Posudek oponenta bakalářské práce

Název: Oprava na spojitost
Autor: Marek Štěpán

Shrnutí obsahu práce

Marek Štěpán sa vo svojej bakalárskej práci venuje problému aproximácie rozdelenia konečného súčtu diskrétných náhodných veličín rozdelením spojitým, špeciálne využitiu tzv. korekcie na spojitosť, ktorá v určitých konkrétných prípadoch zaručuje lepšie empirické výsledky.

V prvej časti práce je teoreticky popísaný samotný problém aproximácie po častiach konštantnej distribučnej funkcie diskrétnej náhodnej veličiny distribučnou funkciou spojitou a analyticky je odvodená tzv. oprava na spojitosť, ktorá je následne aplikovaná na prípad binomického rozdelenia.

Užitočnosť takejto modifikácie pri konštrukcii intervalov spoľahlivosti autor odôvodňuje jednak teoreticky a tiež ilustruje na konkrétnych numerických príkladoch.

V druhej časti práce je analogická úprava na spojitosť aplikována na problém testovania hypotéz, konkrétne na test nezávislosti v kontingenčnej tabuľke. Autor uvádza teoretické základy χ^2 testu nezávislosti, prestavuje modifikáciu testu s opravou na spojitosť a empiricky, pomocou simulačnej štúdie, porovnáva dosiahnutú hladinu testu a skutočnú sílu χ^2 testov s opravou a bez opravy na spojitosť.

Z odborného hľadiska je téma práce pomerne jednoduchá, ale spracovaná je vcelku prehľadne, súvislo a s použitím viacmenej korektne formulovaného matematického textu. Z matematického aj formálneho hľadiska považujem prácu za dostatočne kvalitnú a vyčítam jej len niektoré menšie nedostatky, ako napr. nezavedené, alebo nepresné značenie, prípadne nedostatočný popis niektorých výsledkov. Prácu ale jednoznačne doporučujem odbornej komisii uznať ako bakalársku prácu.

Otázky & pripomienky k obhajobe

- V niektorých prípadoch chýba zavedené značenie, dostatočne podrobný popis, prípadne sú použité formulácie nepresné. Názorný príklad ku každemu bodu: Asi by malo byť uvedené na str. 3, že $\Phi(z)$ je distribučná funkcia normálneho normovaného rozdelenia (definovať pri prvom použití, nie až na str.4). Popisok k Obrázku 2.1. by zase mal obsahovať informáciu o tom, o aké binomické rozdelenie sa jedná (parametre rozdelenia). Podobne, Lévy-Lindebergova centrálna limitná veta dáva tvar aproximace distribučnej funkcie pre výrazne väčšiu triedu rozdělení, než iba binomické, ako plynie z trochu nešťastnej formulácie prvej vety v Sekcii 2.2. Podobných nedostatkov sa v práci objavuje niekoľko.

- Autor sa venuje aproximácii distribučnej funkcie diskrétnej veličiny, ktorá nabýva maximálne spočetne mnoho hodnôt. Samotná distribučná funkcia je ale definovaná pre ľubovolné $x \in \mathbb{R}$. Niektoré formulácie v práci pritom navodzujú pocit, že distribučná funkcia je definovaná pouze pro $x \in \mathbb{N}_0$. Rozloženie problému na situáciu kedy $x \in \mathbb{N}_0$ a $x \in \mathbb{R}$ (Sekcia 2.4.1) preto nepovažujem za najvhodnejšie riešenie.

- V náväznosti na predchádzajúcu poznámku považujem voľbu hodnôt $s \in \{-3, -, -1, 0, -1, 2, 3\}$ v Sekcii 2.2.3 za nedostatočnú a v tomto zmysle aj samotné závery numerickej ilustrácie, prezentované na str.10, nie sú úplne správne. Konkrétne tvrdenie, že "pro velmi nízké n už není (aproximace s opravou na spojitosť) obecně lepší než aproximace bez opravy na spojitost", závisí samozrejme od konkrétnej voľby hodnoty x (resp. hodnoty s). Obecně (chápem ve zmyslu "pro libovolné x (nebo s)") není aproximace s opravou lepší napr. ani pro $p = 0.5$ (a $n = 10$), nebo ani $n = 1000$ (a napr. $p = 0.25, 0.05$).

