Wnioskowanie bayesowskie

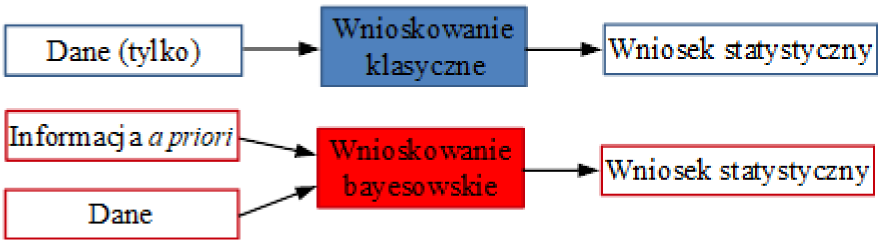
W podejściu klasycznym wnioskowanie statystyczne oparte jest wyłącznie na podstawie pobranej próby losowej. Możemy np. estymować punktowo lub przedziałowo nieznane parametry rozkładów, weryfikować hipotezy dotyczące rozkładów badanych cech, badać różne zależności pomiędzy badanymi cechami i wyznaczać różne związki między nimi. Na podstawie próby losowej wnioskujemy na temat całej badanej populacji.

Oprócz podejścia klasycznego istnieje tzw. podejście bayesowskie, które pozwala analitykowi korzystać nie tylko z wyników zaobserwowanych w próbie, ale także z informacji *a priori* dotyczących rozważanego problemu.

W podejściu klasycznym zakładamy, że parametry rozkładu badanej cechy w populacji, takie jak średnia, wariancja, frakcja elementów określonego rodzaju w populacji są określonymi, choć nieznanymi wielkościami. Natomiast w podejściu bayesowskim zakładamy, że interesujące nas parametry są zmiennymi losowymi. Możemy zatem na podstawie wiedzy eksperckiej (informacji *a priori*) przewidywać ich rozkłady i na podstawie danych z próby modyfikować te przypuszczenia.

W podejściu bayesowskim matematyczny związek prawdopodobieństw wynikających z informacji *a priori* i prawdopodobieństw wynikających z obserwacji próby opiera się na twierdzeniu Bayesa i stąd jego nazwa.

Schematyczne porównanie podejścia klasycznego z podejściem bayesowskim przedstawia poniższy rysunek:



W ogólności podejście bayesowskie nie musi jednak doprowadzić do wniosków dokładniejszych niż podejście klasyczne.

Jeżeli informacja *a priori*, którą dysponujemy jest dokładna, wówczas jej wykorzystanie łącznie z informacją pochodzącą z próby prowadzi do wniosków dokładniejszych od wniosków uzyskanych w podejściu klasycznym. Jeżeli jednak informacja *a priori* jest niedokładna, korzystanie z niej i z informacji z próby może prowadzić do wniosków mniej dokładnych od tych, do których doszlibyśmy pomijając informację *a priori*.

Jeżeli informacja *a priori* jest bezpośrednim rezultatem wcześniejszych badań statystycznych i nie opiera się na czyjejś przypadkowej wiedzy o badanym problemie, to analiza bayesowska jest w pełni obiektywna. Często jednak informacja *a priori* odzwierciedla tylko osobiste poglądy osoby przeprowadzającej analizę lub innej osoby dysponującej tylko pewną przypadkową wiedzą o badanym problemie. Wówczas informacja *a priori* ma charakter subiektywny i dlatego sama idea wykorzystania informacji *a priori* w analizie statystycznej jest często przez niektórych atakowana.

Przypomnimy teraz znane twierdzenie o prawdopodobieństwie całkowitym i twierdzenie Bayesa.

Niech $(\Omega, \mathcal{F}, P)$ będzie dowolną przestrzenią probabilistyczną.

