

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Mnohorozměrná teorie extrémních hodnot

Autor: Renata Šiklová

SHRNUTÍ OBSAHU PRÁCE

Práce pojednává o modelování vícerozměrných extrémních hodnot pomocí teorie jednorozměrných extrémních hodnot a kopul. Prezentované metody jsou pak ilustrovány v případové studii modelování výskytu katastrofických událostí.

CELKOVÉ HODNOCENÍ PRÁCE

Práce svým obsahem i rozsahem odpovídá diplomové práci. Autorka musela nastudovat rozsáhlou literaturu, která pro ni nebyla nijak jednoduchá, vzhledem k množství statistických přednášek, které měla příležitost absolvovat v rámci svého oboru. Mou největší výtka je, že autorka se občas vydává na tenký led a používá pojmy a koncepty, kterým příliš nerozumí. Některé její formulace mně navíc přijdou poněkud nejasné nebo vágní.

Téma práce. Teorie mnohorozměrných extrémních hodnot představuje ve své šíři velmi rozsáhlé a obtížné téma. Autorka zpracovala téma v rozumné hloubce, která odpovídala jejím možnostem.

Vlastní příspěvek. Autorka shrnula teorii jednorozměrných extrémů, kopulí a jejich využití při modelování vícerozměrných extrémů. Popsané metody pak použila na datech o úhrnech srážek.

Matematická úroveň. Matematická úroveň práce je slušná. Některé obraty a vyjádření však nejsou úplně nejšťastnější. Např. *Parametry γ jsou skoro jistě menší než 1.* (str. 44), *Gaussova kopula se jeví jako nejpravděpodobnější model* (str. 53). Někdy také není úplně jasné, co který symbol znamená.

Práce se zdroji. Zdroje jsou sice správně citovány, nicméně často chybí přesnější specifikace. Např. pro důkaz věty 1 se autorka odkazuje na knihu Beirlant et al. (2004), aniž čtenáři jakkoliv dále pomůže v tom, kde v knize danou větu hledat.

Formální úprava. Formální úroveň práce je slušná. Práci se nevyhnuly překlipy a stylistické nedostatky, jejich množství však není nijak výrazné. Dle mého názoru jsou až příliš využívány poznámky pod čarou, např. definice AIC, které je hojně využíváno v aplikační části práce, by si zasloužila místo v hlavním textu.

PŘIPOMÍNKY A OTÁZKY

1. Str. 5: Co znamená $M_n \xrightarrow{P} x_F$?
2. Str. 6: Jak vypadají distribuční funkce $G_{1,\alpha}(x)$ pro $x \leq 0$ a $G_{2,\alpha}(x)$ pro $x > 0$?
3. Str. 6: *Jestliže je splněn předpoklad věty 1...* Který předpoklad?
4. Str. 7: Limita ve tvrzení 3 je dle mého názoru jednostranná.
5. Str. 7: První rovnost ve vyjádření $l(\gamma, \mu, \sigma; M_{n1}, \dots, M_{nm})$ není v pořádku.
6. Str. 7: V jakých proměnných se funkce $l(\gamma, \mu, \sigma; M_{n1}, \dots, M_{nm})$ maximalizuje?

7. Str. 8: *Vhodné u tak identifikujeme z grafu ... Jak?*
8. Str. 9: $j = 1, \dots, N_u$.
9. Str. 9: Proč nazýváte odhad $\widehat{F}(x)$ semiparametrický?
10. Str. 11: $(t_1(X_1), \dots, t_d(X_d))$.
11. Str. 12: Definice 3. Pro každou permutaci čeho? Pro jaká (u_1, \dots, u_d) platí rovnost?
12. Str. 12: Autorka mluví o sdruženém t -rozdělení. Není však jasné, co se tím myslí.
13. Str. 14: *Dle věty 5...* Myslí se tím, dle tvrzení 5. Má autorka nějaký důvod, proč jednou formuluje větu a jindy tvrzení?
14. Str. 15: Co rozumí autorka spojením *plně asymetrické kopuly*?
15. Str. 17: Co se skrývá pod označením $\ell(\theta_C; \dots)$ v kroku 2 Dvoukrokové metody maximální věrohodnosti?
16. Str. 17: *Empirické d.f. nejsou spolehlivým odhadem chvostů rozdělení...* Proč?
17. Str. 17: Co se rozumí tím, že odhad není eficientní?
18. Str. 18: Proč označujete metodou popsanou v podkapitole 2.3.3 jako neparametrickou?
19. Str. 18: Jak vypadá model, pro který doporučujete odhad $\widehat{\theta}_{n,C}^{\text{PKTE}}$?
20. Str. 18: Co znamená, že odhad Kendallova tau je ve vícerozměrných případech velmi náročný?
21. Str. 18: Mluvíte o vychýlení a RMSE odhadů. Co tím myslíte?
22. Str. 19: Co myslíte kletbou dimenzionality?
23. Str. 19: *Čím více parametrů do modelu zahrneme, tím náročnější bude jejich odhad, chybovost předpovědi poroste.* Proč?
24. Str. 20: Symbol $\bigvee_{i=1}^n$ není vysvětlen.
25. Str. 20: *Řešení nastoleného problému se od jednorozměrné teorie liší. Limitní rozdělení afinně transformovaných maxim nelze reprezentovat parametrickou rodinou indexovanou konečně rozměrným parametrem. Mnohorozměrné rozdělení extrémních hodnot proto definujeme pouze obecně, nikoli explicitně analyticky jako ve větě 1.* Co se tím chce říct?
26. Str. 22: Krok od rovnosti $C^n(\mathbf{u}^{1/n}) = C(\mathbf{u}) = C^m(\mathbf{u}^{1/m})$ pro všechna $n, m \in \mathbb{N}$ k rovnosti $C^t(\mathbf{u}^{1/t}) = C(\mathbf{u})$ pro všechna t kladná by si zasloužil zdůvodnění.
27. Str. 23: Důkaz Věty 9. K ověření (3.9) ještě něco chybí.
28. Str. 25: Co to znamená *vhodnou rychlostí*?
29. Str. 25: Proč se zde používá spojení *neparametrický model kopul*?
30. Str. 27: *Preferujeme dvoukrokovou metodu...* Proč?
31. Str. 28: *Primitivním odhadem ...* Proč primitivním?
32. Str. 29: Nerozumím větě, že *Odhad je korektní pro $x \in \mathbb{R}^d \setminus [\mathbf{0}, \mathbf{u}]$.*

