

## POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

**Název:** Exponenciální rozdělení a jeho zobecnění

**Autor:** Vojtěch Vočadlo

### SHRNUTÍ OBSAHU PRÁCE

Cieľom predloženej práce bolo skúmanie niekoľkých menej známych zovšeobecnení exponenciálneho rozdelenia, ktoré boli v posledných rokoch navrhnuté. Autor vychádza z dvoch odborných článkov Gupta a Kundu (2007) a Nadarajah a Haghighi (2011), na základe ktorých predstavuje dve dvojparametrické zovšeobecnenia. V kapitole 1 sú skúmané základné vlastnosti týchto systémov hustôt, hlavne ich monotonicita a prvé momenty. Kapitola 2 pojednáva o odhadoch parametrov týchto rozdelení pomocou metódy momentov a maximálnej vierohodnosti. Kapitola 3 predstavuje simulačnú štúdiu v ktorej sú tieto odhady porovnané.

### CELKOVÉ HODNOCENÍ PRÁCE

**Téma práce.** Téma práce sa mi javí ako pomerne jednoduchá, avšak primeraná.

**Vlastní příspěvek.** V kapitole 1 autor derivuje hustoty a diskutuje, kedy funkcie rastú/klesajú. Záverom kapitoly 2 je zistenie, že ani štandardné momentové odhady, ani maximálne vierohodné odhady pre tieto rozdelenia nie je možné vyjadriť explicitne. Zaujímavejším výsledkom je simulačná štúdiu v kapitole 3 prevedená v programe *Mathematica*, v ktorej sú porovnané momentové odhady parametrov zavedených rozdelení s odhadmi metódou maximálnej vierohodnosti.

**Matematická úroveň.** Práca obsahuje iba niekoľko jednoduchých výpočtov derivácií zložených funkcií a integrálov. Aj v týchto výpočtoch sa však vyskytujú chyby a nepresné formulácie.

**Práce se zdroji.** Zdroje sú citované správne a práca neobsahuje preložené pasáže textu.

**Formální úprava.** Formálna úprava práce by sa síce dala zlepšiť, nemám ale žiadne zásadné výhrady.

### PŘIPOMÍNKY A OTÁZKY

1. Zavádzanie nových tried pravdepodobnostných rozdelení je jednoduché. Akýkoľvek systém nezáporných integrovateľných funkcií je možné predstaviť ako novú triedu rozšírení známych distribúcií. Práve preto je kľúčové, aby bolo každé takéto rozšírenie dobre motivované, ľahko interpretovateľné, alebo výpočetne jednoduché. Je dôležité tieto nové rozdelenia zasadiť medzi dobre známe parametrické rodiny, a diskutovať prečo by takéto rozšírenie mohlo byť zaujímavé. Všetky tieto aspekty sú uvažované v článkoch, z ktorých práca vychádza. Tu sa však autor obmedzuje na
  - uvedenie hustôt vo vzorcoch (1.1) a (1.2),
  - ich derivovanie (tj. vyšetrenie monotonicity) v sekcii 1.1,
  - ich integrovanie (tj. výpočet momentov) v sekcii 1.2, a
  - derivovanie ich logaritmov (tj. odvodenie maximálne vierohodných odhadov) v sekcii 2.3.

Všetky tieto elementárne výpočty sú bohužiaľ predvádzané s minimom diskusie, náhľadov alebo užitočných komentárov.

