{1}{2}=P(C) \quad \text{oraz} \quad P(C|B)=P(C)$$

$$P(A\cap B\cap C)=P(A\cap B)=\frac{1}{4} \quad \text{ale} \quad P(A)P(B)P(C)=\frac{1}{8}$$

Wniosek: Zdarzenia A, B i C są parami niezależne, ale nie są wzajemnie niezależne.

Przykład: P-two wystąpienia przynajmniej jednego z n niezależnych zdarzeń. Dla n = 3:

$$P(A\cup B\cup C)=$$

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

$$=P(A)+P(B)+P(C)-P(B)P(C)-P(A)P(B)-P(A)P(C)+P(A)P(B)P(C)$$

Prościej dochodzimy do tego wyniku, korzystając z p-twa zdarzenia przeciwnego.

Dla n niezależnych zdarzeń mamy:

$$P(A_1\cup A_2\cup\ldots\cup A_n)=1-P\left(\overline{A_1\cup A_2\cup\ldots\cup A_n}\right)=1-P(\bar{A}_1\cap\bar{A}_2\cap\ldots\cap\bar{A}_n)=$$

$$=1-P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n)=1-(1-P(A_1))(1-P(A_2))\cdots(1-P(A_n))$$

Efektywność detektora

Przykład: Efektywność (wydajność) detektora A to liczba p_A określająca p-two rejestracji pojedynczej cząstki, która wpadła do jego wnętrza.

Rozwiązanie teoretyczne: kierujemy na detektor znaną liczbę cząstek N, znajdujemy liczbę cząstek N_A, które zostały zarejestrowane i obliczamy $p_A=N_A/N$.

W praktyce jednak liczba cząstek padających nie jest znana!

Rozwiązanie praktyczne: ustawiamy na wiązce drugi detektor B, który rejestruje N_B cząstek, podczas gdy detektor A zarejestrował N_A cząstek.

Gdybyśmy znali efektywność detektora B wówczas oczywiście $p_A = (N_A/N_B) \cdot p_B$

Zakładamy, że detektory działają niezależnie od siebie i fakt rejestracji cząstki przez jeden z nich nie wpływa w żaden sposób na rejestrację cząstki w drugim detektorze.

Z niezależnego, $P(A \cap B) = P(A) \cdot P(B)$, działania obu detektorów mamy:

$$p_A = \frac{p_{A\cap B}}{p_B} = \frac{N \cdot p_{A\cap B}}{N \cdot p_B} = \frac{N_C}{N_B} \qquad p_B = \frac{p_{A\cap B}}{p_A} = \frac{N \cdot p_{A\cap B}}{N \cdot p_A} = \frac{N_C}{N_A}$$

Ocena liczby cząstek: $p_A = N_C/N_B = N_A/N \Rightarrow N = N_A \cdot N_B/N_C$

Prawdopodobieństwo całkowite

Przykład: Test na chorobę „szalonych krów” BSE.

Definiujemy zdarzenia: T – „wynik testu pozytywny”, B – „krowa jest chora na BSE”.

Wiadomo, że $\mathrm{P}\left(\mathrm{T}\,|\,\mathrm{B}\right)=0.70$ oraz $\mathrm{P}\left(\mathrm{T}\,|\,\bar{\mathrm{B}}\right)=0.10$ a także, że $\mathrm{P(B)} = 0.02$

Szukamy p-twa, że test wykonany dla losowo wybranej krowy, da wynik pozytywny.

$$\mathrm{T}=\left(\mathrm{T}\cap\mathrm{B}\right)\cup\left(\mathrm{T}\cap\bar{\mathrm{B}}\right)\quad\Rightarrow\quad \mathrm{P}\left(\mathrm{T}\right)=\mathrm{P}\left(\mathrm{T}\cap\mathrm{B}\right)+\mathrm{P}\left(\mathrm{T}\cap\bar{\mathrm{B}}\right)$$

Wykorzystując dane p-twa warunkowe możemy napisać:

$$\mathrm{P}\left(\mathrm{T}\right)=\mathrm{P}\left(\mathrm{T}\,|\,\mathrm{B}\right)\cdot\mathrm{P}\left(\mathrm{B}\right)+\mathrm{P}\left(\mathrm{T}\,|\,\bar{\mathrm{B}}\right)\cdot\mathrm{P}\left(\bar{\mathrm{B}}\right)=0.112$$

Definicja: Zbiór zdarzeń $\mathcal{H}=\left\{\mathrm{H}_1,\mathrm{H}_2,\ldots\right\}$ nazywamy podziałem przestrzeni zdarzeń elementarnych jeśli spełnione są warunki $\mathrm{H}_i\cap\mathrm{H}_j=\varnothing$, dla $i\neq j$ oraz $\bigcup_i \mathrm{H}_i=\Omega$

