

# Pytania do egzaminu inżynierskiego z Informatyki – Rafał Bańka

## Metody probabilistyczne i statystyka

1. Prawdopodobieństwo warunkowe, prawdopodobieństwo całkowite, wzór Bayesa, niezależność zdarzeń.

Prawdopodobieństwo warunkowe – metoda obliczania prawdopodobieństwa dla zdarzeń A i B należących do tej samej rodziny zdarzeń (przestrzeni zdarzeń), jest to prawdopodobieństwo zdarzenia A pod warunkiem zajścia zdarzenia B.

Prawdopodobieństwo całkowite - metoda obliczania prawdopodobieństwa zdarzenia, uwzględniająca wszystkie możliwe sposoby jego wystąpienia poprzez sumowanie warunkowych prawdopodobieństw wystąpienia tego zdarzenia przy różnych warunkach. W praktyce oznacza to, że możemy obliczyć prawdopodobieństwo zdarzenia A, biorąc pod uwagę jego zależności i warunki.

Twierdzenie Bayesa opisuje prawdopodobieństwo zajścia danego zdarzenia na podstawie wcześniejszej wiedzy o warunkach, które mogą być związane ze zdarzeniem. Wzór Bayesa stosowany jest głównie wtedy, gdy wynik doświadczenia jest znany i zadawane jest pytanie o przebieg tego doświadczenia.

Niezależność zdarzeń w probabilistyce oznacza, że wystąpienie jednego zdarzenia nie wpływa na drugie. Innymi słowy, gdy dwa zdarzenia są niezależne, to prawdopodobieństwo jednego zdarzenia nie zależy od drugiego. Na przykład, rzut monetą i wyrzucenie kostki są niezależnymi zdarzeniami, ponieważ wynik rzutu monetą nie wpływa na wynik wyrzucenia kostki. W takich przypadkach możemy obliczyć prawdopodobieństwo obu zdarzeń, mnożąc prawdopodobieństwa każdego z osobna.

2. Zmienna losowa i jej rozkład, parametry rozkładu zmiennej losowej, przykładowe rozkłady.

Zmienną losową nazywamy jednoznacznie określoną na zbiorze  $\Omega$  zdarzeń elementarnych taką, że każdemu przedziałowi wartości funkcji oznaczanych jako X odpowiada zdarzenie losowe. Zmienna losowa to inaczej wielkość, która może przyjmować wartości liczbowe z pewnego obszaru zmienności, z określonym prawdopodobieństwem.

Parametry rozkładu zmiennej losowej to wartości opisujące rozkład i pozwalające połączyć rachunek prawdopodobieństwa ze statystyką matematyczną.

Przykładowe rozkłady:

Wartość oczekiwana (wartość przeciętna, wartość średnia, nadzieja matematyczna) - jest to jedna z miar określających "środek" zmiennej losowej.

Odchylenie standardowe – najczęściej używana miara rozrzutu zmiennej losowej wokół wartości średniej. W celu jej określenia definiujemy parametr – wariancję (dyspersję).

3. Rozkład normalny i centralne twierdzenie graniczne.

Rozkład normalny inaczej rozkład Gaussa jest podstawowym, teoretycznym rozkładem wartości. Często nazywany jest krzywą dzwonową, ponieważ jego wykres przypomina kształtem dzwon. Jest to teoretyczny, ponieważ rozkład częstości jest w nim wyznaczony na podstawie wzoru, a nie obserwacji rzeczywistych danych. W wielkim skrócie opisuje on sytuacje w świecie, gdzie większość przypadków jest bliska średniemu wynikowi, a im dany wynik bardziej odchyli się od średniej tym jest mniej reprezentowany. Najwięcej jest

przypadków blisko przeciętnej. Im dalej oddalamy się od średniego wyniku, tym przypadków jest mniej.

Centralne Twierdzenie Graniczne jest fundamentalnym twierdzeniem w statystyce, które mówi, że suma lub średnia z dużej liczby niezależnych próbek z innych rozkładów zbliża się do rozkładu normalnego. To pozwala na stosowanie rozkładu normalnego w analizie danych, nawet jeśli pierwotne dane nie są rozkładem normalnym. W praktyce Centralne Twierdzenie Graniczne oznacza, że im większa będzie liczebność naszej badanej próby w pojedynczym badaniu, tym większa jest szansa, że pochodzi ona z rozkładu średnich z cechy o rozkładzie normalnym. Centralne Twierdzenie Graniczne mówi, że jeśli zbierzesz dużo próbek z dowolnego rodzaju danych i zsumujesz je lub obliczysz średnią, to wynik będzie miał kształt rozkładu normalnego.

#### 4. Metoda największej wiarygodności.

Metoda największej wiarygodności (Maximum Likelihood Estimation) to technika używana w statystyce i teorii prawdopodobieństwa do oszacowania parametrów rozkładu prawdopodobieństwa, które najlepiej opisują dane obserwowane. W skrócie, polega na znalezieniu takich parametrów, które sprawiają, że dane są najbardziej "wiarygodne" zgodnie z przyjętym modelem. W praktyce oznacza to znalezienie tych wartości parametrów, dla których obserwowane dane są najbardziej prawdopodobne. Metoda ta jest szeroko stosowana w analizie danych, w modelowaniu statystycznym, oraz w uczeniu maszynowym. Pomaga w oszacowaniu parametrów w różnych rozkładach prawdopodobieństwa, a także w dostosowaniu modeli do rzeczywistych danych.