Twierdzenie o prawdopodobieństwo całkowitym

Jeżeli $B_1, B_2, \dots, B_n \in \mathcal{F}$ są zdarzeniami losowymi wykluczającymi się, dla których $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$ oraz $P(B_i) > 0$ dla $i = 1,2,\dots,n$, to dla każdego zdarzenia losowego $A \in \mathcal{F}$ zachodzi następujący wzór:

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

Prawdopodobieństwa $P(B_i)$ nazywamy prawdopodobieństwami *a priori* (danymi z góry, przed doświadczeniem).

Wzór Bayesa (Twierdzenie Bayesa)

Niech będą spełnione założenia twierdzenia o prawdopodobieństwie całkowitym. Wówczas dla dowolnego zdarzenia losowego $A \in \Omega$ takiego, że $P(A) > 0$, zachodzi następujący wzór:

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A)}, i = 1, 2, \dots, n.$$

Prawdopodobieństwa $P(B_i|A)$ nazywamy prawdopodobieństwami *a posteriori* (danymi po zbadaniu, po doświadczeniu).

Zdarzenia $B_1, B_2, \dots, B_n$ nazywamy przyczynami, a zdarzenie A skutkiem. Z twierdzenia Bayesa wynika, że można obliczać prawdopodobieństwa *a posteriori*, gdy są znane ich prawdopodobieństwa *a priori* oraz warunkowe prawdopodobieństwa skutku przy danej przyczynie.

Twierdzenie Bayesa pozwala połączyć informację *a priori* z wynikami badań próbnych, czego wynikiem jest informacja *a posteriori* (po pobraniu próby).

Uogólnimy teraz twierdzenie Bayesa na zmienne losowe dyskretne (skokowe). W tym celu wykorzystamy warunkowe prawdopodobieństwa przyporządkowane poszczególnym elementom zbioru wszystkich możliwych wartości dyskretnej zmiennej losowej, które tworzą funkcję wiarygodności.

Funkcja wiarygodności przyporządkowuje warunkowe prawdopodobieństwa $P(x|\theta)$ pojawienia się wyniku obserwacji x możliwym wartościom nieznanego parametru θ.

Korzystając z funkcji wiarygodności i apriorycznych prawdopodobieństw $P(\theta)$ przyjęcia przez parametr θ określonych wartości, otrzymujemy następującą postać twierdzenia Bayesa:

Twierdzenie Bayesa dla dyskretnej zmiennej losowej

Zachodzi wzór

$$P(\theta|x) = \frac{P(x|\theta) \cdot P(\theta)}{\sum_i P(x|\theta_i) \cdot P(\theta_i)},$$

gdzie θ jest nieznanym parametrem populacji (przyjmującym wartości θ_i), który ma być oszacowany na podstawie wyników obserwacji.

Oczywiście sumowanie w mianowniku rozciąga się na wszystkie możliwe wartości interesującego nas parametru θ_i a x reprezentuje dane (wyniki obserwacji).

Ponieważ w podejściu bayesowskim zakładamy, że interesujący nas parametr jest zmienną losową, to możemy sprecyzować posiadaną informację *a priori* o tym parametrze poprzez przypisanie mu apriorycznego rozkładu prawdopodobieństwa.

Następnie, po otrzymaniu wyników obserwacji (danych), możemy ustalić funkcję wiarygodności, czyli funkcję określającą prawdopodobieństwo otrzymania tych danych przy różnych wartościach parametru, generowaną przez założony aprioryczny rozkład prawdopodobieństwa.

Z kolei ta informacja jest przez twierdzenie Bayesa przekształcona w aposterioryczne prawdopodobieństwo interesującego nas parametru. To aposterioryczne prawdopodobieństwo uwzględnia zarówno wyniki obserwacji (dane), jak i informację *a priori*.

Aposterioryczny rozkład prawdopodobieństwa może być następnie wykorzystany do wnioskowania statystycznego. Z wnioskowaniem może się łączyć wyznaczanie przedziałów ufności dla parametrów, które w przypadku analizy bayesowskiej często bywają nazywane zbiorami wiarygodnymi (ang. *credibile sets*) o założonym z góry prawdopodobieństwie aposteriorycznym.