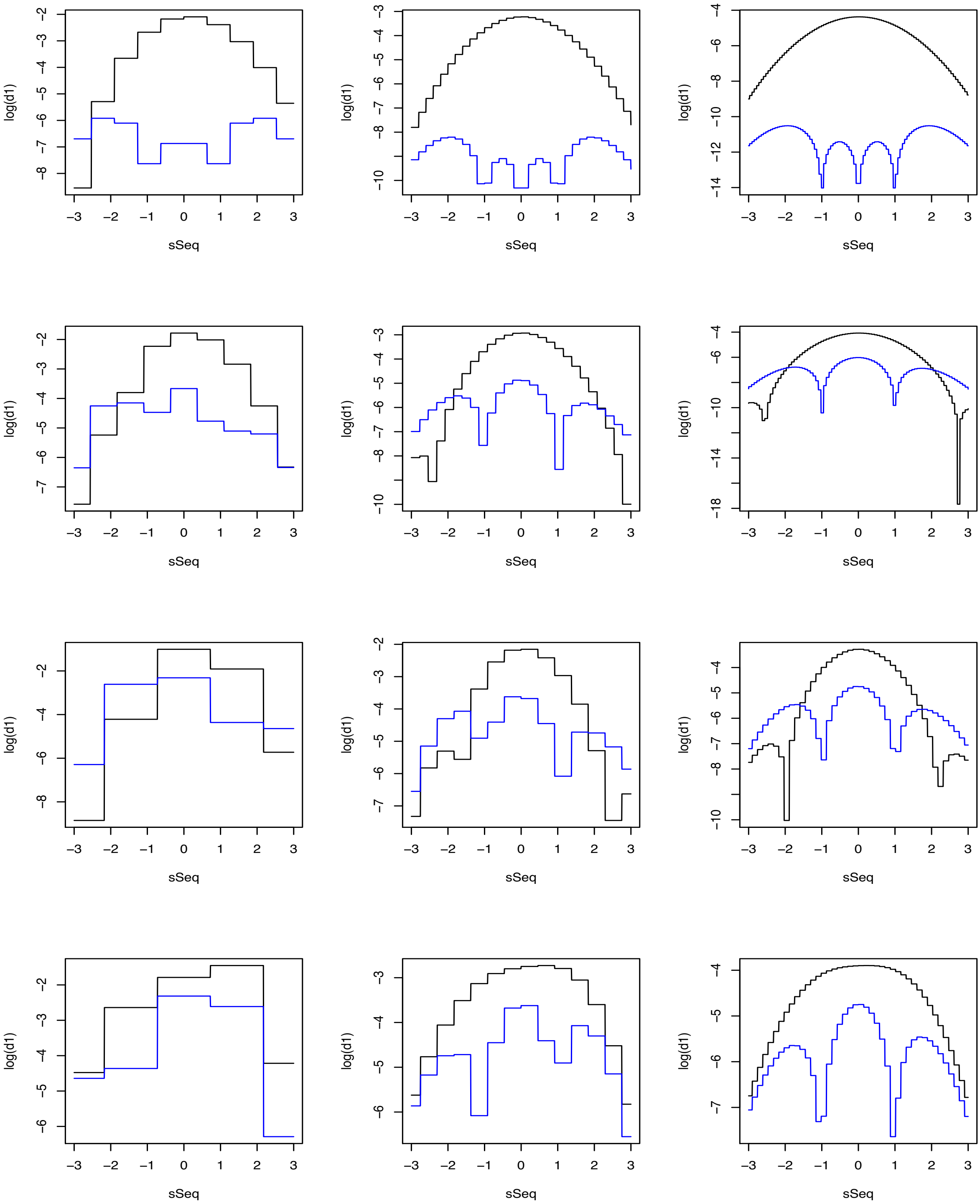
Taktiež tvrdenie, že "aproximace s opravou na spojitost nejeví známky žádného pravidelného chování" není asi celkom pravda. Tohle pravidelné chování lze těžko poznat pro omezenou volbu $s \in \{-3, -, -1, 0, -1, 2, 3\}$ (Obrazek 2.3), ale je podľa mňa celkom evidentné pre detailnejšiu voľbu $s \in (-3, 3)$ – viď rekonstrukci Obrazku 2.3 priloženú na záver.

- ❏ Kapitoly 3 a 4 pôsobia na mňa dojmom, že boli dopracované hodne narýchlo a na poslednú chvíľu: často príliš stručný text, výrazne väčšie množstvo preklepov (napr. zbytočné zátvorky na str.12 a 13), nedostatočné popisky Tabuliek 3.2, 3.3 a 4.1, neúplne formulácie (napr. chýbajúca alternatíva na str.20) a podobne. Naopak, oceňujem vizuálne zaujímavé a pekne vypracované Obrazky 4.2 a 4.3.

- ❏ Aký by bol autorov obecný záver (doporučenie) ohľadom použitia (nepoužitia) navrhovanej korekcie na spojitosť? Bolo by možné v špecifických prípadoch vylepšiť fungovanie opravy na spojitosť napr. použitím inej hodnoty $a \in (0, 1)$, než jedna polovina?

Praha, 05.06.2019

RNDr. Matúš Maciak, Ph.D.
maciak@karlin.mff.cuni.cz



Obrázek 1: Rekonstrukce Obrázku 2.3 pro volbu $s \in (-3, 3)$ se stejným rozložením vzhledem k pravděpodobnostem $p = 0.5, 0.25, 0.05$ a 0.95 (po řádcích) a $n = 10, 100$ a 1000 (po sloupcích).