

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Multivariate Cox point processes

Autor: Noemi Kuželová

SHRnutí OBSAHU PRÁCE

Práce se zabývá odhadu parametrů pro vícerozměrné Coxovy bodové procesy, speciálně log gaussovské Coxovy bodové procesy. Pro speciální parametrický model převzatý z článku Waagepetersen et al (2016) odvozuje odhadovací proceduru založenou na složené věrohodnosti. Odvození tvaru složené věrohodnosti je podrobně provedeno a je rozšířeno z použití na jednorozměrné bodové procesy (kde je standardní metodou) na vícerozměrné procesy. Práce je uzavřena simulačním porovnáním odhadu pomocí odvozené složené věrohodnosti a jiné verze složené věrohodnosti v literatuře také používané.

CELKOVÉ HODNOCENÍ PRÁCE

Práce je logicky uspořádána, ze šířeji formulovaného zadání práce byla vybrána část, která je ucelená a pro magisterského studenta zvladatelná. V kapitolách 4 a 5, kde dochází k podrobnému odvození vzorečků pro složenou věrohodnost, je práce i korektní, vše je velmi podrobně propočítáno a zdůvodněno. Zde se prokázalo, že autorka je schopna samostatně použít znalosti ze svého magisterského studia (str. 23 a 24) a korektně je aplikovat na zkoumaný problém. V šesté kapitole prokázala i schopnost praktické implementace odvozených teoretických postupů a provedení a rozumného zhodnocení malé simulační studie.

Hlavní problém práce spočívá v kapitole 1 a 2, tedy ve shrnutí výchozích definic a teoretických výsledků. Je volně přecházeno a střídáno mezi přesnými formálními definicemi a vágními / neformálními / heuristickými výklady, ovšem používanými jako definice (tj. zavádějícími pojmy, které se v následných formálních definicích používají). Některé i přesně definované pojmy nejsou zavedeny tak, aby byly v následujících navazujících definicích použitelné, jindy se zase používá v dané chvíli ještě nedefinované značení. Ve výkladu není zřejmo, zda jsou popisované skutečnosti předpoklady, důsledky, a nebo se jaksí náhodně přihodilo, že daná věc platí. (Podrobněji viz sekce Připomínky a otázky) Výklad, kterým se autorka snaží formální matematiku vysvětlovat, ji spíš zatemňuje. Způsobuje to silný dojem, že autorka nerozumí tomu, co píše, resp. její představy nejsou přesné, a často jsou i chybné. Problém je o to horší, že všechny pojmy představené v kapitole 1 (i když rozhodně nejsou jednoduché) jsou obsaženy v přednášce magisterského studia Prostorové modelování.

Celkově: bylo odvedeno dost práce na to, aby bylo možno uznat předloženou práci jako diplomovou, ovšem matematická úroveň práce v kapitole 1 není dobrá.

Struktura práce. Práce je rozdělena do šesti kapitol, po úvodním shrnutí potřebné teorie z bodových procesů jsou představeny jedno- a vícerozměrné log Gaussovské Coxovy bodové procesy. Ve čtvrté kapitole, která tvoří těžiště práce, je podrobně odvozena složená věrohodnost prvního i druhého řádu, získaná jako limitní verze věrohodnosti aproximujících diskrétních modelů, a jsou určeny předpoklady, za kterých je tento postup korektní. Do obecného vzorce jsou dosazeny konkrétní tvary momentových hustot odpovídající poměrně komplikovanému modelu z Waagepetersen et al (2016). V páté kapitole je představen jiný druh složené věrohodnosti v literatuře používaný a opět je provedeno dosazení pro zkoumaný model. V závěrečné šesté kapitole je provedeno simulační porovnání obou představených verzí složené věrohodnosti.

Téma práce. Téma práce je přiměřené pro diplomovou práci a je aktuální. Bylo zpracováno v souladu se zadáním práce.

Vlastní příspěvek. Hlavním příspěvkem je podrobné odvození složené věrohodnosti jako limity věrohodnosti aproximujících diskrétních modelů ve čtvrté kapitole. Vlastním příspěvkem je i simulační porovnání chování odhadů získaných pomocí různých verzí složené věrohodnosti v kapitole 6.

Matematická úroveň. Od kapitoly 3 dále dobrá. V kapitole 1 dobrá není, viz popis výše a podrobné připomínky níže. Rovněž chybějící seznam značení by možná pomohl nejen čtenáři, ale i autorce si lépe srovnat, kterou notaci už zavedla a jak, a kterou ještě ne.

Práce se zdroji. Zdroje jsou citovány správně.

Formální úprava. Po formální stránce je práce na dobré úrovni, občas sice utečou věci typu velké písmeno v nadpise (např. kapitola 4.2.1), ale celkově je formální úprava v pořádku.

