Teoria analizy dużych zbiorów - Lista IV

Ryzyko estymacyjne

MM

14 czerwiec 2017

Spis treści

1 Wstęp

W nimniejszym raporcie umieszczone zostały rozwiązania szstej listy zadań z przedmiotu **Teoria analizy dużych zbiorów** prowadzonego przez Panią Profesor Małgorzatę Bogdan we współpracy z Panem Michałem Kosem. Na tejże liście poruszony został problem kontroli FWER, FDR oraz mocy w problemie testowania.

1.1 Definicje

1.1.1 Tabela kontyngencji

Tabela kontyngencji dana jest poprzez:

	zaakceptowana	odrzucona	suma
prawdziwa	U	V	n_0
fałszywa	T	S	n-n_0
suma	n-R	R	n

1.1.2 Familywise error rate

$FWER = P(V \geq 1)$

1.1.3 False discovery proportion

$Fdp = V/R * \mathbb{1}_{R \geq 1}$

1.1.4 False discovery rate

$FDR = E[Fdp]$

1.1.5 Procedury testowania wielokrotnego

1.1.5.1 Bonferroni

Procedura odrzuca daną hipotezę, gdy p-wartośc testu jest mniejsze niż α/n

1.1.5.2 Holm

Posortowane rosnąco p-wartości porównujemy z wyrażeniem $p_{(j+1)} \leq \alpha/(n-j)$. Procedura znajduje pierwszą, najmniejszą, nie spełniającą tej zależności i odrzuca wszystkie hipotezy o mniejszej p-wartośc od znalezionej.

1.1.5.3 Hochberg

Posortowane rosnąco p-wartości porównujemy z wyrażeniem $p_{(j)} > \alpha/(n-j+1)$. Procedura znajduje pierwszą, największą, nie spełniającą tej zależności i odrzuca wszystkie hipotezy o mniejszej p-wartośc od znalezionej.

1.1.5.4 Holm

Posortowane rosnąco p-wartości porównujemy z wyrażeniem $p_{(i)} \leq \alpha * i/n$. Procedura znajduje największą spełniającą tą zależność i odrzuca wszystkie hipotezy o mniejszej lub równej p-wartośc od znalezionej.

2 Zadanie 1

W zadaniu należało porównać FWER, FDR oraz moc (stosunek wykrytych hipotez alternatywnych do wszystkich hipotez alternatywnych) w następującym problemie "niskowymiarowym".

Niech $p = 20$, dla $i = 1, ..., 10$ $\mu_i = \sqrt{2\ln(20/i)}$, dla pozostałych $\mu_i = 10$.

Charakterystki podane powyżej porównujemy dla następujących procedur:

- Korekta Bonferonniego
- Procedura Holma
- Procedura Hochberga
- Procedura Benjaminiego-Hochberga.

Oto otrzymane wyniki:

Metoda	FDR	FWER	Moc
benj	0.0258	0.0700	0.1646
bonf	0.0183	0.0280	0.1084
hoch	0.0197	0.0320	0.1108
holm	0.0190	0.0300	0.1106

Tablica 2: Kontrola FDR, FWER i Mocy

3 Zadanie 2

Analogicznie jak w zadaniu pierwszym, w zadaniu drugim należało porównać FWER, FDR oraz moc. Tym razem rozpatrujemy następujące problemy "wysokowymiarowe".

- A. $\mu_1 = 1.2\sqrt{2\ln(p)}, \mu_1 = ... = \mu_p = 0$,
- B. $\mu_1 = ... = \mu_{1000} = 0.15\sqrt{2\ln(p)}, \mu_{1001} = ... = \mu_p = 0$,
- C. $\mu_1 = ... = \mu_{100} = 2, \mu_{101} = ... = \mu_p = 0$,
- D. $\mu_1 = ... = \mu_{100} = \sqrt{2\ln(p)}, \mu_{101} = ... = \mu_p = 0$.

Przyjmujemy $p = 5000$.

Pozostała część zadania jak w zadaniu 1. Z racji złożoności obliczeniowej do kalkulacji posłużono się pakietem **doParallel**.

Otrzymane wyniki:

Metoda	FDR	FWER	Moc
benj	0.0567	0.0920	0.7200
bonf	0.0360	0.0500	0.7060
hoch	0.0360	0.0500	0.7060
holm	0.0360	0.0500	0.7060

Tablica 3: Problem A

Metoda	FDR	FWER	Moc
benj	0.0360	0.0420	0.0001
bonf	0.0310	0.0340	0.0001
hoch	0.0310	0.0340	0.0001
holm	0.0310	0.0340	0.0001

Tablica 4: Problem B

Metoda	FDR	FWER	Moc
benj	0.0565	0.1460	0.0155
bonf	0.0327	0.0480	0.0080
hoch	0.0327	0.0480	0.0080
holm	0.0327	0.0480	0.0080

Tablica 5: Problem C

Metoda	FDR	FWER	Moc
benj	0.0477	0.9820	0.7835
bonf	0.0017	0.0660	0.3867
hoch	0.0017	0.0660	0.3873
holm	0.0017	0.0660	0.3873

Tablica 6: Problem D

Widoczne jest inne zachowanie procedur dla każdego z problemów. Bardzo ciekawe jest to, że dla każdego z nich nie było różnic pomiedzy metodami Bonferroniego, Hochberga i Holmesa.

Zastanawiające jest to, ze wyniki FDR dla BH wcale nie są lepsze niż dla pozostałych procedur, a to BH kontoluje FDR na poziomie q.

Ponadto, dla każdego problemu wyniki dla trzech ostatnich procedur są takie same, a pokazano, na wykładzie, że procedury Holma i Hochberga kontrolują FWER na tym samym poziomie, ale procedura Hochberga jest mocniejsza.

Nie doszukano się błędów programistycznych.