

STATYSTYKA MATEMATYCZNA

WYKŁAD 1

5 października 2009



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPOJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Model deterministyczny

Model probabilistyczny (stochastyczny)

Model statystyczny

Statystyka:

Statystyka:

statystyka opisowa

Statystyka:

statystyka opisowa

statystyczna analiza danych

Statystyka:

statystyka opisowa

statystyczna analiza danych

wnioskowanie statystyczne

Statystyka opisowa: charakteryzacja *populacji* za pomocą różnych (syntetycznych) wskaźników. Parametry populacji.

Statystyka opisowa: charakteryzacja *populacji* za pomocą różnych (syntetycznych) wskaźników. Parametry populacji. *Populacja* Zbiór elementów etykietowanych. Etykiety jakościowe, liczbowe, wektorowe (wielowymiarowe). Populacje skończone. Populacje nieskończone - rozkład prawdopodobieństwa.

Statystyka opisowa: charakteryzacja *populacji* za pomocą różnych (syntetycznych) wskaźników. Parametry populacji. *Populacja* Zbiór elementów etykietowanych. Etykiety jakościowe, liczbowe, wektorowe (wielowymiarowe). Populacje skończone. Populacje nieskończone - rozkład prawdopodobieństwa. Wskaźniki (parametry): średnia, mediana, kwantyle, rozrzut, symetria, ...

Wskaźniki współzależności cech

Statystyczna analiza danych: formułowanie wniosków o zjawiskach i procesach na podstawie wyników obserwacji tych zjawisk i procesów. Typowe: medycyna kojarzy wskaźniki stanu pacjenta z wystąpieniem określonej choroby lub na podstawie różnych wskaźników stanu zdrowia pacjentów po różnych terapiach ocenia skuteczność tych terapii. Inżynier obserwuje wskaźniki jakości produktów i ocenia jak na tę jakość wpływają różne czynniki; np. jak zawartość węgla w stali wpływa na określone własności konstrukcyjne tej stali.

Statystyczna analiza danych - data mining - statistical data analysis

Wnioskowanie statystyczne (statystyka matematyczna w ścisłym sensie). Na podstawie obserwacji X_1, X_2, \dots, X_n wnioskować o rozkładzie prawdopodobieństwa tych obserwacji, gdy znane są tylko pewne ogólne własności tego rozkładu.

(Populacja, etykieta, losowanie z populacji) – (zmienna losowa, rozkład prawdopodobieństwa)

Przykład: W populacji jest N elementów, w tym nieznaną liczbą M elementów wyróżnionych. Oznaczamy $\theta = M/N$. Z populacji wybieramy losowo n elementów X_1, X_2, \dots, X_n i umawiamy się, że $X_j = 1$, gdy X_j jest jednym z M elementów wyróżnionych oraz $X_j = 0$ w p.p. Na podstawie tej próby oszacować *frakcję* θ . Albo: na podstawie tej próby sprawdzić, czy $\theta < 0.1$.

Zmienna losowa X_j ma rozkład prawdopodobieństwa:

$$P_\theta\{X_j = 1\} = \theta = 1 - P_\theta\{X_j = 0\}$$

Model statystyczny

Przestrzeń próby \mathcal{X} , \mathcal{R} , \mathcal{R}^n .

Rodzina rozkładów prawdopodobieństwa \mathcal{P} , $\{P_\theta, \theta \in \Theta\}$.

Problem statystyki matematycznej:

z jakiej populacji pochodzą obserwacje, które mamy w ręku

Podstawowe twierdzenie statystyki matematycznej

Podstawowe twierdzenie statystyki matematycznej

Dystrybuanta. Dystrybuanta empiryczna

Podstawowe twierdzenie statystyki matematycznej

Dystrybuanta. Dystrybuanta empiryczna

Ogólne twierdzenia o dystrybuancie empirycznej:

SPWL, MPWL: $F_n(x) \rightarrow F(x)$ w danym punkcie x .

Podstawowe twierdzenie statystyki matematycznej

Dystrybuanta. Dystrybuanta empiryczna

Ogólne twierdzenia o dystrybuancie empirycznej:

SPWL, MPWL: $F_n(x) \rightarrow F(x)$ w danym punkcie x .

Gliwienko-Cantelli: $\sup_x |F_n(x) - F(x)| \rightarrow 0$

TWIERDZENIA STATYSTYCZNE

Model statystyczny: $(\mathcal{X}, \mathcal{P})$, $(\mathcal{X}, \{P_\theta, \theta \in \Theta\})$

DUŻO RÓŻNYCH PRAWDOPODOBIENSTW !!!

$$P\{T(X_1, X_2, \dots, X_n) \leq \varepsilon\} \geq \eta \quad \text{dla każdego } P \in \mathcal{P}$$

$$P_\theta\{T(X_1, X_2, \dots, X_n) \leq \varepsilon\} \geq \eta \quad \text{dla każdego } \theta \in \Theta$$

Zadania:

- Estymacja
- Weryfikacja hipotez statystycznych
- Teoria decyzji statystycznych

Przykład: na podstawie próby

oszacować prawdopodobieństwo, że rozważana zmienna losowa przekracza daną wartość (zastosowania: np ubezpieczenia, wielkość rezerw itp)

oszacować wielkość, która będzie przekroczona z zadany (małym) prawdopodobieństwem (np. poziom wody w rzece, V@R)

NASZE NARZĘDZIA

Pakiet R

Excel

Matematyka

Bartoszewicz, J. (1996): Wykłady ze statystyki matematycznej. PWN

Bobrowski, D., Maćkowiak–Łybacka, K. (2006): Wybrane metody wnioskowania statystycznego. Wydawnictwo Politechniki Poznańskiej

Plucińska, A., Pluciński, E. (2000): Rachunek prawdopodobieństwa. Statystyka matematyczna. Procesy stochastyczne. WNT

Silvey, S.D. (1978): Wnioskowanie statystyczne. PWN

Zieliński, R. (1990): Siedem wykładów wprowadzających do statystyki matematycznej. PWN. www.impan.pl/ rziel (aktualizacja 10.III.2005)

Bartoszewicz, J. (1996): Wykłady ze statystyki matematycznej. PWN

Bobrowski, D., Maćkowiak-Łybacka, K. (2006): Wybrane metody wnioskowania statystycznego. Wydawnictwo Politechniki Poznańskiej

Plucińska, A., Pluciński, E. (2000): Rachunek prawdopodobieństwa. Statystyka matematyczna. Procesy stochastyczne. WNT

Silvey, S.D. (1978): Wnioskowanie statystyczne. PWN

Zieliński, R. (1990): Siedem wykładów wprowadzających do statystyki matematycznej. PWN. www.impan.pl/ rziel (aktualizacja 10.III.2005)

Duuuuuuużo !

ZALICZENIA

Własne problemy!

Aktywność

Lista obecności