

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Truncated marked processes

Autor: Bc. Daniela Hrbáčová

SHRNUTÍ OBSAHU PRÁCE

Předložená práce se zabývá použitím kótovaných bodových procesů v neživotním pojištění, kde kóty odpovídají zpoždění mezi výskytem dané škodní události a jejím nahlášením. Cílem je odhad funkce intenzity procesu výskytu škodních událostí na základě procesu nahlášených škod a následně odhad potřebné rezervy na události, které již nastaly, ale ještě nebyly v daném čase nahlášený. Přístup k odhadu je parametrický, využívá předpokladu poissonovského rozdělení procesu nahlášených škod a stejně tak předpokládá konkrétní typ rozdělení pro náhodné veličiny udávající délku zpoždění a výši škody.

Po teoretické části, kde jsou používané metody představeny, následuje praktická část. V ní jsou uvažované metody použity na reálná data z oblasti pojištění motorových vozidel. Je zpracována podrobně, s řadou grafických výstupů a vysvětlujícími komentáři. Nejde tedy o pouhé vykopírování výsledků ze softwaru.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Zadané téma hodnotím jako zajímavé, vhodné pro diplomovou práci na daném oboru. Zadání práce bylo naplněno.

Vlastní příspěvek. Přínos autorky spočívá v provedení analýzy v praktické části. V teoretické části vidím přínos autorky v pečlivě odvozeném testu dobré shody v části 2.2 (s jednou výhradou uvedenou níže).

Matematická úroveň. Matematická úroveň práce je nízká, práce obsahuje řadu chyb a teoretická část vyvolává pochybnosti, zda autorka tématu skutečně rozumí. Práce obsahuje rigorózně formulovaný matematický text, který ale obsahuje chyby. Formulace definic a vět jsou zřejmě převzaty z různých zdrojů a na několika místech spolu nejsou konzistentní jak věcně, tak v terminologii.

Práce se zdroji. Použité zdroje jsou řádně citovány.

Formální úprava. Formální úprava práce je velmi dobrá, s výjimkou chybných odkazů na obrázky v praktické části. Jazyková úroveň by zasloužila vylepšit.

PŘIPOMÍNKY A OTÁZKY

K předložené práci mám tři hlavní výhrady. Zaprvé, práce obsahuje řadu chyb matematické povahy, které komentuji níže. Zadruhé, práce se zabývá useknutými kótovanými procesy, ale useknutí není v použitých metodách nijak reflektováno. V sekci 1.3.2 autorka useknutí obejde tím, že se zaměří na proces reportování škod, který je pozorován kompletní. Pro tento proces reportování autorka bude odhadovat funkci intenzity (to je korektní), a pak pomocí věty 4 spočítá odhad funkce intenzity procesu výskytu škod. To už ale není korektní, protože na konci sledovaného období bude spočtená funkce intenzity procesu výskytu škod podhodnocená (nemáme totiž informace o budoucím reportování škod, které v tomto období nastaly), přesně jak autorka uvádí v praktické části na straně 21 v odstavci nad obrázkem. To vede samozřejmě k podhodnocení potřebných IBNR rezerv.

Podobně by useknutí mělo být reflektováno v sekci 2.1.3, kde autorka odhaduje rozdělení zpoždění. Autorka se k pozorovaným zpožděním chová jako k iid náhodným veličinám. To by bylo v pořádku,

kdybychom nezávisle náhodně vybrali nějakou m -tici vyskytnuvších se škod a zkoumali jejich zpoždění. Jenže proces reportování vede k tomu, že škody s velkým zpožděním mají menší šanci být zahrnuty do našich dat – vidíme jen ty škody, které byly reportovány v daném časovém období.

Zatřetí, na začátku praktické části autorka komentuje tradiční metody řešení daného problému. Proč nejsou pro srovnání také použity v praktické části? Bez takového srovnání není možné hodnotit úspěšnost navrženého postupu.

Níže uvádím podrobné komentáře, které považuji za důležité. Drobnější připomínky s ohledem na délku posudku neuvádím. Pro přehlednost člením komentáře po jednotlivých kapitolách.

Kapitola 1:

