

Oponentský posudok

Názov bakalárskej práce: Mnohonásobné testovanie
Autor: Matěj Kadavý

Študent Matěj Kadavý sa vo svojej bakalárskej práci venuje problému merania chyby v prípade chybných zamietnutí nulových hypotéz pri mnohonásobnom testovaní. Vo svojej práci popísal dva základné a najčastejšie používané postupy k meraniu tejto chyby a to tzv. False Discovery Rate metódu (FDR) a Familywise Error Rate metódu (FWE).

Bakalárska práca poskytuje ucelený a logický prehľad danej problematiky, kvalitatívne opísané oba spomenuté prístupy ale aj základné algoritmy najčastejšie používané v prípade metódy FDR, alebo FWE. Nechýbaju ani dôležité porovnania medzi jednotlivými metódami a popis základných vzťahov medzi nimi.

Druhá časť práce sa venuje simulačnej štúdii, kde sú popísané algoritmy a ich základné vlastnosti testované na simulovaných datach. Prehľadne sú pritom simulácie rozdelené do dvoch kategórií, podľa predpokladu závislosti testových štatistík. V súčasnej dobe by možno bolo vhodné šie venovať väčšiu pozornosť práve prípadu so závislými testovými štatistikami - v práci je tento prípad spomenutý podľa môjho názoru dosť okrajovo a podstatná časť je venovaná práve prípadu s nezávislými testovými štatistikami.

K práci mám len niekoľko drobných pripomienok:

- V niektorých prípadoch dochádza v texte k použitiu značenia, ktoré nie je vysvetlené. Logický je síce jasné, čo chcel autor povedať, prípadne sa dá toto značenie nájsť o niekoľko stránok neskôr, malo by to byť ale formulované ako súvislý text. Napr. str.6 - čo je to m vo vzorci? Podobný problém vidím aj vo formulácii niektorých viet. - napr. veta 2.5.1. Veta by mala mať charakter samostatného a logicky uceleného textu, vrátane všetkých potrebných predpokladov.
- V niektorých prípadoch mi chýba označenie ukončenia dôkazu (napr. Lemma 3.2.1, Lemma 3.2.3). Taktiež si myslím, že by bolo dobré uvádzať alespoň odkaz na literatúru s dôkazom spomínaného tvrdenia (Lemma 3.2.2 - bez odkazu na dôkaz).
- str.8 - vo vzťahu (2.4) sa odkazuje na (2.2) a (2.3), pak správne by ale malo byť: $T_0 = t_{n-k}(p/2)$. Stejně aj o dva riadky vyššie.
- str.10 - Jednokroková Bonferroniho metóda - prepočítané p-hodnoty majú byť značené ako $p_{(i)}^{bonf}$ a sú definované ako minimum z dvojprvkovej množiny $\{mp_i, 1\}$ pre $i = 1, \dots, m$.
- str.19 - v poslednom odseku pred 2.12, je potrebné uviesť, aký je vzťah medzi i a j , ináč porovnania hodnôt $p_{(i)}^{bonf}, p_{(j)}^{bonf}$ a $p_{(i)}^{hoch}, p_{(j)}^{hoch}$ nedávajú zmysel.
- K obrázkom 2.1 a 2.2 by sa skôr hodil iný popis pre os x.
- Doporučil by som tiež podať trochu obsiahnejší popis k niektorým tabuľkám v simulačnej štúdii (napr. tabuľky 4.3, 4.4, 4.6). Prispelo by to k lepšiemu pochopeniu výsledkov, bez nutnosti čítať celý text v tejto sekcii. Tiež by som uvítal komplexnejšie zrovnanie výsledkov pre nezávislé a závislé testové štatistiky, než je len jediná veta referujúca k niektorým tabuľkám (str. 35 - posledný odsek).

• V zozname literatúry som našiel niektoré publikácie, na ktoré sa autor v texte neodkazuje. Taktiež si myslím, že je pomerne zbytočne označovať každú rovnosť a vzorec v texte, ak nedochádza k odkazovaniu sa naň. Pôsobí to potom práve opačne a text sa stáva menej prehľadným.

Niektoré drobné preklepy v texte:

- správna adresa na Dr. Hlávkku je: hlavka@karlin.mff.cuni.cz
- str.11, veta 2.5.1: "pro libovolný počet testových statistik..."
- str.11, indukčný krok pre $m' < m$ a teda ukážeme platnosť pre $m' + 1$ nie $m + 1$!
- str.12, veta pred (2.19) - doporučil by som inú formuláciu, pretože (2.15) je vlastne definícia j_0 a nie nerovnosť, jež je v tejto definícii obsažená!
- str.15 - druhý riadok - odkaz na obrázok 2.1, nie obrázok 2.2!
- str.23 - podnadpis b) - poněkud nelogická formulácia
- str.28-29 - nie celkom súhlasím s formuláciou oboch otázok. Simulácie opakujete 100 000 krát, ale nepočítate v kolkých *opakovaniach* nastáva sledovaný jav, ale v kolkých *pokusech*, resp. *prípadoch* nastáva sledovaný jav.
- str.29 "...ve 100 případech dostaneme právě 50 (nie 5!) p-hodnot menších než α ."
- vo veľmi málo prípadoch je slovo na konci riadku nesprávne rozdelené...
- Literatúra - v prípade viacerých autorov by sa hodilo oddeliť ich mená čiarkou (napr. *Benjamin Y., Hochberg Y., atd.*).

Aj napriek týmto drobným pripomienkám považujem ale bakalársku prácu študenta Matěja Kadavého za veľmi pekne a kvalitne vypracovanú a myslím si, že jednoznačne splňuje nároky a požiadavky na dobrú bakalársku prácu.

Celkové hodnotenie: výborne

Mgr. Matúš Maciak
doktorand KPMŠ MFF UK
20. August 2007