

OPONENTNÍ POSUDEK ZÁVĚREČNÉ KVALIFIKAČNÍ PRÁCE

Autor závěrečné práce: Bc. Martin Špetlík

Název práce: Rekonstrukce rozdělení pravděpodobnosti z odhadu zobecněných momentů

Oponent práce doc. Petr Volf, CSc.

Pracoviště opointa FP TU v Liberci

A. Kvalita abstraktu, klíčová slova odpovídají náplni práce	Výborn• (1)
B. Rozsah a zpracování rešerše	Velmi dob•e (2)
C. Řešení práce po teoretické stránce	Velmi dob•e (2)
D. Vhodnost, přiměřenost použité metodiky	Velmi dob•e (2)
E. Úroveň zpracování výsledků a diskuse	Výborn• mínus (1-)
F. Vlastní přínos k řešené problematice	Výborn• (1)
G. Formulace závěru práce	Výborn• (1)
H. Splnění zadání (cílů) práce	Spln•no s výhradou
I. Skladba, správnost a úplnost citací literárních údajů	Výborn• (1)
J. Typografická a jazyková úroveň (vč. pravopisu)	Výborn• (1)
K. Formální náležitosti práce (struktura textu, řazení kapitol, přehlednost ilustrací)	Velmi dob•e (2)

Komentáře či připomínky:

Viz přiložený list "Posudek DP.."

...pokračuje na straně 2

Celkové zhodnocení:

Např:
Komentář k výše uvedenému hodnocení
Aktuálnost tématu
Náročnost tématu na teoretické znalosti, vyšší úroveň zpracování literární rešerše, práce s vědeckou literaturou
Přínosy práce
Experimenty a jejich vyhodnocení
Význam pro praxi/teorii

Otázky k obhajobě:

Také na priloženém listě

Celková klasifikace:

Práce splňuje požadavky na udělení akademického titulu, a proto ji doporučuji k obhajobě.
Navrhuji tuto práci klasifikovat stupněm **Velmi dobře (2)**

V Liberci

dne 23.6.2020

Podpisem současně potvrzuji, že nejsem v žádném osobním vztahu k autorovi práce

.....

podpis oponenta

Posudek diplomové práce "*Rekonstrukce rozdělení pravděpodobnosti z odhadů zobecněných momentů*" Bc. Martina Špetlíka, Ústav nových technologií a aplikované informatiky FM TU v Liberci.

Cílem oponované diplomové práce je navrhnout a ověřit metodu rekonstrukce hustoty rozdělení pravděpodobnosti na základě znalosti konečného počtu R (či odhadu) jejích momentů. Protože řešení není jednoznačné, hledá se hustota s největší entropií. Tím je definována úloha optimalizace s R vazbami, která je převedena na řešení soustavy rovnic pro R Lagrangeových multiplikátorů. Odvození, implementace i ověření tohoto postupu je nejcennějším výsledkem, řešení tohoto úkolu je věnována podstatná část práce. Z hlediska "vyladění" algoritmu považuji za cenné zvláště části 3 (zlepšení podmíněnosti matice druhých derivací pro iterační algoritmus) a regularizaci hledaných řešení pomocí penalizace za přílišnou variabilitu v kap. 6. Kapitola 6.6 pak přináší výsledky vycházející z momentů různých typu rozdělení. Použitelnost metody je ukázána graficky, porovnáním s původní hustotou. Výsledky potvrzují, že metoda funguje a je schopna (alespoň za podmínek v této práci nastavených) tvar hustoty rekonstruovat. Samozřejmě, výsledek závisí na počtu použitých momentů (zpravidla jich bylo použito několik desítek).

A zde narážím na úskalí této práce. Ačkoli základní problém byl úspěšně vyřešen, jak jsem uvedl výše, jsem přeci jen obsahem práce mírně zklamán:

