**Thomas Bayes** (ur. ok. [1702](http://pl.wikipedia.org/wiki/1702) w [Londynie](http://pl.wikipedia.org/wiki/Londyn) — zm. [17 kwietnia](http://pl.wikipedia.org/wiki/17_kwietnia) [1761](http://pl.wikipedia.org/wiki/1761)) [brytyjski](http://pl.wikipedia.org/wiki/Wielka_Brytania) [matematyk](http://pl.wikipedia.org/wiki/Matematyk) i duchowny [prezbiteriański](http://pl.wikipedia.org/wiki/Prezbiterianizm), znany ze sformułowania opublikowanego pośmiertnie [twierdzenia Bayesa](http://pl.wikipedia.org/wiki/Twierdzenie_Bayesa), które to zapoczątkowało dział [statystyki](http://pl.wikipedia.org/wiki/Statystyka).

W jaki sposób szacujemy prawdopodobieństwo, że przykład X należy do klasy C?

Korzystamy z twierdzenia Bayesa, które brzmi następująco: P(C|X) = (P(X|C) \* P(C))/P(X). Gdzie P(C) oznacza prawdopodobieństwo wystąpienia klasy C, czyli prawdopodobieństwo, że dowolny przykład należy do klasy C).

Prawdopodobieństwo P(X|C) oznacza prawdopodobieństwo, że X należy do klasy C, natomiast P(X) oznacza prawdopodobieństwo wystąpienia przykładu X.

Prawdopodobieństwo P(X) jest stałe dla wszystkich klas, zatem klasa Ci, dla której wartość P(Ci|X) jest największa, to klasa Ci, dla której wartość P(X| Ci) \* P(Ci) jest największa. Jeżeli chodzi o prawdopodobieństwo wystąpienia klasy Ci, mamy dwie możliwości. Możemy założyć, w bardzo dużym uproszczeniu, że wystąpienie każdej klasy posiada to samo prawdopodobieństwo. Innymi słowy, możemy przyjąć założenie, że prawdopodobieństwo P(C1) = P(C2) = ... = P(Cm).

W jaki sposób obliczyć P(X|Ci)? Dla dużych zbiorów danych, o dużej liczbie deskryptorów, obliczenie P(X|Ci) będzie operacją bardzo kosztowną. Wymaga ono bowiem oszacowania ogromnej liczby prawdopodobieństw i jest rzędu k^p, gdzie p oznacza zmienne, natomiast k oznacza liczbę wartości tych zmiennych, np. dla p=30 zmiennych binarnych (przyjmujących tylko dwie wartości) musielibyśmy oszacować liczbę prawdopodobieństw rzędu 2^30 czyli około 10^9. Rozwiązaniem tego problemu jest przyjęcie założenie o niezależności atrybutów. (ang. class conditional independance). Przypomnijmy, że mówiliśmy wcześniej, że możemy przyjąć, że wszystkie zmienne są warunkowo niezależne przy danych klasach. Wówczas możemy zastąpić prawdopodobieństwo warunkowe P(X|Ci) iloczynem prawdopodobieństw zgodnie z formułą przedstawioną na slajdzie

Podsumowując prezentację naiwnego klasyfikatora Bayes’a, założenie o niezależności atrybutów znacznie redukuje koszt obliczeń. Dodatkowo, jeżeli założenie jest spełnione, naiwny klasyfikator Bayes’a jest optymalny, tzn. zapewnia najlepszą dokładność klasyfikacji w porównaniu z innymi klasyfikatorami. Założenie to jest bardzo rzadko spełnione w praktyce, jednakże naiwny klasyfikator Bayes’a jest zadziwiająco dokładny w porównaniu z innymi metodami klasyfikacji.