

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Semiparametrický model aditívneho rizika

Autor: Adéla Zavřelová

SHRNUTÍ OBSAHU PRÁCE

Diplomová práca študentky, bakalárky Adély Zavřelové, sa venuje semiparametrickému odhadovaniu rizika v rôznych štatistických modeloch analýzy prežitia. Práca je členená do štyroch kapitol. V prvej kapitole sú uvedené základné pojmy z analýzy prežitia a definované je potrebné značenie. V nasledujúcich dvoch kapitolách sú popísané dva štandardné semiparametrické modely rizika – multiplikatívny Coxov model proporčného rizika a aditívny, resp. lineárny model rizika od autorov Lin a Ying (1994). V štvrtej kapitole sú oba modely uvažované súčasne v tzv. kombinovanom modeli z článku Lin a Ying (1995). V prípade všetkých troch modelov sú odvodené základné štatistické vlastnosti, pričom hlavný teoretický prínos autorky je v podrobnom odvodení týchto vlastností pre prípad aditívneho a kombinovaného model. Doplnené sú niektoré teoretické podrobnosti potrebné pre praktickú aplikáciu, napr. postup odhadovania neznámych parametrov, alebo konkrétny spôsob implementácie v programe SAS. Kombinovaný model je zároveň posudzovaný aj z empirického hľadiska, pomocou autorkou navrhutej podrobnej simulačnej štúdie.

Téma diplomovej práce je pomerne náročná, ale jednoznačne vhodná pre diplomovú prácu. Vyžaduje mnohé netriviálne znalosti z oblasti matematiky, pravdepodobnosti a matematickej štatistiky. Samotné vypracovanie v podaní autorky ale pôsobí trochu schizofrenický. Na jednej strane komplikované podrobné odvodenia niektorých štatistických vlastností, na druhej strane triviálne chyby a často podstatné nedostatky, ktoré sa týkajú viac-menej všetkých aspektov práce: teoretickej stránky, formulácie matematického aj nematematického textu, ale aj celkovej prezentácie a úpravy.

V tomto smere považujem za najviac problematickú prvú a štvrtú kapitolu. Prvá pôsobí len ako náhodné okopírovanie skript doc. M. Kulicha k predmetu NMST531 (naviac zdroj chýba v zozname literatúry), kde autorka v podstate bez logickej náväznosti zredukovala 20 strán z pôvodných skript na dve strany prvej kapitoly. Nutne tak vzniklo množstvo nedefinovaného značenia, nezavedených pojmov, alebo nezmyselných formulácií. Zároveň chýba formálne rozlíšenie teoretického modelu a samotných pozorovaných dát – realizácií. Štvrtá kapitola, špeciálne od sekcie 4.3, vyzerá, že sa pod ňu podpísal už výrazný nedostatok času. Viac-menej formálne korektné teoretické odvodenia obsahujú množstvo preklepov, chýbajúcich znakov, alebo textu, ktorý akoby zostal zabudnutý z predchádzajúcich verzii práce. Ocenil by som aj prídanie aspoň stručnej interpretácie a vysvetlenie/zdôvodnenie u niektorých časti práce – napr. k uvedeným predpokladom, alebo niektorým krokom teoretického odvodzovania.

Z čisto teoretického hľadiska hodnotím prácu ako náročnú a teoretické spracovanie ako dobré. Z formálneho hľadiska je ale vypracovanie práce podpriemerné, obsahuje množstvo chýb, chýbajúce, alebo nevysvetlené značenie, ale aj mnohé ďalšie formálne nedostatky.

POZNÁMKY A PRIPOMIENKY

- Názov práce je *Semiparametrický model aditívneho rizika*. V úvode (druhá veta, str. 2), je napísané, že práca sa bude “*zabývať odhadem parametru v rôznych semiparametrických modelech rizika*”. Po úvode nasleduje 40(!) strán textu, na ktorých sa slovo *semiparametrický* vyskytuje

jedinýkrát (na str. 6, kde autorka uvádza, že multiplikatívny model je semiparametrický model). V samotnom závere práce potom autorka hneď v prvej vete zhrňuje, že v “práci jsme se seznámili se semiparametrickým modelem aditivního rizika” (str. 42). Ktoré z troch modelov predstavených v práci sú aditívne? Ktoré z troch predstavených modelov sú semiparametrické? Ktoré z troch predstavených modelov sú semiparametrické modely aditivného rizika? Vpodstate sa čitateľ v práci vôbec nedozvie, čo je to *semiparametrický model aditivního rizika*.

