

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Funkcionální ANOVA

Autor: Viktor Dolník

SHRnutí OBSAHU PRÁCE

Autor v práci podrobne popisuje odvodenie obálkových testov pre funkcionálne dáta, a ich aplikáciu na problém funkcionálnej analýzy rozptylu. Následne sú tieto testy použité na analýzu reálnych dát, a porovnané v krátkej simulačnej štúdiu.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Téma práce je zaujímavá a aktuálna, ale tiež pomerne náročná. Autor prehľadne spracoval informácie z niekoľkých nedávnych odborných článkov. Zadanie práce bolo splnené.

Vlastní příspěvek. Jedná sa o prácu kompilačného charakteru. V časti 4.2 autor aplikoval popísané testy na reálne dáta, a v časti 4.3 replikoval simulačnú štúdiu z literatúry.

Matematická úroveň. Matematická úroveň práce je uspokojivá. Rigoróznosť textu, definícií a značenia by sa miestami dala zlepšiť, práca ale neobsahuje žiadne zásadné chyby.

Práce se zdroji. Zdroje sú citované primerane a správne.

Formální úprava. Práca obsahuje iba drobné formálne nedostatky primerané jej rozsahu.

PŘIPOMÍNKY A OTÁZKY

1. str. 10, r. -3: výrok “if H_0 is not true, V_N tends to infinity for $N \rightarrow \infty$ ” nie je všeobecne pravda. Neuvažujeme totiž iba spojité funkcie (so spojitými strednými hodnotami).
2. Nikde v texte nie je explicitne stanovené, akým spôsobom sa z usporiadania \prec vo všeobecnom prípade určuje poradie.
3. str. 16: podľa Definition 3.4 môžeme definovať \prec ako $f \prec g$ pre každé f a g . Pre takúto reláciu triviálne platí podmienka (3.1) v Theorem 3.2. Poradie všetkých funkcií bude ale zrejme skoro iste konštanta (ktorá závisí na presnej definícii poradia, viď predchádzajúci bod). Platí teda Theorem 3.2?

4. sekcia 4.1: popis funkcionálneho ANOVA testu sa mi nezdá byť úplný. V algoritme v časti 4.1.2 je síce popísaný výpočet testovej štatistiky T , nikde ale nevidím popis konštrukcie kritického oboru testu, alebo jeho p -hodnoty. Tie by sa podľa všetkého mali vzťahovať k testu z časti 3.2, spätná referencia tu ale chýba.
5. sekcia 4.2: p -hodnoty testu v uvedenom príklade vychádzajú medzi 0.000 a 0.085. Rozhodnutie o zamietnutí nulovej hypotézy na základe priemeru týchto dvoch hodnôt sa mi nezdá byť dostatočne zdôvodnené. Prinaajmenšom by si tento výsledok zaslúžil nejakú diskusiu.
6. str. 26 a 27: výsledky simulačnej štúdie sú pomerne zvláštne — napríklad v tabuľke 4.1, model M1, sa takmer všetky výsledky pohybujú výlučne blízko 0, alebo 0.25 (predpokladám, že hladina testu mala byť $\alpha = 0.05$ — v popise simulačnej štúdie v časti 4.3 však hodnota parametru α nie je uvedená). Ako by sa tieto hodnoty dali vysvetliť? Celkove, výsledky v tabuľkách 4.1 a 4.2 by si zaslúžili lepšiu interpretáciu.
7. Voľba vysokých hodnôt σ v simulačnej štúdii sa zdá byť prehnaná. Napríklad, v modele M4 sú stredné hodnoty skupín 1.02, 1.04 a 1.06. K týmto stredným hodnotám sa pripočítajú nezávislé normálne rozdelené náhodné veličiny s rozptylom až 25. Čo nám takáto simulácia dáva?
8. Ako sa budú testy správať pre počet diskretizovaných bodov funkcií idúci do nekonečna?

Nasledujú niektoré ďalšie, menej závažné pripomienky a pozorovania.

1. Oceňujem dobrú úroveň anglického jazyka práce. Pozor však na chyby ako “nominator” (str. 10) alebo “estimator/estimate” (str. 22 a inde). Prvé vety dôkazov by mali začínať veľkými písmenami. Odkazy na literatúru menom autorov v texte by mali byť bez zátvoriek, napr. Cuevas et al. [2004], nie [Cuevas et al., 2004].
2. Časť značenia používaná v práci nie je definovaná (výrazy typu $\bar{X}_{+,+}$ pre funkcie na str. 10, alebo $R_1(t)$ na str. 14). Časť značenia sa v rôznych kapitolách mení (n a k v sekcii 1 sú označené ako N a K inde; rozdelenie náhodných veličín je v sekcii 1 označované ako F_i , ďalej ako \mathbb{P}_{X_i} ; norma funkcií na str. 10 je označená ako $\|\cdot\|$ aj $\|\cdot\|_2$). Riadky vo výrazoch by nemali končiť symbolmi $=$, $<$, $+$, atď. Tieto symboly sa typicky uvádzajú iba na začiatku nasledujúceho riadku.
3. str. 5, Theorem 1.3: vo výrazoch (1.1)–(1.3) je nesprávne uvedený menovateľ vo výraze pre strednú hodnotu.
4. str. 9, 23, 24 a inde: kvalita obrázkov by mohla byť omnoho lepšia. Tiež, v popiskoch obrázkov by malo byť v každom prípade jasne uvedené čo je na obrázku znázornené. Napríklad, z popiskov nie je vôbec jasné čo je vykreslené na Figure 4.1 a 4.2.

5. str. 10: σ^2 ako je definované tu je náhodná veličina, nie celkový rozptyl.
6. str. 10, r. -3: Cuevas et al. (2004) netvrdia, že V_N konverguje k rovnakému rozdeleniu ako čitateľ F_N .
7. str. 13, Definition 3.1: vo výraze pre hornú obálku by malo byť $X_{upp}^{(k)}(t) = X(t)_{(s-k+1)}$, nie $X_{upp}^{(k)}(t) = X(t)_{(s-k)}$.
8. str. 15, dôkaz Theorem 3.1: vo výraze po $\mathbb{P}(H_0 \text{ is rejected})$ by malo byť $\exists t$ namiesto $\forall t$. Tiež, druhý riadok rovnice sa v dôkaze nevyužíva.
9. str. 24, modely (M1)–(M4): rozsahy indexov i a j sú obrátené.
10. str. 26 a 27: nefunkčné odkazy na tabuľky 4.1 a 4.2.
11. Článok citovaný ako Cuevas (2004) je v skutočnosti Cuevas, Febrero a Fraiman (2004).

ZÁVĚR

Napriek väčšiemu množstvu pripomienok prácu považujem za dobrú, a odporúčam ju uznať ako bakalársku prácu.

Stanislav Nagy
KPMS MFF UK
05.06.2018