

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Intervaly spolehlivosti pro korelační koeficient

Autor: Martin Farda

SHRNUTÍ OBSAHU PRÁCE

První kapitola shrnuje základní poznatky o (Pearsonově) korelačním koeficientu. V druhé kapitole autor popisuje metodu zobecněných pivotů, které předpokládá dvourozměrnou normalitu. Ve třetí kapitole autor vysvětluje využití empirické věrohodnosti. V poslední kapitole pak pomocí simulační studie porovnává různé metody konstrukce intervalů spolehlivosti pro korelační koeficient.

Téma práce. Téma je přiměřené svou náročností bakalářskému studiu na oboru Finanční matematika a bylo splněno.

Vlastní příspěvek. Autor podrobně vysvětlil využití metody zobecněných pivotů a empirické věrohodnosti pro konstrukci intervalů spolehlivosti pro korelační koeficient. Pro metodu empirické věrohodnosti pak provedl některé kroky důkazu (úplný důkaz by byl nad rámec bakalářské práce). Za ocenění pak stojí také provedení netriviální simulační studie. Netriviálnost této studie pak spočívá v tém, že jak metoda zobecněných pivotů, tak zejména metoda založená na empirické věrohodnosti jsou značně výpočetně náročné.

Matematická úroveň. Matematická úroveň práce je solidní, avšak s jistými nedostatky. Např. z modelu na začátku kapitoly 2.1 vyplývá, že X_1, \dots, X_n jsou náhodné veličiny. Přesto však u řady výsledků uváděných v této kapitole to vypadá, jako by autor bral X_1, \dots, X_n jako fixní čísla. Některé místa textu jsou také napsány poněkud nešikovně, takže není úplně jasné, co autor přesně myslí.

Práce se zdroji. Zdroje jsou citovány správně. Občas by ovšem mohly být citovány přesněji, viz např. reference na článek Hall a La Scala (1990) na straně 20. Na kolik mohu posoudit, tak autor doslovně nepřejímal a snažil se, aby šlo opravdu o jeho vlastní text.

Formální úprava. Formální úroveň práce je slušná. Práce obsahuje sice jistý (nikoliv však neúměrně velký) počet překlepů. Nejasná a zřejmě neúplná je reference Flynn (1999). Vylepšit by se dalo také strukturování textu, aby se v něm lépe orientovalo (viz např. kapitola 2.2).

CELKOVÉ HODNOCENÍ PRÁCE

Práce je logicky uspořádána. Za uznání kromě simulační studie jistě stojí vcelku srozumitelné popsání docela komplikovaných metod a konceptů. Práci bych svým obsahem radil k mírně nadprůměrně náročným pracím.

Některé pasáže jsou však dle mého názoru nadbytečné (např. řada výsledků uvedených v kapitole 1) a některé jsou zmatečné (zejména kapitola 2.1). Vzhledem k nadprůměrnému rozsahu práce je škoda, že se autor nesoustředil jenom na to opravdu podstatné.

PŘIPOMÍNKY O OTÁZKY

1. **2 Věta 2:** Pro
2. **3 – 4:** Autor značí výběrový korelační koeficient se začátku r_{XY} , pak ale přejde ke značení r .

3. **4:** Co myslí autor větou: *Konstrukce intervalů spolehlivosti pomocí Fisherovy z-transformace využívá její asymptotické normality...*
4. **6:** Do prvního předpokladu **Věty 11** se vloudil překlep.
5. **8:** Domnívám se, že tvrzení c) **Věty 13** není v kontextu daného modelu pravdivé.
6. **9:** Nerozumím, co autor myslí větou: *Abychom mohli metodu rozšířit i na data nepocházející z normálního rozdělení, bylo by třeba najít alespoň asymptotická rozdělení funkce statistiky $(n-1)S_X^2$ a parametru σ_X^2 ...* Takové asymptotické rozdělení je o několik řádků výše.
7. **15:** Autor sice postupoval dle předlohy, ale mně by přišlo přirozenější definovat $\widehat{V}(W_i)$ jako

$$\widehat{V}(W_i) = \frac{X_i - \bar{X}_n}{\hat{\sigma}_X} \frac{Y_i - \bar{Y}_n}{\hat{\sigma}_Y},$$

kde $\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ a podobně pro $\hat{\sigma}_Y^2$. Potom by se suprema v definici $\widehat{L}_0(r)$ nabývalo pro $w_1 = \dots = w_n = \frac{1}{n}$.

8. **18¹⁷:** Vysvětlete podrobněji první nerovnost pro zbytkový člen $|r_2| \leq K \sum_{i=1}^n \dots$. Také není úplně jasné, proč platí, že $\frac{1}{n} \sum_{i=1}^n |\widehat{V}(W_i) - \rho|^3 = o(\sqrt{n})$. A co vůbec znamená pravá strana této rovnosti, když na levé straně jsou náhodné veličiny?
9. **19⁹:** Nerozumím, odkud se vzal člen $-o_p(1/n)$ ve vzorci pro $\widehat{l}(\rho)$.
10. **20:** Dle mého názoru není úplně přímočaré nahlédnout, že odhady $\hat{\sigma}_0^2$ a $\hat{\sigma}_v^2$ jsou konzistentní.
11. **23:** Ve vzorcích pro \bar{d} a *opp* by se asi nemělo sčítat do n .
12. **24 – 25:** Která z metod by pro příklad 2 mohla dávat jiné výsledky než pro příklad 1?
13. **4. Simulace:** Stejně jako se v (1.2) používá \sqrt{n} místo $\sqrt{n-3}$, tak by možná metodám, které jsou založené na empirické věrohodnosti, pomohla také nějaká podobná korekce (*finite sample correction*). Například, že by se místo kvantilu $\chi_1^2(1-\alpha)$ použil kvantil F -rozdělení o 1 a $n-2$ stupních volnosti.

DOPORUČENÍ K OBHAJOBĚ:

Připravte si odpovědi na připomínky č. 3, 5, 8 a 12.

ZÁVĚR

I přes výše uvedené výhrady a nedostatky splňuje práce všechny požadavky kladené na bakalářskou práci na oboru Finanční matematika a doporučuji ji za ni uznat.

doc. Ing. Marek Omelka, Ph.D.

KPMS MFF UK

16. června 2021