

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Kalibrační odhady ve výběrových šetřeních

Autor: Petr Klička

SHRNUTÍ OBSAHU PRÁCE

Práce sa zameriava na odhadovanie populačného úhrnu vo výberových šetreniach v prípade, že máme k dispozícii pomocné informácie. V prvej časti autor popisuje odvodenie tzv. kalibračných odhadov úhrnu, a diskutuje podmienky ich asymptotickej normality. V druhej časti odvodené vzťahy porovnáva na reálnych dátach RADIOPROJEKT poskytnutých agentúrami na marketingové výskumy.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Téma práce je vhodná a zaujímavá. Aplikčná časť sa ale zdá byť významne obmedzená povahou uvažovaných reálnych dát. Zadanie práce bolo splnené.

Vlastný príspevek. Vlastný príspevek autora spočíva v praktickej časti, ktorej úroveň ale dokážem zhodnotiť iba čiastočne. V práci sa totiž uvádza iba minimálny popis dát¹, krátka diskusia o použitých metódach, a rutinný výpis výsledkov. K práci nie je priložený použitý dátový súbor, zdrojové kódy implementácií, ani úplný popis dát a vstupov štúdie.

Matematická úroveň. Teoretická časť práce obsahuje niekoľko originálnych rigorózných odvodení. Tejto časti by prospel systematickejší prístup. Technická úroveň textu je ale dobrá, a nevidím žiadne závažné nedostatky.

Práce se zdroji. Použitá literatúra je citovaná primerane a správne.

Formální úprava. Text obsahuje preklepy a drobné chyby, celková úprava je ale uspokojivá.

PŘIPOMÍNKY A OTÁZKY

1. Štýl práce je monotónny a repetitívny. Autor sa zdržiava diskusie, príkladov, a vlastných náhľadov. Namiesto toho zdĺhavo opakuje sady predpokladov, detailne popisuje jednoduché úpravy rovníc, a často zdôrazňuje informácie zrejmé z kontextu.

¹čo by sa čiastočne dalo vysvetliť dôvernosťou poskytnutých dát. V tom prípade by tento problém ale mal byť v práci diskutovaný.

2. V teoretickej časti práce v kapitolách 2 a 3 je formulovaných viacero predpokladov, ktoré zaručujú dobré vlastnosti uvedených odhadov ((a)–(c) na str. 8, (A) na str. 10, (I)–(IV) na str. 16). Okrem toho sa ale na rôznych miestach textu vyskytuje rada ďalších predpokladov, ktoré sú rozumné, ale nezriedka plynú z ostatných. Teoretickej časti práce by prospel systematickejší prístup, t.j. v tomto poradí

- jasne popísať model a všetko jeho predpoklady;
- dokázať tvrdenia;
- diskutovať nakoľko sú predpoklady silné a ako by sa dali oslabiť.

Uvediem napríklad:

str. 21, posledný odstavec: aby bol porušený predpoklad (IV), nie je nutné aby $\sigma_{zN}^2 = 0$. Z predpokladu na str. 20 ale vieme, že $\lim_{N \rightarrow \infty} \sigma_{zN}^2$ existuje a je kladná, a z toho už plynie platnosť (IV).

V teoretickej časti tiež veľmi chýba diskusia — výklad sa obmedzuje na technické odvodenia vzťahov, často bez ďalších komentárov.

