Posudek oponentky bakalářské práce

Název: Useknuté čítací procesy
Autor: Ondřej Přítel

Shrnutí obsahu práce

Práce se zabývá problémem odhadu intenzity počtu pojistných událostí v situaci, kdy mohou být pojistné události hlášeny se zpožděním, a tedy v daném časovém okamžiku nemáme k dispozici data o všech pojistných událostech do té doby vzniklých, ale jen o těch, které vznikly a zároveň už byly nahlášeny. Za předpokladu, že pojistné události jsou realizací nehomogenního Poissonova procesu a rozdělení doby do nahlášení jednotlivých událostí závisí jen na čase jejich vzniku a je nezávislé s Poissonovým procesem, platí, že i časy nahlášení událostí jsou realizací jiného nehomogenního Poissonova procesu a jeho intenzita je přesně určena intenzitou procesu událostí a rozdělením dob do nahlášení. Tento fakt v kombinaci s maximálně věrohodnými odhady intenzity nahlášených událostí a rozdělení doby do nahlášení je využit k odvození odhadu intenzity procesu pojistných událostí, resp. k predikci počtu pojistných událostí v budoucnosti.

Práce se snaží vyložit teorii Poissonových procesů, potřebnou k formálnímu zdůvodnění tohoto postupu a poté ilustrovat použitelnost navrženého odhadu na skutečných pojistných datech.

Celkové hodnocení práce

Téma práce. Téma je na bakalářskou práci na finanční matematice náročné, oponentka by řekla, že až příliš náročné. Z odevzdané práce je totiž zřetelné, že autor nemá dostatečné zkušenosti s náhodnými procesy na to, aby zvládl teorii v práci pojednanou bez větších chyb či nepochopení. Stejně tak v aplikační statistické části se projevuje omezený rozsah výchozích znalostí. Nicméně zadání práce bylo splněno.

Vlastní příspěvek. Mimo sepsání nutné teorie z různých zdrojů je vlastním příspěvkem autora propočítání několika konkrétních příkladů pro různé volby parametrické intenzity procesu událostí v kapitole 3 a provedení v práci navržené odhadovací procedury na reálných datech.

Matematická úroveň. Práce obsahuje rigorózně a korektně zformulovaný matematický text, ale není ho mnoho. Matematická úroveň práce není dobrá. V práci se vyskytují rozpory v definicích, nedefinované pojmy, chybějící (důležité) kvantifikátory, zavádějící i nepravdivé formulace. Podrobněji viz připomínky níže.

Práce se zdroji. Zdroje jsou správně citované, ale seznam literatury by zasloužit revizi po stránce typografické.

Formální úprava. Formální úprava práce by se jistě dala zlepšit — práce obsahuje nezanedbatelné množství pravopisných chyb a překlepů, popisy obrázků a grafů mají zčásti chybějící diakritiku.

Práce je dobře strukturovaná, ale v mnoha místech obtížně čitelná, respektive formulace jsou nesrozumitelné („problematika vytvoření takovéto predikce tkví v charakteru dat“, „hustota času“, „motivuje nás k zavedení Kingmanova Displacement theorem“, „testy ověřující spolehlivost maximálně věrohodných odhadů“, ...).

K aplikační části je důležité říci, že pokud se analyzují data a používá nějaké softwarové řešení, je třeba oboje k práci přiložit. Jinak nelze ověřit, co se vlastně skutečně při té analýze dat dělo (a jestli to bylo to, co je v práci popsáno, resp. částečně popsáno).

