

Teoretyczne podstawy analizy dużych zbiorów danych

Materiał na kolokwium

1. Testowanie globalnej hipotezy zerowej.

- a) Korekta Bonferroniego. Dowód kontroli błędu pierwszego rodzaju.
- b) Igła w stogu siana - model matematyczny, postać testu ilorazu wiarygodności (z wyprowadzeniem), próg detekcji, optymalność korekty Bonferroniego (bez dowodów).
- c) Test kombinacyjny Fishera. Rozkład przy H_0 (z dowodem).
- d) Test chi-kwadrat. Rozkład przy H_0 (z dowodem).
- e) Matematyczny model dla rozproszonych sygnałów, postać testu ilorazu wiarygodności (z wyprowadzeniem), próg detekcji, optymalność testu chi-kwadrat (bez dowodów).
- f) Test Simesa z dowodem kontroli błędu pierwszego rodzaju.
- g) Model rzadkiej mieszaniny, próg detekcji (bez dowodu).
- h) Test podwyższonego krytycyzmu - definicja, wartość krytyczna, optymalność (bez dowodu).
- i) Optymalność korekty Bonferroniego w modelu "rzadkiej mieszaniny" (bez dowodu).
- j) Porównanie różnych testów dla hipotezy globalnej i odpowiednich modeli matematycznych.

2. Testowanie wielokrotne

- a) Definicja FWER.
- b) Dowód kontroli FWER przez korektę Bonferroniego.
- c) Procedura Holma z dowodem kontroli FWER.
- d) Zasada domknięcia z dowodem.
- e) Procedury Holma i Hochberga jako wyniki zasady domknięcia (bez dowodów).
- f) FDR - definicja, relacja z FWER.
- g) Procedura Benjaminiego Hochberga - kontrola FDR gdy statystyki są niezależne, adaptacja do sytuacji dowolnej zależności (bez dowodów).
- h) Optymalność reguły BH z punktu widzenia optymalizacji ryzyka Bayesowskiego i minimalizacji błędu estymacji - znajomość definicji (ryzyko bayesowskie, ryzyko minimaksowe) i faktów (bez szczegółowego przytaczania założeń).

3. Estymator Jamesa-Steina.

- a) Tożsamość Steina (bez dowodu).
- b) Nieobciążony estymator ryzyka Steina (SURE).
- c) Estymator Jamesa-Steina - dowód, że dominuje on nad estymatorem największej wiarygodności (z wykorzystaniem tożsamości Steina)
- d) Bayesowski estymator średniej w modelu normalnym (z wyprowadzeniem).
- e) Związek estymatora Jamesa-Steina z empirycznym estymatorem bayesowskim (co oznacza "empiryczny estymator bayesowski"?)