Pokażemy teraz na przykładzie wykorzystanie twierdzenia Bayesa w przypadku, gdy interesującym nas parametrem jest frakcja elementów określonego rodzaju w badanej populacji.

Załóżmy, że analityk rynku chce określić udział pewnej firmy w rynku, czyli chce oszacować frakcję ludności mieszkającej na danym terenie nabywającej produkt tej firmy,. Niech S będzie parametrem, który ma zostać oszacowany, czyli oznacza udział firmy w rynku w całej populacji na danym obszarze.

Na podstawie poprzednich badań nad podobnymi zagadnieniami oraz informacji o branży analityk sformułował aprioryczne prawdopodobieństwa różnych udziałów firmy w rynku, podane w poniższej tabeli:

S	$P(S)$
0,1 (10%)	0,05
0,2 (20%)	0,15
0,3 (30%)	0,20
0,4 (40%)	0,30
0,5 (50%)	0,20
0,6 (60%)	0,10
-	1,00

Zatem tabela ta zawiera rozkład apriorycznych prawdopodobieństw zmiennej losowej S. Przyporządkowuje ona różnym możliwym wartościom parametru S stopnie przekonania analityka (wyrażone jako prawdopodobieństwa), że parametr przybierze właśnie tę wartość.

Analityk pobiera teraz próbę złożoną z 20 osób i stwierdza, że 4 spośród nich nabywa produkt interesującej go firmy. Analityk chce teraz połączyć za pomocą twierdzenia Bayesa informację o apriorycznym rozkładzie prawdopodobieństwa udziału firmy w rynku z wynikami badań próbnych i dojść do aposteriorycznych prawdopodobieństw udziału w rynku.

Zauważmy, że przy podejściu klasycznym, wszystko, co można byłoby zrobić, sprowadzałoby się do wykorzystania informacji o udziale firmy w rynku zaobserwowanym w próbie

$$\hat{S} = \frac{x}{n} = \frac{4}{20} = 0{,}2 = 20\%$$

do oszacowania tego udziału w całej populacji i ewentualnie do wyznaczenia przedziału ufności dla tego parametru lub sprawdzenia odpowiedniej hipotezy.

Korzystając z twierdzenia Bayesa dla dyskretnych zmiennych losowych analityk aktualizuje swoją informację *a priori*, łącząc ją z wynikami obserwacji pochodzącymi z próby.

Analityk oblicza prawdopodobieństwa warunkowe $P(x|S)$, które tworzą funkcję wiarygodności. Obliczając je, analityk odpowiada na następujące pytania:

Na ile prawdopodobne byłoby uzyskanie takiego wyniku w próbie, jaki uzyskaliśmy (4 „sukcesy” w ciągu 20 doświadczeń), gdyby prawdopodobieństwo sukcesu w jednym doświadczeniu, czyli udział w rynku w całej badanej populacji, było odpowiednio równe 0,1, 0,2, 0,3, 0,4, 0,5 i 0,6?

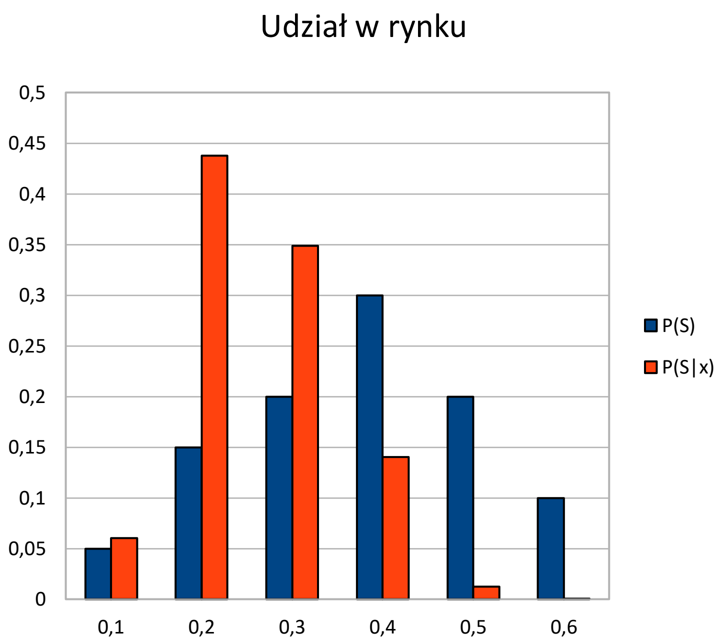
Do obliczenia tych prawdopodobieństw wykorzystujemy oczywiście rozkład dwumianowy ($n = 20$, k=4, $p = 0{,}1$ $(0{,}2;\ 0{,}3;\ 0{,}4;\ 0{,}5;\ 0{,}6$ odpowiednio)).