□ V úvode práce je napísane, že funkcia určujúca riziko “závisí na daných regresorech skrz neznámý parametr” a že práca sa bude “zabývať odhadem tohoto parametru”. Naozaj se jedná o jeden neznámy parameter? Podobných “drobných” nepresnosti je len v samotnom jednostranovom úvode niekoľko. A analogicky to vyzerá aj vo zvyšku práce. Uvádzam niekoľko príkladov len z prvých dvoch strán prvej kapitoly:

- Pre ktoré $t \in \mathbb{R}$ platí Definícia 1?
- Funkcia prežitia je $S(t)$, nie $F(t)$ (Lemma 1).
- V Definícii 4 sa jedná o náhodný proces $N(t)$.
- Na začiatku str. 4, nie sú definované dva procesy v (1.1) a (1.2), ale dohromady $2n$ procesov. Navyše, čo je to T_i , pre $i = 1, \dots, n$?
- V akom zmysle platí rozklad vo Vete 2?
- Čo je $N(s)$ a $Y(s)$ v definícii filtrácie \mathcal{F}_t ? Ani jedno nie je zadané. Síce autorka tvrdí, že $N(t)$ je definované v (1.1), ale není to pravda (tam sú definované pouze $N_i(t)$, pre $i = 1, \dots, n$). Taktiež není jasné, čo je to príslušný pozorovací proces $Y(t)$. V (1.1) a (1.2) mali byť asi definované teoretické procesy $N(t)$ a $Y(t)$.
- Čo je to podmínka nezávislého cenzorování v Lemme 3?
- Netreba niečo predpokladať pre martingál $M(t)$ v Definícii 5?
- Ako sú definované $\Delta M(0)$ a $\Delta N_i(t)$ v Lemme 4 a Definícii 6?
- V Definícii 6 není jasné, ako je definovaný príslušný mnohorozmerný čítací proces. Asi by malo byť uvedené, že sa jedná o náhodný proces $(N_1(t), \dots, N_n(t))$. Formulácia Definície 6 implikuje, že ktorýkoľvek proces $N_i(t)$ môže byť sám osebe mnohorozmerným čítacím procesom, čo nie je pravda.
- Celkovo mám pocit, že na strane 4 (resp. v prvej kapitole) došlo k nechcenému pomiešaniu teoretického modelu a konkrétnych realizácii.
- Niekoľko krát sa objavujú chýbajúce čiarky a bodky na konci viet, resp. nezmyselné/neštandardné začiatky viet.

Žiaľ, podobný výčet poznámok/pripomienok/chýb/nepresnosti by sa dal viac-menej urobiť pre ľubovoľné dve náhodné vybrané strany z práce.

- Teoretické predpoklady uvádzané v práci sú často nekompletné a nie vždy je zrejmé o čo konkrétne sa jedná a prečo. Napr. čo sa predpokladá o funkcii $\mathbf{b}(t)$ v P1.1, alebo P2.1 (funkcia odkiaľ kam)? Alebo, nie je potrebné predpokladať nejaké obmedzenie na kovariáty \mathbf{W} (str. 9)?
- V práci je popísaná implementácia v programu SAS, ale pri samotnej simulačnej štúdií je použitý program R. Z môjho pohľadu to pôsobí trochu nekonzistentne. Nebolo by lepšie popísať implementáciu v programe R, alebo využiť v empirickej časti program SAS?

OTÁZKY

- ❑ Vysvetlite, čo konkrétne sa v práci rozumie pod pojmom *semiparametrický model aditívneho rizika*. Čo sa myslí pod pojmom *semiparametrický model*?
- ❑ V práci sa viackrát objavuje výraz, že *martingálový integrál nahradíme nulou* (napr. str. 11, alebo str. 19). Prečo a na základe čoho je takýto krok opodstatnený?

- ❑ Vysvetlite, čo sa myslí značením

$$(U_1, \dots, U_r) \implies (W_1(t), \dots, W_r(t))$$

na $D^\tau \langle 0, \tau \rangle$, pre $n \rightarrow \infty$.

- ❑ Stručne vysvetlite a čo ide v predpokladoch P3.1, P3.2 a P3.3.

Praha, 26.06.2020


RNDr. Matúš Maciak, Ph.D.
maciak@karlin.mff.cuni.cz