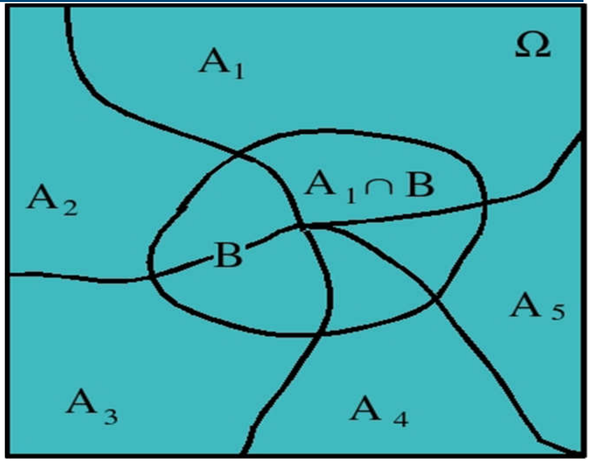
Definicja: Dysponując dwoma podziałami $\mathcal{H}=\left\{\mathrm{H}_1,\mathrm{H}_2,\ldots\right\}$ oraz $\mathcal{K}=\left\{\mathrm{K}_1,\mathrm{K}_2,\ldots\right\}$ mówimy, że podział $\mathcal{H}$ jest bardziej szczegółowy niż $\mathcal{K}$ jeśli dla każdego zdarzenia $\mathrm{H}_i\in\mathcal{H}$ istnieje zdarzenie $\mathrm{K}_j\in\mathcal{K}$ takie, że $\mathrm{H}_i\subset\mathrm{K}_j$.

Przykład: Dysponując dowolnym zdarzeniem A można utworzyć podział $\mathcal{K}=\left\{\mathrm{A},\bar{\mathrm{A}}\right\}$

Przykład: Para zdarzeń A i B pozwala zdefiniować zbiory, które tworzą podział $\mathcal{H}$ (będący podziałem bardziej szczegółowym niż podział $\mathcal{K}$ z poprzedniego przykładu):

$$\mathrm{H}_1=\mathrm{A}\cap\mathrm{B},\quad \mathrm{H}_2=\mathrm{A}\cap\bar{\mathrm{B}},\quad \mathrm{H}_3=\bar{\mathrm{A}}\cap\mathrm{B},\quad \mathrm{H}_4=\bar{\mathrm{A}}\cap\bar{\mathrm{B}},$$

Prawdopodobieństwo całkowite



Twierdzenie (o prawdopodobieństwie całkowitym):

Jeśli $\mathbf{B}$ jest dowolnym zdarzeniem, zaś zdarzenia $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ ($\mathbf{P}(\mathbf{A}_i) > 0$ dla $i = 1, 2, \ldots, n$), są parami rozłączne ($\mathbf{A}_i \cap \mathbf{A}_j = \emptyset$ gdy $i \neq j$) oraz ich suma jest zdarzeniem pewnym ($\mathbf{A}_1 \cup \mathbf{A}_2 \cup \ldots \cup \mathbf{A}_n = \Omega$) to p-two zdarzenia $\mathbf{B}$ jest równe:

$$\mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{B}|\mathbf{A}_1)\mathbf{P}(\mathbf{A}_1) + \mathbf{P}(\mathbf{B}|\mathbf{A}_2)\mathbf{P}(\mathbf{A}_2) + \ldots + \mathbf{P}(\mathbf{B}|\mathbf{A}_n)\mathbf{P}(\mathbf{A}_n)$$

Dowód:

$$\mathbf{P}(\mathbf{B}) = \mathbf{P}\big((\mathbf{B}\cap\mathbf{A}_1)\cup(\mathbf{B}\cap\mathbf{A}_2)\cup\ldots\cup(\mathbf{B}\cap\mathbf{A}_n)\big) =$$

$$= \mathbf{P}(\mathbf{B}\cap\mathbf{A}_1) + \mathbf{P}(\mathbf{B}\cap\mathbf{A}_2) + \ldots + \mathbf{P}(\mathbf{B}\cap\mathbf{A}_n) =$$

$$= \mathbf{P}(\mathbf{B}|\mathbf{A}_1)\mathbf{P}(\mathbf{A}_1) + \mathbf{P}(\mathbf{B}|\mathbf{A}_2)\mathbf{P}(\mathbf{A}_2) + \ldots + \mathbf{P}(\mathbf{B}|\mathbf{A}_n)\mathbf{P}(\mathbf{A}_n)$$