2. Častejšie než predstavené nové systémy zovšeobecnení sa v praxi vyskytujú iné rozšírenia exponenciálneho rozdelenia: Erlangovo rozdelenie, gama rozdelenie, alebo Weibullovo rozdelenie. Tieto triedy majú úzky súvis so skúmanými rozšíreniami. Nie sú však v práci vôbec popísané. Porovnanie nových rozšírení s týmito známymi by práci veľmi prospelo.
3. Vo vete 1 na str. 4 autor využíva iba pozorovanie, že  $\lambda$  parameter merítka pre rozdelenie  $X_1$ . To ale platí aj pre rozdelenie  $X_2$ , a táto diskusia mohla byť vedená súčasne pre obe rozdelenia. Ďalší triviálny dôsledok tohto faktu je vzťah pre hustotu  $X_1$  s parametrom  $2\lambda$  na str. 5, opäť uvedený iba pre  $X_1$ . Rovnako v kapitole 2 sú veľmi podobné výpočty predvádzané dvakrát, alebo iba pre jedno z rozdelení bez zmienky druhého prípadu (str. 13). Celkove, práci by prospel systematickejší prístup a prehľadnejšie spracovanie.
4. Výpočty v práci obsahujú chyby a nepresnosti:
  - str. 5: Pre  $\alpha = 1$  je derivácia hustoty  $X_1$  záporná aj v bode  $t = 0$  sprava.
  - str. 6: Hustota  $X_2$  klesá iba pre  $t > 0$ , nie všade.
  - str. 9: Predstavenie odhadu metódou momentov nie je korektné. Čo ak riešenie rovníc nie je jednoznačné? Je skutočne spojitosť  $h_i$  postačujúca pre konzistenciu odhadov?
  - str. 14: Platí veta 4 skutočne bez ďalších predpokladov?
  - str. 15: Vo všetkých rovniciach (2.10)–(2.13) sú vierohodnosti nesprávne derivované.
  - str. 16: Z výsledkov simulačnej štúdie určite nevidíme *rozptyl odhadov* našich parametrov. V tabuľkách na str. 17–19 nie je k dispozícii skutočné *vychýlenie odhadov* tak, ako je uvedené v záhlaví tabuliek.
5. Text miestami obsahuje nepresné a nematematické formulácie. Uvediem príklady:
  - str. 4: Hustoty a distribučné funkcie v základných definíciách (1.1) a (1.2) nie sú definované pre  $t < 0$ .
  - str. 11: Sú funkcie  $h_{X_j}$  pre  $j = 1, 2$  monotónne v  $\alpha$ ? Odkiaľ vieme, že ich inverzie existujú?
  - str. 11: Čím je významná hodnota 1.6 v pozorovaní 3? Znamená to, že funkcia  $h_{X_2}$  nemá pre  $\alpha < 1.6$  inverziu? Čo to pre naše odhady parametrov znamená?
  - str. 13: Je funkcia  $d(\alpha)$  skutočne kladná, alebo to iba usudzujeme z obrázku?
  - str. 23 a 24: Čo sú definičné obory týchto funkcií? Z predpisu pre funkciu gama sa zdá, že ide o komplexné čísla. Čo teda znamená integrál od  $x$  do nekonečna pre  $x$  komplexné v predpise pre hornú gama funkciu? Čo znamená  $\gamma$  a  $O$  v predpise pre  $H_x$  na konci str. 23? Nedá sa exponenciálna integrálna funkcia vyjadriť pomocou hornej gama funkcie? Je zeta funkcia skutočne definovaná pre každé  $z$  komplexné, resp. reálne? Vzťahy uvedené v kapitole 5 by si zaslúžili vhodné odkazy na literatúru, alebo aspoň omnoho detailnejší popis.
6. str. 21: Zdá sa, že dáta obsahujú 44 pozorovaní a nie 41 ako je uvedené. Maximálne pozorovanie má hodnotu 1776 mesiacov do smrti pacienta s rakovinou, čo zodpovedá takmer 150 rokom. Je to skutočne tak? Ako takéto pozorovanie ovplyvňuje uvedené odhady? Celkove, tak výsledky simulačnej štúdie, ako aj aplikácia na reálne dáta postrádajú ucelenejšiu diskusiu. Popis výsledkov simulácií a analýza reálnych dát nesmie byť obmedzená na mechanické popisy tabuliek s výsledkami, resp. uvedenie vektoru odhadnutých parametrov. Ktoré odhady sú lepšie a prečo? Ktorý z uvedených modelov je popisuje reálnych dáta lepšie, a ktorý z dvoch uvedených odhadov jeho parametrov je vhodnejší? Je použitie nových tried rozdelení štatisticky významne lepšie než použitie štandardného exponenciálneho rozdelenia, alebo nejakých jeho klasických zovšeobecnení (Weibullovo rozdelenie, gama rozdelenie)?

7. Čo je to *hustota dat*, ktorá sa spomína v abstrakte a v závere práce?
8. Prečo je v poznámke na str. 4 uvedená šikmosť a špicatosť exponenciálneho rozdelenia? Ako tento fakt zapadá do textu?
9. Referencia [7] bola publikovaná aj v odbornom časopise.
10. Bolo by možné uvažovať aj intervalové odhady neznámych parametrov, alebo testy hypotéz? Bolo by možné konštruovať napr. test hypotézy exponenciality rozdelenia, oproti alternatíve zovšeobecnenej triedy rozdelení typu  $X_1$ , alebo  $X_2$ ?

#### ZÁVĚR

Predložená práca skúma pomerne zaujímavý problém, autor sa však k nemu stavia mechanicky a odvodzuje iba jednoduché vlastnosti daných hustôt s naznačením možnosti odhadovania ich parametrov. Matematická časť práce je obmedzená na derivovanie a integrovanie zložených funkcií. Práci chýba nadhľad a precíznosť vo formuláciách výsledkov. Ako zaujímavejšia sa javí kapitola 3, ktorej obsahom je ucelená simulačná štúdia.

Vzhľadom k vyššie uvedeným komentárom považujem predloženú bakalársku prácu za pomerne slabú.



Stanislav Nagy  
KPMS MFF UK  
15. júna 2021