

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Testy Poissonova rozdělení

Autor: Filip Trusina

SHRNUTÍ OBSAHU PRÁCE

V práci je, na základě simulačnej štúdie, porovnaných niekoľko známych testov hypotézy o tom, že náhodný výber pochádza z Poissonovho rozdelenia proti všeobecnej alternatíve.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Téma práce je jednoduchá, určite má však radu aplikácií. Okrem simulačnej štúdie bol súčasťou zadania aj podrobný popis testov používaných v literatúre. Táto časť zadania bola splnená iba vo veľmi obmedzenom rozsahu.

Vlastní příspěvek. Vlastný príspevok autora spočíva v simulačnej štúdiu. Práca však prakticky neobsahuje teoretickú časť. Okrem triviálnej lemy 1 sa popis všetkých testov obmedzuje na uvedenie testovej štatistiky a jej kritického oboru známych z literatúry, s minimom ďalších komentárov. Napríklad pri uvedení testovej štatistiky v sekcii 4.1 by určite mala byť diskutovaná voľba odhadov $\hat{p}(k)$ a $\hat{F}(k)$ a ich zdôvodnenie. Pre jednoduchšie testy by bolo určite prínosné aspoň v náznaku dokázať platnosť tvrdení na ktoré sa autor odkazuje, prípadne detailne interpretovať význam testových štatistík, a popísať rozdiely medzi testami.

Matematická úroveň. Vzhľadom k minimu vlastných odvodení autora matematickú úroveň práce nedokážem posúdiť.

Práce se zdroji. Zdroje sú citované správne.

Formální úprava. Text obsahuje niekoľko preklepov a chýb v značení. Celková formálna úroveň je ale dobrá.

PŘIPOMÍNKY A OTÁZKY

1. Interpretácia výsledkov simulačnej štúdie je pomerne slabá. To je najlepšie vidieť v sekcii 5.3, ktorá sa obmedzuje na nezaujímavé porovnanie čísel v tabuľke 6. Omnoho cennejšie by boli interpretácie a vlastné zdôvodnenia výsledkov prezentovaných v tabuľkách.
2. Práca je rozdelená na dve časti — v kapitolách 2 a 3 sú porovnané asymptotické testy, a v kapitolách 4 a 5 testy založené na princípe bootstrapu. Chýba však akékoľvek porovnanie asymptotických testov s bootstrapovými. Rozdelenie testových štatistík z kapitoly 2 mohlo byť aproximované aj pomocou metódy bootstrap, a tieto dva prístupy mohli byť porovnané. Takto práca pôsobí neucelene.
3. Ako z (2.3) plynie približná normalita veličiny N ako je tvrdené na str. 6?
4. Ako sa využíva linearita strednej hodnoty pre odvodenie strednej hodnoty zmesi rozdelení (str. 9)?

5. V tabuľkách 3 a 4 sú porovnané priemery 100 odhadov z nezávisle generovaných náhodných výberov a skutočné hodnoty parametrov rozdelenia. To dáva dobrú predstavu o vychýlení odhadov. Z tabuliek ale nevidíme nič o rozptyle odhadov, alebo ich (približnom) rozdelení. Ak chceme niečo usudzovať o kvalite odhadov, tabuľky by mali obsahovať aspoň informáciu o ich smerodajnej odchýlke.
6. Čo je mienené „najsilnejším testom“ na str. 21?
7. Aký rozsah výberu n je volený v obrázkoch 2–5? Ako by sa tieto silofunkcie menili v závislosti na n ? Vo všetkých týchto obrázkoch by sa hodilo vyznačiť aj hladinu testu 0.05.
8. Ako by sa dalo vysvetliť zvláštne správanie silofunkcie pre T_{An} v obrázku 3?
9. Integrály v definíciách 2–4 na str. 3 nemusia existovať. V definícii 4 nie je ošetrené prípadné delenie nulou.
10. Na str. 4, ale aj ďalej v texte, sa nepresne píše „zvolíme náhodné číslo z množiny“, pričom sa myslí náhodné číslo s rovnomerným rozdelením na množine. Podobne napr. pri popise princípu bootstrapu na str. 13 sa píše o generovaní náhodných výberov, nie je už ale uvedené že tieto výbery musia byť nezávislé. Hovorí sa *rozsah náhodného výberu*, nie veľkosť. Výraz „asymptotičnosť rozdelenia“ je zvláštny.
11. Vo funkcii `test2` v prílohe práce je testová štatistika FD počítaná s $\sqrt{2}$ v menovateli na rozdiel od formuly (4.6). Premenná `poz2` v prílohe zrejme označuje konštantu zo sekcie 5.1 práce. Tam je táto konštanta volená ako $2(\lceil \bar{X} \rceil + 5)$. (Napríklad) na riadku 242 prílohy však vidíme `poz2 = int(prum //2*2 + 5) * 2`. Jedná sa o rovnakú konštantu?
12. Veta ktorá začína „Abychom získali pravdepodobnosť...“ na str. 10 je nezrozumiteľná. Rovnako veta ktorá začína „Musíme si uvědomit...“ na str. 11. Čo je to „krok Poissonova rozdění“ v popisku tabuľky 4?
13. Na str. 8 má byť v predpise pre binomické rozdelenie $k = 0, 1, \dots, m$. V ďalšej formulii chýba bodka. Ďalšie bodky ukončujúce vety chýbajú po popiskoch tabuliek a obrázkov, a po poznámkach pod čiarou. Niektoré poznámky začínajú veľkým písmenom, niektoré nie. Kritické obory testov sú občas uvádzané s ostrou (sekcia 2.4) a občas s neostrou (sekcia 2.3) nerovnosťou.
14. V prvej formulii dôkazu lemy 1 chýbajú indexy sumy; priemer je vo formulii (2.2) na str. 5 označený inak ako v (2.1); „Pro pro“ na str. 8; „asymptotický rozděléním“ na str. 10; predpis pre $\hat{F}(k)$ nie je v sekcii 4.1 uvedený; písmo pre k v tabuľke 3; „za pomocí“ na str. 14; $\sum_0^{i=14}$ na str. 17.

ZÁVĚR

Predložená práca sa radí skôr k slabším. Jej hlavným prínosom je jednoduchšia simulačná štúdia na ktorej sú porovnané testy prebraté z literatúry. Myslím však, že je možné ju uznať ako bakalársku prácu.

Stanislav Nagy
 KPMS MFF UK
 3. júna 2019