3. Vo vzťahu na str. 16 autor používa pre vektory, ktoré označíme $\boldsymbol{\beta}$ a \mathbf{X} , nerovnosť $|\boldsymbol{\beta}^\top \mathbf{X}| \leq |\boldsymbol{\beta}^\top| \sum_{j=1}^J |\mathbf{X}^j|$, kde \mathbf{X}^j je j -ty prvok vektoru \mathbf{X} . O akú nerovnosť sa jedná? Čo znamená absolútna hodnota vektoru na pravej strane nerovnosti?
4. Z predpokladu $\limsup_{N \rightarrow \infty} n_N/N \in (0, 1)$ ešte neplynie existencia limity tohto výrazu, ako je tvrdené na str. 39.
5. V práci je niekoľkokrát zmieneny algoritmus *iterative proportional fitting* (najmä str. 37). Nikde v texte však nie je vysvetlené ako tento algoritmus funguje, alebo aspoň čo je jeho vstupom a výstupom.
6. Praktická časť práce (kapitoly 5 a 6) nie je dostatočne popísaná. Práca nemá prílohu. Okrem prezentovaných tabuliek výsledkov teda nemám možnosť posúdiť vhodnosť implementácie uvažovaných metód. Popis dátového súboru je neadekvátny. Napríklad, iba vo vzorci pre c na str. 39 sa čitateľ dozvie, že uvažujeme $N = 8\,795\,000$. Nikde v práci nie je spomenuté aké konkrétne pomocné informácie sú využité pri konštrukcii odhadov, ani to, ktoré premenné nie sú v porovnaní s originálnym prieskumom k dispozícii (pozn. 11, str. 39).
7. Popis záverov praktickej časti práce uvedený v kapitolách 5 a 6 nie je príliš zaujímavý. Obmedzuje sa na rutinné porovnanie hodnôt uvedených v tabuľkách, bez ďalších náhľadov a zdôvodnení.
8. Z obr. 5.1 sa zdá, že sady váh w_k a g_k , $k \in s$, môžu byť podobné. Nebolo by vhodnejšie porovnávať dvojice hodnôt (w_k, g_k) priamo pre každé $k \in s$ (napr. využiť Q-Q plot)?

9. V simulačnej štúdií (kapitola 6) vidíme, že (bodový) odhad metódou M dáva vždy najlepší výsledok. Platí podobné tvrdenie všeobecne? T.j., za platnosti predpokladov 0.-2. na str. 44 a 45, je bodový odhad metódou M (v nejakom zmysle) optimálnym odhadom populačného úhrnu?

Nasleduje zoznam niektorých menších chýb na ktoré som narazil:

1. str. 10, v znení Lemmy 1 nemá byť „Pro k -tou jednotku“; str. 10, r. -2 $v_{kN} = y_k$ má byť $v_{kN} = y_{kN}$ a str. 11 po (2.13) obdobne; str. 11 v Pozn. 4 po „Chceme ukázať, že“ chyba \mathcal{P} nad \rightarrow ; str. 13 v (3.2) je vpravo \in namiesto $=$; str. 14 v (3.10) a inde nie je vhodné indexovať cez k viackrát v jednom výraze; str. 15 a inde: v trojuholníkovej schéme s rozdielnymi hodnotami N by mali aj rozsah výberu s a rozptyl σ_k^2 záležať na N . Ďalej neplatí $U_N = \{y_{1N}, \dots, y_{NN}\}$, ale $U_N = \{1, \dots, N\}$; str. 20, po (3.22): x_{kjN} nemá byť transponované, nejde o vektor; str. 21 v strede strany „populačného průměru \mathbf{X}_N “ by malo byť „populačného úhrnu“ a rovnako pre Z_N ; str. 21: v rovnici dole chyba v menovateli $\max_{j=1, \dots, J}$; str. 22, po (3.24) výraz $\hat{\beta}_{sN}$ by mal byť $\hat{\beta}_{sN}$; str. 26, po (3.31) nie je uvedená dĺžka vektoru λ ; str. 28 hore vo výraze pre \hat{Y}_{kal} chyba suma; str. 40 hore: „populační úhrn X “ namiesto \mathbf{X} ; str. 40 a 41 vo výrazech (5.7) a (5.8) je použité nesprávne znamienko na pravej strane; str. 45 bod 2 algoritmu má byť „nevážený odhad P populačného úhrnu“ namiesto „nevážený populační úhrn“.
2. str. 2 „za pomoci výběrových šetření“; str. 3 veta začínajúca „Odhad parametru“ nie je zrozumiteľná; str. 11 „větu o spojitě transformace“; str. 28 hore „druhé derivace podle,“; str. 32 „věkové strukturu“; str. 39 po (5.5) má byť bodka.

ZÁVĚR

Práca je pomerne dobre technicky zvládnutá. Určite by sa ale dal zlepšiť jej štýl, a spôsob prezentácie. Odporúčam ju uznať ako diplomovú prácu.

Stanislav Nagy
KPMS MFF UK
14. augusta 2018