33. Str. 29: Kopula v C v podkapitole 3.3.3 je tedy symetrická?
34. Str. 30: Jak má člověk rozumět rovnosti $\{X \leq x\} = \{F(X) \leq F(x)\}$?
35. Str. 30 : Nerozumím větě *Speciální případy nezávislé a komonotónní kopuly jako obsaženy jako v (2.6)..*
36. Str. 32: Co je hodnota v riziku na hladině spolehlivosti 99.5% ?
37. Str. 38: Co to znamená, že data jsou homogenní?
38. Str. 39: Nerozumím spojení *iid rozdělení*.
39. Str. 39: Mohla byste slovy vysvětlit, jakým způsobem tedy nahrazujete chybějící hodnoty?
40. Str. 39: *Na první pohled vypadají řady závislé na dolním konci rozdělení. Proč?*
41. Str. 39: *Na Obrázku 4.9 vidíme převážně vysokou volatilitu ročních naměřených hodnot. Z čeho tak usuzujete?*
42. Str. 42: Jak rozumět poznámce pod čarou č.7? *Kdyby 5-denní úhrny srážek netvořily shluky, tak v průměru pouze 6 pozorování za měsíc.*
43. Str. 44 a Obrázek 4.12: Máte nějakou představu o tom, jak se počítá zmiňovaný intervalový odhad?
44. Str. 44: Co se myslí *mírně větší extrémní závislosti*?
45. Str. 44: Co se myslí tím, že Gaussova kopula *reflektuje celkový (lineární) vztah mezi pozorováními . . . ?*
46. Str. 44: Co se myslí tím, že *Opravdové rozdíly mezi kopulami extrémů nejsou. . . ?*
47. Str. 44: V práci jsem nenalezl vysvětlení, co se přesně rozumí *indexy a škodnými poměry*.
48. Str. 45: Co značí symbol F_{Klatovy}^1 ?
49. Str. 45: Co rozumíte pojmem *spillover*?
50. Str. 45: Nerozumím tvrzení *Gaussova a nezávislá kopula jsou ve střední hodnotě blízké kopulám extrémů, jsou však méně volatilní (údaje v závorkách, viz. také histogram na Obrázku 4.18).*
51. Str. 46: *S druhým nejnižším AIC skončil překvapivě model (d). Proč překvapivě?*
52. Str. 46: *Pravděpodobnosti, že jsou škody na všech portfoliích nad 99.5% kvantilem počítat nebudeme . . . Proč?*
53. Str. 53: *Grafy průměrných excesů nad mezí u tuto volbu nevyvracejí. Které grafy? Kde je čtenář nalezne?*
54. Str. 53–54: Pokud tomu správně rozumím, tak při analýze měsíčních dat vycházejí výsledky, které se dost liší od výsledků analýzy ročních dat. Výsledky analýza měsíčních dat se navíc ještě dost liší v závislosti na použité metodě. Autorka tyto rozdíly téměř nevysvětluje ani nekomentuje.
55. Str. 57: *Všechny přisuzovaly ve srovnání s Gaussovou kopulou vyšší pravděpodobnost sdruženým extrémům, nižší naopak extrémům marginálním, malým a středním hodnotám. Nerozumím tomu, jak různé kopuly mohou přisuzovat různé pravděpodobnosti marginálním extrémům. Nerozumím také tomu, co se myslí závěrem věty malým a středním hodnotám.*

56. Oceňuji, že přílohou k práci je kód pro výpočetní prostředí R pro provedení analýz a líbí se mi, že autorka přidala také kód pro vykreslení obrázků. Už jen z množství a netriviality kódu je patrné, že autorka v aplikaci odvedla nemalý kus práce. Bohužel, běh kódu se neobejde bez zásahů uživatele a některé části se mně nepodařilo rozchodit vůbec. Např. nepodařilo se mně zjistit, kde jsou spočítány proměnné `skody.hist` (soubor MEV.R) a `s85` (soubor POT.R.). Chybou také skončil příkaz `f.optimalizace.kopuly(pseudo4,pseudo5)`.

ZÁVĚR

I přes značné množství připomínek se domnívám, že předložená práce **splňuje** podmínky kladené na diplomovou práci pro obor Finanční a pojistná matematika a **doporučuji ji za ni uznat**. Pokud by se však při obhajobě ukázalo, že autorka není schopna na tyto připomínky reagovat anebo nerozumí základním pojmům, na kterých svoji práci staví, toto „doporučení“ by se mohlo změnit na „nedoporučení“.

Literatura

Beirlant, J., Gogebaur, Y., Teugels, J., and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, New York.

Ing. Marek Omelka, PhD.
KPMS MFF UK
3. června 2013