Uwaga: W przypadku kiedy podział przestrzeni zdarzeń elementarnych jest nieskończony, ale przeliczalny, to twierdzenie o p-twie całkowitym zapisujemy w postaci:

$$\mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{B}|\mathbf{A}_1)\mathbf{P}(\mathbf{A}_1) + \mathbf{P}(\mathbf{B}|\mathbf{A}_2)\mathbf{P}(\mathbf{A}_2) + \ldots$$

Prawdopodobieństwo całkowite

Przykład: Dane jest n+1 urn z białymi oraz czarnymi kulami, przy czym stosunek liczby białych kul do wszystkich dla i-tej urny wynosi i/n (i=0, 1, 2, … , n). Z losowo wybranej urny pobieramy jedną kulę. Ile wynosi p-two, że wybrana kula jest biała?

P-two wyboru każdej z urn wynosi: $\mathrm{P}(0) = \mathrm{P}(1) = \mathrm{P}(2) = \cdots = \mathrm{P}(n) = \frac{1}{n+1}$

P-two (warunkowe) wylosowania kuli białej z i-tej urny jest równe: $\mathrm{P}(\mathrm{B} \mid i) = \frac{i}{n}$

Szukane p-two wylosowania kuli białej wyznaczamy z formuły p-twa całkowitego:

$$\mathrm{P}(\mathrm{B}) = \mathrm{P}(\mathrm{B} \mid 0) \cdot \mathrm{P}(0) + \mathrm{P}(\mathrm{B} \mid 1) \cdot \mathrm{P}(1) + \cdots + \mathrm{P}(\mathrm{B} \mid n) \cdot \mathrm{P}(n) = \frac{1}{n+1} \sum_{k=0}^{n} \frac{k}{n} = \frac{1}{2}$$

Przykład: Test na chorobę „szalonych krów” BSE.

Zakładając, że wynik testu jest pozytywny, jakie jest p-two, że dana krowa ma BSE?

Wielkości dane: $\mathrm{P}(\mathrm{T} \mid \mathrm{B}) = 0.70, \quad \mathrm{P}(\mathrm{T} \mid \overline{\mathrm{B}}) = 0.10, \quad \mathrm{P}(\mathrm{B}) = 0.02$

Szukamy p-twa warunkowego:

$$\mathrm{P}(\mathrm{B} \mid \mathrm{T}) = \frac{\mathrm{P}(\mathrm{B} \cap \mathrm{T})}{\mathrm{P}(\mathrm{T})} = \frac{\mathrm{P}(\mathrm{T} \mid \mathrm{B}) \cdot \mathrm{P}(\mathrm{B})}{\mathrm{P}(\mathrm{T} \mid \mathrm{B}) \cdot \mathrm{P}(\mathrm{B}) + \mathrm{P}(\mathrm{T} \mid \overline{\mathrm{B}}) \cdot \mathrm{P}(\overline{\mathrm{B}})}$$

Otrzymujemy: $\mathrm{P}(\mathrm{B} \mid \mathrm{T}) = 0.125$ oraz $\mathrm{P}(\mathrm{B} \mid \overline{\mathrm{T}}) = 0.0068$

Prawdopodobieństwo całkowite

Twierdzenie: Niech zdarzenia $\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_n$ spełniają warunki twierdzenia o p-twie całkowitym, oraz $\mathbf{A}$ i $\mathbf{B}$ będą zdarzeniami takimi, że $\mathbf{P}(\mathbf{B}) > 0$, wówczas zachodzi

$$\mathbf{P}(\mathbf{A}|\mathbf{B}) = \sum_i \mathbf{P}(\mathbf{A}|\mathbf{B}\cap\mathbf{H}_i)\mathbf{P}(\mathbf{H}_i|\mathbf{B})$$

Dowód:

$$\sum_i \mathbf{P}(\mathbf{A}|\mathbf{B}\cap\mathbf{H}_i)\mathbf{P}(\mathbf{H}_i|\mathbf{B}) = \sum_i \frac{\mathbf{P}(\mathbf{A}\cap\mathbf{B}\cap\mathbf{H}_i)}{\mathbf{P}(\mathbf{B}\cap\mathbf{H}_i)}\frac{\mathbf{P}(\mathbf{H}_i\cap\mathbf{B})}{\mathbf{P}(\mathbf{B})} =$$

$$= \frac{1}{\mathbf{P}(\mathbf{B})}\sum_i \mathbf{P}(\mathbf{A}\cap\mathbf{B}\cap\mathbf{H}_i) = \frac{1}{\mathbf{P}(\mathbf{B})}\mathbf{P}\left(\mathbf{A}\cap\mathbf{B}\cap\bigcup_i \mathbf{H}_i\right) = \frac{\mathbf{P}(\mathbf{A}\cap\mathbf{B})}{\mathbf{P}(\mathbf{B})} = \mathbf{P}(\mathbf{A}|\mathbf{B})$$