Analityk może już obliczyć szukane prawdopodobieństwa *a posteriori*. Wyniki tych obliczeń zawiera poniższa tabela:

S	$P(S)$	$P(x\|S)$	$P(S)P(x\|S)$	$P(S\|x)$
0,1	0,05	0,0898	0,00449	0,06005
0,2	0,15	0,2182	0,03273	0,43785
0,3	0,20	0,1304	0,02608	0,34895
0,4	0,30	0,0350	0,01050	0,14043
0,5	0,20	0,0046	0,00092	0,01236
0,6	0,10	0,0003	0,00003	0,00036
-	1,00	-	0,07475	1,00000

Porównując wartości w kolumnie drugiej z wartościami w kolumnie piątej tej tabeli, widzimy jak zmieniły się pierwotne prawdopodobieństwa różnych możliwych udziałów w rynku po uwzględnieniu, za pomocą wzoru Bayesa, informacji pochodzącej z próby.

Wpływ apriorycznych przekonań analityka o udziale firmy w rynku na prawdopodobieństwa aposterioryczne przedstawia także poniższy rysunek:



Widzimy, że prawdopodobieństwa aposterioryczne koncentrują się wokół trzech wartości udziału: 0,2, 0,3, i 0,4. Suma prawdopodobieństw dla tych trzech wartości udziałów jest równa 0,92723.

Te trzy sąsiednie wartości, którym odpowiadają najwyższe aposterioryczne prawdopodobieństwa, tworzą zbiór wiarygodnych wartości szacowanego parametru S, o aposteriorycznym prawdopodobieństwie bliskim standardowo przyjmowanemu poziomowi ufności 0,95. Zatem możemy oczekiwać z pewnością około 95%, że udział firmy w rynku znajdzie się między 0,2 i 0,4. Przy podejściu bayesowskim możemy jednak temu wnioskowi przypisać określone prawdopodobieństwo aposterioryczne, którego obliczenie uwzględnia zarówno informację *a priori*, jak i dane z pobranej próby.

Jedną z większych zalet podejścia bayesowskiego jest możliwość przeprowadzenia analizy w sposób sekwencyjny. Informacje uzyskane w ciągu pewnego badania próbnego można wykorzystać jako informacje *a priori* w następnym badaniu gdy uzyskamy nową informację.

Tą nową informacją mogą być dane uzyskane w wyniku kolejnych badań próbnych, które zostaną powiązane z informacją *a priori* za pomocą twierdzenia Bayesa. Otrzymany stąd rozkład prawdopodobieństw aposteriorycznych znów będzie mógł być wykorzystany jako rozkład prawdopodobieństw apriorycznych, gdy uzyskane zostaną nowe dane, itd. Zobaczmy sposób działania takiej analizy sekwencyjnej kontynuując nasz przykład.

Załóżmy, że analitykowi rynku po zakończeniu badania pierwszej próby udało się zbadać 16 dalszych osób, wśród których znalazło się 3 nabywców interesującego go produktu.

Tą nową informację (pochodzącą z drugiej próby) analityk chce połączyć z wszystkim co wie do tej pory o udziale firmy w rynku.

