

Posudek oponenta na disertační práci

Mgr. Matúš Maciak

Flexibility, Robustness and Discontinuities in Nonparametric Regression Approaches

Předložená disertační práce se zabývá velmi aktuální a užitečnou problematikou-robustními neparametrickými regresními odhady. Tématicky patří práce do regresní analýzy. Klasické regresní techniky jsou založeny na regulární metodě nejmenších čtverců. Mgr. M. Maciak se však v předložené práci zabývá takovými postupy, které umožňují nejen přítomnost odlehklých pozorování, ale také případy, kdy chyby mají rozdělení s těžkými chvosty. Cílem práce je vytvořit vhodnou statistickou teorii pro flexibilní a robustní odhady regresních funkcí. Tento přístup spočívá v zobecnění lokálních lineárních M-odhadů. Tímto zobecněním jsou lokálně polynomiální M-odhady.

Disertační práce je rozdělena do sedmi kapitol. První kapitola je úvodem do neparametrické regrese.

Druhá kapitola je věnována lokálním polynomiálním M-odhadům. Nejdříve se autor zabývá homoscedastickým modelem a dokazuje pro tento případ statistické vlastnosti těchto odhadů (konsistence, asymptotická normalita). Další odstavec této kapitoly je věnován heteroscedastickým modelům. V tomto případě je variabilita modelu vyjádřena funkcí $\sigma(\cdot)$, kterou je třeba odhadnout. Pro tento odhad existují dva přístupy. První přístup používá odhadu pro homoscedastický přístup, kdy je třeba zvlášť odhadnout tuto funkci. V druhém případě je funkce $\sigma(\cdot)$ přímo implementována do minimalizačního problému. Takové odhady se nazývají studentizované M-odhady. Všechna tvrzení týkající se statistických vlastností odhadů jsou v této kapitole podrobně a pečlivě dokázána.

Další problém, kterým se autor v této kapitole zabývá, jsou M-odhady pro model se závislými chybami, který se často vyskytuje v praxi. V odstavci 2.4 je popsán „ α -mixing“ model a jsou dokázány statistické vlastnosti pro případ „ α -mixing“ závislosti.

Problematicke „bodů změny“ (change-points) je pro heteroscedastický model věnována třetí kapitola. Nejdříve autor pojednává o případech, kdy body změny jsou předem známy. Na druhé straně se uvažují případy, kdy nemáme a priori žádné znalosti o výskytu těchto bodů.

V úvodní části autor zdůrazňuje fakt, že v případě výskytu bodů změny, aplikace dříve uvedeného M-odhadu (s předpokladem nespojitosti), ale bez další dodatečné modifikace, má za následek nekonzistenci tohoto odhadu. Pozornost je také věnována M-odhadům s nesymetrickými jádry. V této kapitole je vybudována vhodná statistická teorie a jsou navrženy testy pro statistickou významnost bodů změny.

Asymptotické vlastnosti výše uvedených M-odhadů závisí na limitních distribucích, která závisí na neznámých veličinách. Jednou z možností je užít myšlenky plug-in a odhadnout neznámé veličiny dodatečným způsobem, ale získané výsledky nejsou z asymptotického hlediska vyhovující. Z tohoto důvodu je v práci navržen přístup založený na „bootstrap“ metodě. A čtvrtá kapitola je zaměřena právě na aplikaci „bootstrap“ přístupu pro M-odhady. Jsou zde navrženy algoritmy: *Smooth Residual Bootstrap* (homoscedasticity), *Smooth Residual Bootstrap* (heteroscedasticity) a *Moving Block Bootstrap* (α -mixing dependence).

Autor také formuloval a dokázal dvě důležité věty, z nichž je zřejmé, že bootstrap distribuce jsou vhodnou asymptotickou aproximací přesné, ale neznámé distribuce. Na základě

výsledků této kapitoly lze také konstatovat, že bootstrap přístup je vhodnější než plug-in přístupy.

Pátá kapitola je zaměřena na volbu dalších parametrů M -odhadů, a to na volbu vyhlazovacího parametru h , funkce $\sigma(\cdot)$ a ztrátové funkce $\rho(\cdot)$. Pokud jde o volbu vyhlazovacího parametru je uvedena nejčastěji užívaná metoda křížového ověřování (CV(.)) a její zobecnění RCV(.)- robustní funkce křížového ověřování. Pokud jde o funkci σ , jsou uvedeny nejdříve známé odhady a pak je uveden lokální M -odhad pro funkci $\sigma(\cdot)$ v daném bodě založený na po sobě jdoucích diferencích proměnné Y . V odstavci 5.3 je uvedena tabulka vhodných ztrátových funkcí $\rho(\cdot)$.

Šestá kapitola je věnována výpočetním aspektům M -odhadů. Je provedena rozsáhlá simulační studie zahrnující čtyři různá rozdělení chyb, čtyři různé ztrátové funkce, tři různé typy rozsahu výběru a tři různé stupně aproximace. Mgr. Maciak vhodně zvolil testovací funkci a zabývá se kvalitou odhadu ve třech bodech různého charakteru: bod spojitosti třetího řádu, bod nespojitosti první derivace a bod nespojitosti funkce. Výsledky studie jsou přehledně uspořádány do tabulek. V závěru této kapitoly je uvedena aplikace na reálná data včetně diskuze o použití M -odhadů pro konečný počet dat v daném výběru. Simulační studie ukazuje, že vlastnosti M -odhadů pro konečný rozsah souboru jsou v souladu s uvedenými asymptotickými výsledky.

V závěrečné sedmé kapitole autor ukazuje, že použití lokálně polynomiálních M -odhadů je přínosem v neparametrické regresi a zdůrazňuje flexibilitu tohoto přístupu.. Uvádí také směry dalšího výzkumu.

Práce je napsána přehledně a srozumitelně, i když formát práce je poněkud neobvyklý.

K práci mám pouze následující připomínky:

1. Uvažoval autor použít v simulační studii hladší jádra než je Epanečnikovo jádro, např. kvartické jádro?
2. Testové statistiky T jsou podle vztahu (3.4) nezáporné hodnoty (viz také poznámka 37 na str. 91), ale v tabulkách nabývají tyto statistiky i záporných hodnot.
3. V Tabulkách 1(a) (c)- 9(a),(c) jsou uvedeny hodnoty v závorkách, ale není jasné, jaký je jejich význam.

Závěr

Disertační práce obsahuje celou řadu původních výsledků. Zejména jsou to výsledky týkající se rozšíření lokálních lineárních M -odhadů na lokální polynomiální M -odhady. Dále jsou to nové výsledky týkající se heteroscedastických modelů a modelů se závislými chybami. Autor rovněž navrhl nové bootstrap algoritmy včetně algoritmů pro data se závislými chybami. Přínosem je také rozsáhlá simulační studie, která vhodně doplňuje a objasňuje získané teoretické výsledky.

Odborná úroveň práce je vysoká a výsledky práce jsou přínosem v oblasti robustní neparametrické regrese. Disertační práce jasně prokazuje, že Mgr. Matúš Maciak má předpoklady k samostatné tvůrčí vědecké práci.

Předložená disertační práce „Flexibility, Robustness and Discontinuity in Nonparametric Regression Approaches“ splňuje zcela jistě požadavky kladené na disertační práci, a proto navrhuji, aby po úspěšné obhajobě, v souladu se zákonem a platnými předpisy, byla Mgr. M. Maciakovi udělena vědecko-akademická hodnost Ph.D. v oboru doktorského studia Pravděpodobnost a matematická statistika.