Twierdzenie: Niech będą dane dwa podziały $\mathcal{H}$ i $\mathcal{H}'$, przy czym podział $\mathcal{H}'$ jest bardziej szczegółowy niż podział $\mathcal{H}$. Jeśli $\mathbf{H}_i = \mathbf{H}'_{i_1} \cup \ldots \cup \mathbf{H}'_{i_k}$ to dla dowolnego zdarzenia $\mathbf{A}$, p-twa warunkowe względem podziałów $\mathcal{H}$ i $\mathcal{H}'$ spełniają relację:

$$\mathbf{P}(\mathbf{A}|\mathbf{H}_i) = \sum_{j=1}^{k} \mathbf{P}(\mathbf{A}|\mathbf{H}'_{i_j})\mathbf{P}(\mathbf{H}'_{i_j}|\mathbf{H}_i)$$

Dowód:

$$\mathbf{P}(\mathbf{A}|\mathbf{H}_i) = \mathbf{P}(\mathbf{A}\cap\mathbf{H}_i|\mathbf{H}_i) = \mathbf{P}\left(\mathbf{A}\cap\bigcup_{j=1}^{k}\mathbf{H}'_{i_j}\middle|\mathbf{H}_i\right) = \sum_{j=1}^{k}\mathbf{P}(\mathbf{A}\cap\mathbf{H}'_{i_j}|\mathbf{H}_i) = \sum_{j=1}^{k}\frac{\mathbf{P}(\mathbf{A}\cap\mathbf{H}'_{i_j}\cap\mathbf{H}_i)}{\mathbf{P}(\mathbf{H}_i)} =$$

$$= \sum_{j=1}^{k}\frac{\mathbf{P}(\mathbf{A}\cap\mathbf{H}'_{i_j})}{\mathbf{P}(\mathbf{H}_i)} = \sum_{j=1}^{k}\frac{\mathbf{P}(\mathbf{A}|\mathbf{H}'_{i_j})\mathbf{P}(\mathbf{H}'_{i_j})}{\mathbf{P}(\mathbf{H}_i)} = \sum_{j=1}^{k}\mathbf{P}(\mathbf{A}|\mathbf{H}'_{i_j})\frac{\mathbf{P}(\mathbf{H}'_{i_j}\cap\mathbf{H}_i)}{\mathbf{P}(\mathbf{H}_i)} = \sum_{j=1}^{k}\mathbf{P}(\mathbf{A}|\mathbf{H}'_{i_j})\mathbf{P}(\mathbf{H}'_{i_j}|\mathbf{H}_i)$$

Twierdzenie Bayes'a

Twierdzenie Bayes'a: Jeżeli B jest dowolnym zdarzeniem o dodatnim p-twie, zaś zdarzenia $A_1, A_2, \ldots, A_n$ spełniają warunki twierdzenia o p-twie całkowitym, to p-two warunkowe $P(A_k|B)$ zdarzenia A_k $(k=1, \ldots, n)$ przy warunku B dane jest przez:

p-two zaobserwowania danych przy założeniu prawdziwości hipotezy

p-two a priori, tzn. przed zaobserwowaniem danych

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} = \frac{P(B|A_k)P(A_k)}{P(B)}$$

p-two a posteriori, tzn. po zaobserwowaniu danych

normalizacja – suma po wszystkich hipotezach

Przykład: Studentowi zadano pytanie i przedstawiono n możliwych odpowiedzi, z których tylko jedna jest poprawna. Jeśli student zna tę odpowiedź to na pewno ją wybierze, ale jeśli jej nie zna to wybiera jedną na chybił-trafił. Wykładowca zakłada, że z p-twem $P(Z)=0.5$ student zna poprawną odpowiedź. Jakie jest p-two, że student istotnie zna poprawną odpowiedź, jeśli taką wybrał?

$$P(W|Z)=1, \quad P(Z)=\frac{1}{2}, \quad P(W)=1\cdot\frac{1}{2}+\frac{1}{n}\cdot\frac{1}{2} \quad \Rightarrow \quad P(Z|W)=\frac{P(W|Z)\cdot P(Z)}{P(W)}=\frac{n}{n+1}$$

Twierdzenie Bayes'a - przykład

<u>Przykład:</u> Detektor cząstek ustawiony jest na wiązce składającej się w 90% z pionów i w 10% z protonów. Detektor ma wydajność detekcji protonu 95%, ale w 2% przypadków błędnie identyfikuje pion jako proton. Jakie jest p-two, że przez detektor faktycznie przeszedł proton?