Zatem ostatni rozkład aposteriorycznych prawdopodobieństw analityk potraktuje jako aktualny rozkład prawdopodobieństw apriorycznych, w stosunku do danych uzyskanych dzięki nowej próbie.

W poniższej tabeli widać, jak ta informacja została przekształcona w nowy rozkład apriorycznych prawdopodobieństw, po uwzględnieniu wyniku nowej próby.

S	$P(S)$	$P(x\|S)$	$P(S)P(x\|S)$	$P(S\|x)$
0,1	0,06005	0,1423	0,0085480	0,049074
0,2	0,43785	0,2463	0,1078384	0,619103
0,3	0,34895	0,1465	0,0511193	0,293477
0,4	0,14043	0,0468	0,0065734	0,037738

0,5 0,6	0,01236 0,00036	0,0085 0,0008	0,0001056 0,0000003	0,000606 0,000002
-	1,00000	-	0,1741850	1,0000000

Zauważmy, że po pobraniu drugiej próby najwyższe aposterioryczne prawdopodobieństwo, wynoszące 0,619103, zostało przypisane udziałowi firmy w rynku $S = 0{,}2$.

Po każdym dodatkowym pobraniu próby prawdopodobieństwa aposterioryczne będą się coraz bardziej koncentrowały wokół wartości udziałów zaobserwowanych w próbie (4/20=0,2; 3/16=0,1875; 7/36=0,1944), a wpływ apriorycznego rozkładu będzie coraz mniej znaczący.

Analiza bayesowska pozwala danym „mówić za siebie" i korygować nasze aprioryczne mniemania, jeśli odbiegają one od rzeczywistości.

W przypadku, gdy dostęp do danych jest ograniczony, to analiza bayesowska kompensuje doraźny brak danych korzystaniem z informacji wcześniejszej (uzyskanej dzięki poprzednim próbom czy też w inny sposób).

Możemy jeszcze zadać sobie pytanie: co stałoby się, gdyby analityk zdecydował się na połączenie wyników obu prób przed połączeniem ich z posiadaną (wyjściową) informacją *a priori*?

Okazuje się, że po posłużeniu się połączoną próbą (o liczebności 20+16=36, w której liczba nabywców produktu wyniosła 4+3=7) i uwzględnieniu informacji *a priori* rozkład prawdopodobieństw aposteriorycznych byłby taki sam jak ostatnio. Wyniki te są przedstawione w poniższej tabeli.

S	$P(S)$	$P(x\|S)$	$P(S)P(x\|S)$	$P(S\|x)$
0,1 0,2 0,3 0,4 0,5 0,6	0,05 0,15 0,20 0,30 0,20 0,10	0,039319 0,165343 0,058784 0,005039 0,000121 0,000001	0,0019659 0,0248014 0,0117568 0,0015118 0,0000243 0,0000001	0,049074 0,619103 0,293477 0,037738 0,000606 0,000002
-	1,00	-	0,0400603	1,000000

Fakt ten pokazuje, że podejście bayesowskie trafnie uwzględnia kolejne porcje informacji. Bez względu na to, jak dana informacja została włączona do modelu, rozkład aposteriorycznych prawdopodobieństw uwzględni całą informację dostępną w danym momencie.

Zauważmy, że w podejściu klasycznym otrzymalibyśmy (wykorzystując całą informację z próby) następujące oszacowanie naszego paremetru S:

$$\hat{S} = \frac{x}{n} = \frac{7}{36} = 0{,}1944 = 19{,}44\%.$$

Twierdzenie Bayesa możemy także rozszerzyć na przypadek ciągłych rozkładów prawdopodobieństwa. Wówczas wartości S mogłyby należeć do całego przedziału $(0, 1)$. Podejście bayesowskie wykorzystywane jest także w wyznaczaniu naiwnego klasyfikatora bayesowskiego w zagadnieniach klasyfikacji oraz wykorzystywane jest w popularnych sieciach bayesowskich.