

## OPONENTSKÝ POSUDEK NA DIPLOMOVOU PRÁCI

**Název:** Competing risk models in survival analysis

**Autor:** Jaromír Macoun

### Shrnutí:

Diplomová práce Jaromíra Macouna se zabývá metodami pro analýzu přežití se soupeřícími riziky. První kapitola obsahuje úvod do problematiky. Druhá kapitola se zabývá neparametrickými odhady. Je zde uveden a zkoumán Kaplan-Meierův odhad funkce přežití pro první pozorovanou událost, Nelson-Aalenův odhad kumulativního specifického rizika pro jeden typ události a neparametrický odhad kumulativní incidence jednoho typu události (CIF). Na konci druhé kapitoly je uveden důkaz slabé konvergence odhadu CIF. Třetí kapitola se zabývá regresními modely pro specifické riziko a subdistribuční riziko. Ve čtvrté kapitole jsou shrnuty výsledky simulačních studií.

- **Zadání práce**, tj. „samostatně nastudovat a přehledně shrnout“ toto téma, bylo bohatě naplněno. Diplomant vytvořil pěkně zpracované a promyšlené dílo, které spočívá v kritické kompilaci několika zdrojů, ale obsahuje i výsledky jeho samostatné práce.
- **Prísspěvek diplomanta** nespočívá jen v sestavení přehledu výsledků roztroušených v několika různých zdrojích, ale hlavně v úsilí předložit vlastní propracované podrobné důkazy klíčových tvrzení. Autor nachází a objasňuje chyby v publikovaných článcích, čímž prokazuje, že nepřijímá publikované zdroje nekriticky a přemýšlí o nich. Toto vše lze jen chválit.
- **Matematická úroveň práce** je výborná. Jaromír Macoun má snahu předvést podrobné a rigorózní výsledky, uvažuje o předpokladech používaných tvrzení, uvádí je a ověřuje je, používané termíny definuje, o výsledcích přemýšlí a interpretuje je. Občas toto úsilí není dotaženo úplně ideálně – o tom více v druhé části posudku.
- **Faktické chyby** se v práci téměř nevyskytují, snad až na chybu ve znaménku v tvrzení klíčové věty 13.
- **Zdroje jsou citovány řádně**, autor vždy připomíná, jaký zdroj právě používá, a kdy předkládá svou samostatnou práci. Mírnou výtku si zaslouží Bibliografie, kde dochází k nekonsistentnímu používání křestních jmen autorů.<sup>1</sup> Překlepy se v práci vyskytují, ale není jich příliš, občas se najdou různé stylistické neobratnosti v anglických větách (takže je jasné, že je formuloval sám autor a nikoli ChatGPT).

### Zásadnější připomínky:

1. *Str. 16, Věta 7:* Tvrzení je triviální, lze jej dokázat na jednom řádku.
2. *Str. 19, Věta 9:* Formulace věty není zcela korektní. Není vysvětleno, jakou roli v ní hraje ono  $\tau$  a co to vůbec je (v následujících výsledcích také ne), není uvedeno, pro jaká  $t$  tento rozvoj platí ani to, jestli ono  $o_P(1)$  má konvergovat bodově nebo stejnoměrně.
3. *Str. 20–21, důkaz věty 12:* Jakmile se do důkazu začne psát  $o_P(1)$ , které se příležitostně ztratí anebo se napíše do výrazů typu  $\int_0^t o_P(1)dCIF_j(u)$ , důkaz přestává být rigorózním, ale stává se heuristickým. To není žádná ostuda, ale je třeba to přiznat.
4. *Str. 22–25, důkaz věty 13:* Tento důkaz pro mne byl zprvu dosti neprůhledný. Očekával bych, že bude vycházet z Věty 12, ale místo toho se používá poznámka ze str. 22, která ovšem neobsahuje žádné tvrzení a v klíčovém výrazu jsou chyby  $(\bar{Y}^2(u)$  místo  $\bar{Y}(u)$ ). Je zde špatné znaménko u  $CIF_1(t)$ , které neodpovídá Větě 12, a tato chyba (je-li to chyba a ne mé přehlédnutí) se přenáší i do tvrzení Věty 13. Mám tedy pochybnosti o správnosti této věty. V důkazu Věty 13 by bylo dobré čtenáři sdělit, že

<sup>1</sup>DY Lin má jen zkratku, Bob Gray vystupuje též jako Robert J Gray, Bradley Efron křestní jméno vůbec nemá.

$\sqrt{n}(\widehat{CIF}_1(t) - CIF(t)) = U_1(t) + U_2(t) + o_p(1)$ . V poslední části důkazu není třeba používat funkcionální CLV, stačilo by odvolat se na funkcionálního Craméra-Wolda.

### Drobné připomínky:

- (a) Str. 7, Proposition 3: Specifikace for  $t \in (x_{k-1}, x_k]$  zjevně patří na začátek důkazu, nikoli do tvrzení pozice. V důkazu by bylo dobré napovědět, že  $\bar{Y}(x_1) = n$ .
- (b) Str. 8, ř. 13: Suma přes  $j$  od 1 do  $n$  ve výrazu pro  $\hat{S}(t-)$  je nejasná – co se sčítá a jak se na to přišlo?
- (c) Str. 8, výraz (1.2): Značení  $\lambda^{CS}$  ještě nebylo zavedeno.
- (d) Str. 11, Lemma 4: Není mi jasné, co jsou přesně ty „assumptions for competing risks“.
- (e) Str. 14, ř. 12 a 16: Velká  $\Lambda$  by měla být  $\Lambda^{CS}$ . Také se tu přechází mezi indexy  $j \in \{1, \dots, K\}$  a  $k \in \{1, \dots, K\}$ .
- (f) Str. 16, výraz (2.1): Chyba Gooleyho opravdu spočívá jen v posunu indexu při tisku, ale je dobře, že si toho diplomant všiml a spočítal to správně.
- (g) Str. 35, ř. 3: Zde asi má být  $C_i \geq u > t$  namísto  $C_i \leq u > t$ .
- (h) Str. 39–40, Kap. 4.1: Není zde uvedeno, kolik simulací bylo provedeno.
- (i) Str. 43, Obr. 4.2: Nevím, jaké intervaly spolehlivosti jsou na obrázku zakresleny. Skutečné hodnoty parametrů uvedené na obrázku (chápu-li správně nevysvětlené horizontální přerušované čáry) nejsou v souladu s hodnotami uvedenými na str. 41.
- (j) Výsledky simulací nejsou příliš užitečné. Člověk by očekával, že simulace poslouží např. k ověření platnosti Věty 13 a ukáže, jak ji použít v praxi. To tu ale vůbec není.

### Otázky k obhajobě:

Odpovězte na zásadnější připomínky 1, 2, a 4.

Předloženou práci Jaromíra Macouna celkově hodnotím jako vysoce nadprůměrnou a jednoznačně doporučuji ji uznat za práci diplomovou.

doc. Mgr. Michal Kulich, PhD.

KPMS MFF UK

2. června 2024