A – zdarzenie polegające na rejestracji cząstki przez detektor
p – cząstką padającą był proton; π – cząstką padającą był pion

P(p) = 0.10 **P(π) = 0.90** **P(A|p) = 0.95** **P(A|π) = 0.02**

$$P(A) = P(A \cap \Omega) = P(A \cap (p \cup \pi)) = P((A \cap p) \cup (A \cap \pi)) =$$
$$= P(A \cap p) + P(A \cap \pi) = P(A|p)\cdot P(p) + P(A|\pi)\cdot P(\pi) = 0.95\ 0.1 + 0.02\ 0.9 = 0.113$$

P-twa że przez detektor faktycznie przeszedł proton lub pion wyznaczamy z twierdzenia Bayes'a:

$$P(p\,|\,A) = \frac{P(A\,|\,p)\cdot P(p)}{P(A)} = \frac{0.95\cdot 0.1}{0.113} = 84.1\%$$

$$P(\pi\,|\,A) = \frac{P(A\,|\,\pi)\cdot P(\pi)}{P(A)} = \frac{0.02\cdot 0.9}{0.113} = 15.9\%$$

Gdyby P(A|p)=0.1 to wówczas P(p|A) = 51.4% oraz P(π|A) = 48.6%

Twierdzenie Bayes'a w „Idź na całość"

W programie telewizyjnym „Idź na całość" prezentowano uczestnikowi troje drzwi: za jednymi z nich jest ukryty samochód, a za pozostałymi dwoma znajduje się „zonk". Uczestnik programu wygrywa to co znajduje się za wybranymi przez niego drzwiami. Nim jednak otworzone zostają wskazane przez uczestnika drzwi, prowadzący program, który wie co znajduje się za wszystkimi drzwiami, otwiera te drzwi z dwóch pozostałych za którymi znajduje się „zonk". Następnie zezwala grającemu zmienić decyzję co do wybranych przez siebie drzwi. Co powinien zrobić uczestnik programu?

Możliwe są trzy (I, II, III) układy drzwi (A, B, C) i przedmiotów (S, Z):

	A	B	C
I	S	Z	Z
II	Z	S	Z
III	Z	Z	S

Przed otwarciem jednych z drzwi, p-two każdego z układów jest takie samo i wynosi **P(I) = P(II) = P(III) = 1/3** (*a priori*).

Załóżmy, że wybieramy początkowo drzwi A, natomiast prowadzący program, który <u>wie gdzie jest samochód</u>, otwiera drzwi B.

Interesuje nas, które p-two warunkowe (*a posteriori*) jest większe:

P(I|B) – p-two, że wystąpił układ I jeśli prowadzący otwarł drzwi B

P(III|B) – p-two, że wystąpił układ III jeśli prowadzący otwarł drzwi B

Twierdzenie Bayes'a w „Idź na całość"

P-two, że prowadzący otworzył drzwi B wynosi:

$$P(B)=P(B|I)\cdot P(I)+P(B|II)\cdot P(II)+P(B|III)\cdot P(III)=\frac{1}{2}\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}=\frac{1}{2}$$

Obliczamy oba p-twa warunkowe:

$$P(I|B)=\frac{P(B|I)\cdot P(I)}{P(B)}=\frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}}=\frac{1}{3} \qquad P(III|B)=\frac{P(B|III)\cdot P(III)}{P(B)}=\frac{1\cdot\frac{1}{3}}{\frac{1}{2}}=\frac{2}{3}$$

Wniosek: Należy zmienić pierwotny wybór !

Uwaga: Jeśli prowadzący nie wie za którymi drzwiami jest samochód i otworzył drzwi B (za którymi okazało się, że nie ma samochodu) to nie ma znaczenia czy zmienimy nasz pierwotny wybór czy nie:

$$P(B)=P(B|I)\cdot P(I)+P(B|II)\cdot P(II)+P(B|III)\cdot P(III)=\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{3}$$

$$P(I|B)=\frac{P(B|I)\cdot P(I)}{P(B)}=\frac{\frac{1}{3}\cdot\frac{1}{2}}{\frac{1}{3}}=\frac{1}{2} \qquad P(III|B)=\frac{P(B|III)\cdot P(III)}{P(B)}=\frac{\frac{1}{3}\cdot\frac{1}{2}}{\frac{1}{3}}=\frac{1}{2}$$