Vybrané připomínky a otázky

1. Na str.3 v Definici 2 je rozpor, protože množina přirozených čísel $\mathbb{N}$ neobsahuje 0.
2. Na str.3 poslední bod: stacionární proces a proces se stacionárními přírůstky jsou velmi odlišné věci.
3. Na str.4 první bod — pokud v matematice definujeme nějakou odvozenou (náhodnou) veličinu a lze to udělat explicitně, tak to děláme explicitně. Jak by šlo definovat U_i pomocí náhodných veličin $\{Z_n, n \in \mathbb{N}\}$? A lze formálně (tj. rovnicí, předpisem) definovat Z_n pomocí $\{N(t), t \geq 0\}$?
4. Za Definicí 5 tvrdíte, že z ní triviálně plyne, že $EN(t) = \lambda t$. Bylo by možné toto triviální odvození předvést?
5. To, co je napsáno před Tvrzením 1, není důkaz Tvrzení 1, ale povídání o tom, jak by se to dalo dokázat, kdyby se to udělalo pořádně. Rozhodně není pravda, že díky stacionaritě a nezávislosti přírůstků bude mít $N(t)$ binomické rozdělení. Přece dokazujete, že bude mít Poissonovo rozdělení. A oboje najednou mít nemůže.
   Co přesně tvrdí tvrzení nazývané ,,Poissonova aproximace“? A jak z něj tedy přesně vyplývá, že $N(t)$ najednou má Poissonovo rozdělení? A jaké Poissonovo rozdělení?
6. V Tvrzení 1 nesedí ona vložená interpretace náhodné veličiny $(N(t+s) - N(t))$.
7. V Tvrzení 2 by bylo dobré úplně specifikovat ona rozdělení - tedy včetně hodnot jejich parametrů.
8. Jak přesně v důkazu Tvrzení 2 plyne první rovnost ve druhé vysazené formuli z nezávislosti přírůstků procesu $N(t)$ přes pevné disjunktní intervaly? A stačí na druhou rovnost v té samé formuli opravdu stacionarita přírůstků $(N(t+\delta) - N(t))$?
9. V Definici 6 je potřeba doplnit, co je to s a z jaké množiny je. Resp. pro která s požadujeme platnost bodů (iii) a (iv).
10. Ve Větě 3 je potřeba napsat, pro jaká n ta rovnost (1.1) platí.
11. Důkaz Věty 3 — pokud se v matematickém důkazu objeví nový symbol (třeba $\triangle s$), je potřeba napsat, co to je, a odkud.
12. Důkaz Věty 3 — opravdu ,,veličina $p_n(s)$“ závisí jen na s? Pro všechna $t \geq$ je její hodnota stejná? Nebo máte to t zafixované a pro toto konkrétní t provádíte důkaz? To je třeba rozlišovat.
13. První řádek na str.8 — opravdu je to ekvivalence? A potřebujete, aby to byla ekvivalence?
14. Str.8, ř.2 a 3 — ,,řešením diferenciální rovnice bude pravděpodobnostní rozdělení veličiny“ — řešením diferenciální rovnice bývá obyčejně funkce a pravděpodobnostní rozdělení náhodné veličiny je dle definice tohoto pojmu pravděpodobnostní míra na $\mathbb{R}$. Co tedy je $p_n(s)$? Oboje to být nemůže.
15. Str.8, druhý odstavec — diferenciální rovnici řešíte běžným způsobem. Dokonce postupně řešíte spočetně mnoho diferencíálních rovnic běžným způsobem. Proto je potřeba ta matematická indukce.
16. Str.8, první rovnice na stránce — jak jste ji získal? Je to speciální případ rovnice na konci str.7? Pak to ale nesedí, resp. musíte dodefinovat $p_{-1}(s)$... ?

17. V matematickém textu nelze zavést značení (náhodných veličin T_i) příběhem o cyklistech. Je možné význam již zadefinovaných a označených veličin osvětlit příkladem, ale ne je jím definovat.

18. Str.13, ř.1 — jak rozdělení s těžkými chvosty zajistí to, o čem píšete? A jaký význam má to rozdělení s těžkými chvosty pro Vaši aplikaci?

19. V kapitole 4 jste použil jinou parametrizaci, než v kapitole 3. Má to nějaký důvod? Proč byste nemohl zachovat parametrizaci z kapitoly 3, odhadnout parametry τ funkce γ, odhady dosadit do λ_M, pomocí ní odhadnout zbývající parametry θ, a získat pak odhad původní λ_N? Co přesně se děje s useknutím při Vaší proceduře s otočením času? Je to nyní useknutí zleva? A předpovídáte intenzitu na intervalu $(0, \frac{1}{4})$? A sedí funční tvar intenzit procesu pojistných událostí a procesu časů nahlášení jako v příkladech v kapitole 3, nebo to nyní bude obráceně?

20. Str.20, odhad hustoty γ — při pohledu na obrázek 4.3 — co myslíte, jak moc vadí, že některé rozdíly W_i jste nepozoroval, protože některé škody ještě nebyly nahlášeny? Vyšel by Vám odhad hustoty γ s úplnými daty stejně, nebo nějakým podstatným způsobem jinak?

Při obhajobě by bylo vhodné reagovat například na dotazy 4, 13, 18, 19.

Závěr

Práci považuji za podprůměrnou, ale doporučuji ji uznat jako bakalářskou práci.

Návrh klasifikace sdělím předsedovi zkušební komise.

Jméno oponentky: Michaela Prokešová
Pracoviště: KPMS MFF UK
Datum: 3. září 2023