Jazyková úprava. Práce je psaná anglicky a myslím, že by jí prospělo, kdyby nebyla. Často se vyskytují chyby ve členech, předložkových konstrukcích, shodě podmětu s přísudkem, použití slov s jiným než zamýšleným významem a podobně. Všeobecně je použita angličtina velmi "počeštěná" a ztěžuje četbu a porozumění tomu, co vlastně chtěla autorka říci.

PŘIPOMÍNKY A OTÁZKY

1. str. 5, poslední řádek — vypadlo \mathcal{B} . Formulace před vzorcem je nepochopitelná.
2. str. 6 uprostřed — "abbreviated notation $\Psi(B) \dots$ " "We should realize ... $\Psi(B)$ " – označuje v obou větách $\Psi(B)$ stejný objekt?
3. definice 5 — opravdu jste chtěla napsat tuto definici takto? (odkud je převzata?) Jakým způsobem je tento náhodný element měřitelný? A nešlo by to udělat jednodušeji a správně (obzvláště, když ji nakonec využijete jen pro multivariate procesy?)
4. definice 6 — definice symbolu $N_X(B)$ tímto způsobem je možná jen pro JEDNODUCHÉ bodové procesy. Jinak by definice Λ nemusela být korektní. Předpoklad jednoduchého bodového procesu je potřeba ve všech definicích z kapitol 1.3 a 1.4. Bez něj není možné používat identifikaci bodového procesu jako náhodné míry a jako náhodné lokálně konečné množiny. Jak by bylo třeba zadefinovat symbol $N_X(B)$, aby byla definice 6 korektní pro každý bodový proces na \mathbb{R}^d ?
5. str. 7 dole — slovo "statistika" má v oboru statistiky přesný význam (jaký?) Charakteristiky ρ a $\rho^{(2)}$ nejsou statistikami.
6. definice 11 — definice má být definice. Není možné zavádět matematický pojem pouze slovním popisem za použití dosud nepoužitých a nevysvětlených (event) slov.
7. definice 12 — je proces X z definice "multitype point process" zavedený za definicí 5? Proč tedy není tak označen? Nebo jím není?
8. definice 13 — striktně vzato, nebyla zatím ρ_i formálně definována. A co když bude $\rho_i(u) = 0$, jak bude potom definována $g_{i,j}$?
9. definice 14 — proč tato definice (která má být zobecněním definice 11) vypadá úplně jinak než definice 11?
10. definice 15 — notace $N_X(B)$ byla zavedená na str.7. Ovšem v této definici takto zavedená $N_X(B)$ nebude obecně stačit. Pro jaké typy Poissonova procesu je potřeba předefinovat $N_X(B)$ obecněji, aby definice 15 byla správná?

11. str. 14, ř. 4 — "Then $\mu(u) = \mu, \forall u \in \mathbb{R}^2$ " – to je důsledek něčeho (a čeho), nebo předpoklad?
12. str. 15 — obecně je v matematickém textu lepší nejdříve funkce formálně zavést, a až poté vysvětlovat jejich význam.
13. str.16, nahoře — opravdu je míněno, že závislost kovarianční funkce pouze na délce rozdílu argumentů plyne ze stacionarity a izotropie? A stacionarity a izotropie čeho? Nebo to plyne z něčeho jiného?
14. str. 21, první vysazená formule — co je W (nebylo zavedeno)? Opravdu je potřeba omezenost ρ_θ na celém \mathbb{R}^2 ? Jak ji zajistíte ve vaší aplikaci, když neznáte hodnoty kovariát na celém \mathbb{R}^2 ?
15. str. 21 dole — proč je rozumné aproximovat $p_s^n(\theta) \approx \rho_\theta(\mathbf{u}_s^n) |C_s^n|$? Použití všech ostatních aproximací máte v textu zdůvodněné (proč jsou rozumné, a co jejich použitím zanedbáváte), tuto ne. Rozumím, že je převzata z článku, ale bylo by dobré si rozmyslet, proč je rozumná a co se při jejím použití zanedbává. Obdobně u $p_{s,t}$ na str. 26.
16. str. 22 uprostřed — to odůvodnění není dobře sepsáno. Tak jak je napsané je zavádějící.
17. str. 29 poslední odstavec — máte nějaké odůvodnění, proč by to tak mělo být (že body ležící příliš daleko od sebe nenesou žádnou informaci o hodnotě interakčních parametrů) ?
18. str. 32, Model simplification — co znamená předpoklad $\delta = \delta_i = \epsilon_i$ pro data? Třeba na příkladě rozmístění stromů v lese?
19. str. 41, před definicí relBias — je to relativní vychýlení čeho? Parametru θ , θ_0 nebo něčeho jiného?
20. Conclusion — a který postup byste tedy použila, když byste měla odhadovat parametry pro nějaká konkrétní data?

ZÁVĚR

Práci považuji za (pod)průměrnou a **doporučuji ji uznat jako diplomovou práci na MFF UK.**

RNDr. Michaela Prokešová, Ph.D.
 KPMS MFF UK
 3.6.2022