- Za prvé jsem čekal, že práce zmíní praktický případ, kde skutečně je třeba z pozorovaných několika momentů odhadnout tvar rozdělení. Něco o takovém problému jsem čekal už v úvodu, možná místo zbytečně obecných prvních dvou odstavců.
- Mé zklamání se prohloubilo, když se ukázalo, že pro rekonstrukci se vychází z momentů vzhledem k bázevým funkcím, z nichž je následně hustota sestrojena. Chápu, že použití přímo centrálních či ještě lépe necentrálních momentů by vedlo v rámci navrženého řešení k aproximaci logaritmu hustoty polynomem, což není vhodná volba z praktického hlediska (jak i autor uvádí). Proto se preferují ortogonální báze funkcí, zde konkrétně Legendreovy polynomy. Ale pak je právě sporná praktická cena metody, neboť vyžaduje znalost odhadů momentů právě vzhledem k těmto bázevým funkcím.
- Dále, v práci se jako vstup používají momenty přesně spočtené a následně zatížené malým šumem, nikoli momenty spočtené z (generovaných) dat. Dobře, pro ověření fungování metody je to vhodný postup, mohu při něm kontrolovat rozptyl šumu, ale zas je to krok dál od praktické situace. Navíc, poměrně rozsáhlé povídání o Monte Carlo a MLMC generování dat je zde vlastně zbytečné, neboť se MC nepoužije.
- Generování dat je pak sice použito v části 7, ta ale vlastně s námětem práce nesouvisí. Tam jde o rekonstrukci distribuční funkce, z kumulovaných relativních četností, které jsou spočteny ve zvolených bodech q_j , $j=1,2,\dots,S$. Autor je volí ekvidistantně (možná že by bylo vhodnější je volit úměrně empirickým $j/(S+1)$ kvantilům), získané body prokládá (interpoluje) B-spliny. Ne přesně, používá kombinaci s jádrovým vyrovnaním (na str. 58 nahoře, autor zde přestal číslovat vzorce), čímž se (celkem správně) dostal na půl cesty k vyrovňovacím (smoothing) splinům. To už je vlastně regresní metoda používající váženou metodu nejmenších čtverců, a tedy značně jednodušší než metoda zde uvedená. Autor dále uvádí, že důležitý je parametr vyrovnaní δ , ale kupodivu se vůbec nezabývá optimální volbou počtu bodů S . Ten opět řídí variabilitu (a případnou přeúčtenost) výsledné křivky, a na jeho optimalizaci existuje opět několik penalizačních kritérií. Navíc je i vyvinuto několik kritérií na porovnávání distribučních funkcí, jako např. Kolmogorov-Smirnov (supremální) či Cramér-von Mises (v L_2 normě). Celé téma neparametrické rekonstrukce distribuční funkce z částečné znalosti realizace příslušné náhodné veličiny by stačilo na diplomovou práci samo o sobě, zde je však jen načrtnuta 1 metoda, a není popsána blíže její optimalizace.

Jinak, jak jsem uvedl, tato kapitola moc nezapadá do rámce práce, i když je jistě zajímavé srovnat výsledky rekonstrukce v obou případech. Jde ale o to, že v kap. 7 máme k dispozici přímo data (více či méně agregovaná do intervalů), ne "jen" odhady momentů.

Poslední kapitola 8 ukazuje možnosti řešení rekonstrukce pro případ 2-rozměrné (anglicky "bi-variate", český překlad "bivarietní" asi není nejvhodnější) hustoty, což je jistě zajímavý směr, a v rámci zde formulované úlohy bude řešení nejspíš narážet hlavně na výpočetní problémy. Ale, jak jsem řekl, rád bych tu viděl ještě jednu kapitolu, která pomocí generovaných dat bude blízko reálným situacím (i s reálnou chybou v odhadech) a bude zkoumat úspěšnost rekonstrukce z prvních R (nepříliš velkého počtu) empirických necentrálních momentů.

Drobné připomínky:

- Grafy pro porovnání výsledků by měly být větší, je v nich zpravidla mnoho různých křivek, které je obtížné sledovat.
- str. 14 dole: V pravděpodobnosti se pro formu zpravidla připouští možnost, že daná vlastnost neplatí na množině míry 0. Takže $p=q$ skoro všude.
- str. 17 dole "J. Březina zavedl..", chybí [odkaz].
- odhady pro lognormální (i Cauchyho) rozdělení by asi bylo vhodné srovnávat na menším intervalu, aby byly vidět detaily.

SHRNUTÍ: Tato práce přináší cenné výsledky ve formulaci i řešení úlohy rekonstrukce hustoty rozdělení pravděpodobnosti ze zvolených momentů tohoto rozdělení. Formální stránka práce je pěkná, logicky dobře uspořádaná, překlady se téměř nevyskytují.

Soudím, že práce splňuje podmínky kladené na práci diplomovou a navrhuji ji k obhajobě. Protože k práci mám i řadu připomínek, které jsem uvedl výše, navrhuji hodnocení známkou 2 (velmi dobře).

V Liberci 23.6.2020

Petr Volf
FP TUL

Otázky pro obhajobu:

1. Není možné odvodit nějakou vhodnou transformací (rozepsáním Legendreových polynomů třeba) metodu rekonstrukce hustoty z běžných necentrálních momentů ?
2. Parametr regularizace alfa byl optimalizován na základě krosvalidace. Tento postup, ač odvozen teoreticky, nebývá v praxi příliš úspěšný. Byl zkoumán i výběr alfa prostým posouzením variability výsledné funkce? Podle výsledků soudím, že by vedl k vyšší váze penalizace.
3. Jak byl v kapitole 7 volen počet bodů S, a jak byly lokalizovány v případě rozdělení s nekonečnými chvosty (což je v této práci většina)?
