

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Modelling of segment process in the plane

Autor: Milan Pultar

SHRNUTÍ OBSAHU PRÁCE

Předložená bakalářská práce se věnuje konečným procesům úseček v kruhové rovinné oblasti. Rozdělení těchto procesů je dáno hustotou vzhledem k rozdělení Poissonova procesu úseček. Ve tvaru hustoty se kromě neznámých číselných parametrů objevuje i neznámá pravděpodobnostní hustota délek. Cílem je všechny tyto neznámé ingredience odhadnout z jedné pozorované realizace. To je provedeno aplikací metody maximální věrohodnosti a jádrového odhadu hustoty. Autor provedl simulační studii, na které ukazuje vlastnosti odhadů. V závěru jsou ještě navrženy dva složitější modely, které je možné simulovat pomocí MCMC metod.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Zadané téma vyžaduje znalosti bodových procesů, prostorové statistiky a metod MCMC, což jsou všechno oblasti, které se přednáší až v navazujícím magisterském studiu. Student tak musel potřebnou teorii nastudovat z odborné literatury. Téma byla zpracováno v souladu se zadáním práce.

Vlastní příspěvek. Hlavním přínosem práce je návrh a implementace semiparametrického přístupu k odhadování ve zkoumaném modelu. Chování odhadů je prozkoumáno pomocí počítačové simulace. Vlastním příspěvkem je také zavedení složitějších modelů a generování z nich. Tyto modely jsou motivovány reálným datovým vzorkem, který však již není součástí práce.

Matematická úroveň. Kapitola 1 obsahuje seznam potřebných definic a tvrzení (bez důkazu), v kapitole 2 je na začátku dokázáno tvrzení o tvaru hustoty Palmova rozdělení kót, na kterém je pak založen odhad referenční hustoty délek. Dále jsou v práci popsány jednotlivé modely a konstrukce odhadů. Vše je provedeno na solidní matematické úrovni, text je korektně zformulovaný. Ovšem na některých místech by autor měl být preciznější. Několikrát dochází k záměně N a \mathbf{X} (řádek před definicí 6, definice 7, definice 12). Slovní popis k \mathbf{M}^f na str. 6 spíše patří k \mathbf{M}_k . Definice dědičnosti (definice 15) není v pořádku. V algoritmu 1 by se zatím nemělo mluvit o úsečkách. V definici 16 je $P^y(G)$ číslo a ne rozdělení. Kromě toho se v práci vyskytuje pár překlepů ve vzorcích, např. λ místo R ve větě 2, S místo Y v definici 12, n místo m_j na str. 14, chybějící e_a ve vzorci (2.10), x_i místo u_i ve vzorci (3.1), přebytečné $\lambda^n(Y)$ na str. 19 dole. Pro interval možných délek úseček se používají různé symboly L_o a L_0 , také je tento prostor jednou roven $(0, e_a]$ a jindy $[0, e_a]$.

Práce se zdroji. Použité zdroje jsou řádně citovány. Uvedení položek v seznamu literatury není provedeno jednotným stylem a jejich řazení je poněkud zvláštní. V práci se nikde nevyskytuje odkaz na položku [4]. Jedná se o preprint autora a dvou dalších spoluautorů. Očekával bych proto v úvodu nebo závěru komentář, jak obsah bakalářské práce souvisí s tímto preprintem a jaký je vlastní podíl autora.

Formální úprava. V práci se vyskytuje několik typografických prohřešků (záměna spojovníku a pomlčky; matematické funkce jako sin, cos, supp, min nejsou stojatým písmem; někdy chybí

mezera po čárce; nesprávný způsob psaní symbolu pro normu). Práce je psaná obstojnou angličtinou, zlepšit by se daly některé neobratné formulace nebo použití členů.

PŘIPOMÍNKY A OTÁZKY

1. Jak je zaručeno, že součet T ve větě 1 je dobře definovaný?
2. Před začátkem podkapitoly 2.1 se zmiňují f (což je později f_1) a τ , které zatím nebyly vysvětleny. Pak se prohlásí, že se bude uvažovat hustota vzhledem k jednotkovému Poissonovu procesu, ale v podkapitole 2.1 se opět pracuje s mírou intenzity λ . Jak je to tedy ve skutečnosti?
3. Úsečka se někdy chápe jako prvek prostoru $Y = B \times (0, e_a] \times [0, \pi)$ a někdy jako množina v \mathbb{R}^2 , např. se píše $u \subset B$ při zavedení $d(u)$, přičemž o něco výše je $u = (y, r, \varphi) \in Y$. Podobně proces úseček je někdy bodový proces v Y a někdy příslušná náhodná množina vzniká sjednocením úseček, viz $\mathbf{x} \subset B$ ve (2.6). Toto je používáno bez dalšího vysvětlení, což někdy působí zavádějícím dojmem.
4. Je potřeba procesy v kapitole 2 simulovat algoritmem rození a zániku? Nebylo by efektivnější generovat nehomogenní Poissonův proces přímo?
5. V práci se významně využívá toho, že proces s hustotou (2.6) je izotropní. Bylo by proto vhodné definovat izotropii procesu úseček. Definice 5 se týká pouze bodových procesů.
6. Rozdělení do mezikruží ekvidistantním způsobem může znamenat, že v některých částech bude hodně pozorování a v jiných málo. Bylo zkoumáno, jaký to má vliv na výsledné odhady?
7. Odkud pochází volba $\rho(x)$ na str. 14?
8. Jaký cíl má rozlišení 20 a 40 realizací? Proč se rovnou neuvažuje 60 realizací?
9. Hodnota $\text{dist}(u_1, u_2)$ zavedená v podkapitole 3.2 nedefinuje metriku, a tak není moc vhodné ji označovat jako vzdálenost dvou množin.

ZÁVĚR

I přes některé uvedené výhrady považuji práci Milana Pultara za velmi dobrou a doporučuji ji uznat jako bakalářskou práci.

V Praze, 13. června 2017

doc. RNDr. Zbyněk Pawlas, Ph.D.
KPMS MFF UK