1. Ve výkladu dochází ke střídání pojmů *intensity*, *rate*, *intensity function* (platí pro celou práci). Zřejmě jde o důsledek přebírání definic a vět z různých zdrojů.
2. Definice 3 – třetí vlastnost nejde splnit, pokud I je omezený interval. Druhá vlastnost platí pro všechna i nebo jen pro některá? Pravděpodobnost v druhé odrážce je podle mě vždy rovna jedné a do definice nic nepřináší.
3. Definice 4 – definice kótovaného procesu využívá bodový proces na součinném prostoru. Takový proces ale nebyl v práci zaveden. Nejde se odvolávat na definici 3, protože tu na součinném prostoru nejde použít.
4. Definice 5 – značení N pro kótovaný bodový proces je v konfliktu s definicí 4, kde šlo o nekótovaný bodový proces.
5. Věta 1 – měla by být formulována pečlivěji, například nejsou uvedeny předpoklady.
6. Definice 7 – homogenní Poissonův proces z definice 6 splňuje podmínky v definici 7 a měli bychom mu proto říkat nehomogenní Poissonův proces. To je terminologický konflikt. Dále, v poznámce za touto definicí by mělo být uvedeno, pro jaká s, t vzorec platí.
7. Theorem 4 – v předpokladech není uvedeno, na jakém intervalu je proces definován. Integrační obor ve vzorci pro funkci intenzity transformovaného procesu naznačuje, že na celém \mathbb{R} , ale to neodpovídá definici 7. Podobně, pro jaká y platí uvedený vzorec pro λ_{transf} ?
8. Vzorec (1.1) na straně 8 – použité značení není vhodné, co je to i a proč na něm pravá strana závisí a levá ne?
9. Příklad 2 na straně 9 – transformovaný proces bude definován na intervalu $[0, \infty)$.

Kapitola 2:

1. Odvození odhadů v sekci 2.1.2 – parametrický prostor by měl být uveden explicitně. Proč je podrobněji rozebrán zrovna model $e^{\varrho_1 + \varrho_2 t}$, když v praktické části je pak používán jiný model?
2. Vzorec (2.5) na straně 13 – t je integrační proměnná, uvedené podmínky $t \in (0, \infty)$, $s \in (-\infty, t)$ nedávají smysl.
3. Při odvození testu dobré shody na straně 14 není nutně splněno, že $\int_0^T \lambda(t, \varrho) dt = m$. V modelu nahoře na straně 14 je například vhodnou volbou ϱ_1 možné zařídit, že tento integrál bude mít libovolnou kladnou hodnotu.
4. Strana 15 – ve vzorci (2.6) se odečtou výše škod se stejným indexem, což nejsou nutně ty škody, které byly reportovány. Pokud jsou výše škod iid, nehraje to v dalším postupu roli, ale věcně vzato tento vzorec není správně.

Kapitola 3:

1. Odhad funkce intenzity v části 3.3.1 využívá specifický pětiparametrický model. Proč je zrovna tento model relevantní pro daný problém? Proč se nehodí použít jednodušší model, třeba vynechat kvadratický člen? Má to věcné důvody nebo jednodušší modely neprošly testem dobré shody? Pokud byly zkoumány i jednodušší modely, mělo by to být v práci komentováno, spolu s výsledky testů dobré shody.
2. Od strany 20 nefunguje žádný odkaz na obrázek.
3. Obrázky 3.2 a 3.3 – co znamená *empirical intensity calculated on a weekly basis*?
4. Sekce 3.3.2 – považují za vhodné zařadit graf odhadnuté hustoty.
5. Sekce 3.3.3, vzorec (3.4) – pro jaká s vzorec platí?
6. Sekce 3.3.4 – uvedený histogram rozhodně neodpovídá log-normálnímu rozdělení. Směs dvou log-normálních rozdělení se jeví jako vhodnější. To samé je vidět na obrázku 3.5, kde nemůžu s autorkou souhlasit, že odhadnutá hustota (u nuly) popisuje data dobře.
7. Sekce 3.4 – v integrálu na konci stránky je s jak spodní mezí integrálu, tak integrační proměnnou. Dále, v tomto místě je nesmyslný odkaz „hodnota byla získána numerickou integrací pomocí Wolfram Research“ (autorka uvádí firmu, ne použitý software).

Dotazy, které mohou zaznít u obhajoby:

1. Jak by správně měly vypadat definice 3 a 4 (jednoduchý bodový proces a kótovaný bodový proces), aby spolu byly konzistentní, resp. aby v definici 4 nic nechybělo?
2. Jak formulovat definici Poissonova bodového procesu, aby nevznikal terminologický konflikt mezi homogenním a nehomogenním Poissonovým procesem? (viz připomínku k definici 7 výše)
3. Ve výkladu na straně 7 je uvedena hustota pro nehomogenní Poissonův proces. Vzhledem k jaké referenční míře je tato hustota vyjádřena?
4. Jak při odvození testu dobré shody na straně 14 zohlednit skutečnost, že nemusí platit $\int_0^T \lambda(t, \varrho) dt = m$? (viz připomínku výše)
5. Proč má model pro funkci intenzity v části 3.3.1 zrovna tento tvar? (viz připomínku výše)
6. Strana 23, poslední řádek – autorka hovoří o predikci na rok 2019, jak byla tato predikce získána, když v tomto časovém období nemáme k dispozici hodnoty $\hat{\lambda}_{transf}$?
7. Tabulka 3.3 a obrázek 3.7 na straně 25 – co znamenají záporné hodnoty IBNR a jak by šel postup upravit, abychom se záporným hodnotám vyhnuli?

ZÁVĚR

Práci považuji za podprůměrnou, moje hodnocení bude záviset na průběhu obhajoby a odpovědích na otázky položené výše.

V Helsinkách, dne 19. 8. 2023

RNDr. Jiří Dvořák, Ph.D.