# Posudek oponenta bakalářské práce

**Název:** Operační riziko a značkovaný Poissonův proces
**Autor:** Karla Váchová

## Shrnutí obsahu práce

Předložená práce představuje základy teorie operačního rizika a základní definice Poissonova procesu v čase, jak ve verzi bez značek (kót), tak se značkami. Autorka ukazuje možnost jeho využití k modelování procesu výskytu a výše škod. Představené modely jsou pak využity k analýze simulovaných dat z veřejně dostupného balíku.

## Celkové hodnocení práce

**Téma práce.** Téma hodnotím jako přiměřené, svým obsahem i náročností vhodné pro obor Finanční matematika. Zadání práce bylo naplněno.

**Vlastní příspěvek.** Vlastní přínos autorky spočívá v shrnutí potřebných definic a vět v jednotném značení a v analýze simulovaných dat v kapitole 3.

**Matematická úroveň.** Práce obsahuje rigorózně formulovaný matematický text. Matematická úroveň práce není vysoká, množství chyb s ohledem na rozsah práce je vyšší, než by se dalo mlčky tolerovat.

**Práce se zdroji.** Použité zdroje jsou řádně citovány.

**Formální úprava.** Formální úroveň práce je vysoká, jazyková úroveň je dobrá.

## Připomínky a otázky

Kapitola 1 představuje koncept operačního rizika a s ním související pojmy a nástroje. Je sepsaná rozumně, zasazuje práci do kontextu. Je více povídací než matematická, což je v tomto případě v pořádku.

Matematické jádro práce leží v kapitole 2. Zde se bohužel vyskytuje řada problémů a chyb, které podle mého názoru ukazují, že autorka dané pojmy nepochopila pořádně, případně není při psaní tak pečlivá, jak by matematický text zasloužil. Podrobné komentáře uvádím níže. Na tomto místě jen dodám obecný komentář, že řada použitých vyjádření je matoucí či nešikovná, což ilustruje například sekce 2.3.2 o délce čtyř odstavců, kde se vyskytují formulace „na kartézsky násobených prostorech“, „kompletní separabilní metrický prostor“, „značkované procesy s událostmi v $\mathbb{R}^2$“ (jde o bodový proces v čase se značkami v $\mathbb{R}^2$ nebo o proces v $\mathbb{R}^2$ s blíže neurčenými značkami?). Na druhou stranu oceňuji sekci 2.4, tam se řada věcí pěkně vysvětlí.

Kapitola 3 pak popisuje aplikaci představeného modelu na simulovaná data. V první chvíli není jasné, co „aplikace“ znamená, protože v dosavadním textu jsou uvedeny definice, věty a popis modelu, ale žádná metoda odhadu jeho parametrů. Není tedy co „aplikovat“. Odhad parametrů nakonec autorka popisuje právě ve třetí kapitole, výrazně vhodnější místo by ale bylo ve druhé kapitole – jde o obecný výklad, není specifický pro zkoumaný dataset. Chápu použití simulovaných dat, protože dostat se k reálným není snadné, ale považuji za velmi nešťastné použít simulovaná data, o kterých nevíme, jak vznikla. Správnější by bylo připravit si vlastní simulovaná data, která umožní porovnat odhadnuté charakteristiky s těmi „správnými“. Vidím snad jediný argument, který by zdůvodnil použití cizích simulovaných dat bez znalosti modelu, který je vygeneroval, a to porovnání vlastních výsledků s předchozími analýzami téhož datasetu. To se zde ale neděje.

Podrobné připomínky:

1. strana 8, definice 2: čtvrtá vlastnost využívá pojem, který nebyl definován (body). Je tato vlastnost vůbec v definici potřeba?
2. s. 8, věta 1: je formulována jako definice, ne věta, kromě toho vůbec nefunguje (posloupnost náhodných veličin nemůže být neklesající a současně být tvořena nezávislými náhodnými veličinami).
3. s. 10 nahoře: co je to „diskrétní množina stavů se spojitým časem"?
4. s. 10, druhý odstavec: tyto formulace by asi měly být ekvivalentní s definicí 5, a ne s definicí 1.
5. s. 10, dole: pokud autorka tvrdí, že se bod 3 z definice 5 v praxi obtížně ověřuje, pochopí z toho čtenář, že body 3 a 4 z věty 2 se v praxi ověří snadněji. Je to tak?
6. s. 11, před větou 4: formulace „Poissonův proces lze chápat jako součet posloupnosti nezávislých náhodných veličin s exponenciálním rozdělením" je nesmyslná. Součet je jedno číslo, ale Poissonův proces je proces. Ve větě 4 se to pak vyjasní.
7. s. 11, definice 7, předpoklad „$\lambda(t)$ je integrovatelná funkce": prosím upřesněte a vysvětlete, zda $\lambda(t)$ takové, že $\int_0^\infty \lambda(t)\,\mathrm{d}t = \infty$ je přípustné nebo ne.
8. s. 11, dole: $\Lambda(t)$ je zavedeno jako funkce, vysvětlete prosím podrobně, proč se také dá nazývat míra intenzity.
9. s. 12, věta 5: formulace je zmatená, sčítá se zde konečný nebo nekonečný počet procesů? Pokud nekončený, není potřeba ještě podmínka na dílčí intenzity, aby byl výsledný proces dobře definován?
10. s. 12, věta 7: kromě chybného indexu v sumě, poslední věta je zmatená. Náhodná veličina s Poissonovým rozdělením má parametr, ne intenzitu, jak by to tedy bylo správně?
11. s. 13, definice 8: pojem bodového procesu nebyl zaveden.
12. s. 13, pod definicí 8: bodový proces je zde uvažován poprvé jako míra, to je nutné vysvětlit.
13. s. 14, nahoře: $\lambda_{N^*}(A \times M)$ není intenzita procesu, ale míra intenzity. Podobně v definici 9 se zavádí míra intenzity, ne intenzita. Jakou roli zde hraje $t$?
14. s. 14, pod definicí 11: formulace „značky jsou podmíněně iid" potřebuje podrobnější vysvětlení.
15. s. 14: definice 12 NENÍ analogií definice 4, tam šlo o stacionární přírůstky. Zde musí čtenář uhodnout, že $\mathcal{X} = \mathbb{R}$, aby ty posuny dávaly smysl. Dosud byl v této části prostor obecný, navíc pro kontext předchozích částí, kde bylo $\mathcal{X} = [0, \infty)$, ty posuny smysl nedávají.
16. s. 15, nahoře: dochází k drsnému konfliktu značení, náhodná míra je označena symbolem $\lambda$, přestože tento symbol je dávno použitý pro jeden z ústředních pojmů práce (intenzita, případně funkce intenzity).
17. s. 15, věta 8: co je to podmíněné rozdělení značek? Nebylo zavedeno.
18. s. 15, komentáře pod větou 8: jakou roli hraje $A$ v symbolech $\lambda(A)$ a $p_i(A)$?

19. s. 15, definice 13: jak $X(t)$ souvisí s $S(t)$? Upozorňuji, že NEJDE o speciální případ značkovaného Poissonova procesu.

20. s. 17, popis dat: vyjádření „buňka obsahuje časovou značku“ je v rozporu s dosavadní terminologií, tou značkou má být výše ztráty.

21. s. 19, dole: rozhodování o délce jednotkového časového intervalu je nesmyslné, jde jenom o škálování, nedostaneme žádnou další nebo přesnější informaci, když zvolíme třeba roční délku místo měsíční – vyzkoušejte si to a porovnejte odhady pro měsíc a pro rok.

22. s. 22: testování odhadnutých modelů: považuji za nevhodné používat k odhadu parametrů i k testování stejnou vlastnost dat (zde vyjádřenou AD statistikou). Takové testy bývají konzervativní a zbytečně mají malou sílu. Vhodnější by bylo provést jiný test.

23. s. 23: hodnotu VaR můžeme díky simulacím odhadnout, ale ne „spočítat“. Dále, hodnota nebude odhadnuta moc přesně, je založena na 10 nejvyšších hodnotách. Odhad bude velmi variabilní.

24. s. 23: zde se odhaduje $\mathrm{VaR}_{0,999}$, ale na s. 24 se odhaduje $\widehat{\mathrm{VaR}}_{99,5}$. Rozdíl je 1) v použití stříšky, 2) v normování na 1 vs. normování na 100, 3) ve volbě hodnoty 0,999 vs. 0,995. Odhadnuté hodnoty nepůjdou porovnat.

25. s. 24: porovnání se skutečně realizovanými ztrátami mohlo být mnohem podrobnější, díky nezávislosti v Poissonově procesu a iid značkám šlo vynechat libovolný měsíc (leave-one-out krosvalidace). Současný postup (jedna realizovaná hodnota vyšla menší než odhadnutý 99,5% kvantil rozdělení) nese velmi slabou informaci.

Dotazy, které mohou zaznít u obhajoby:

1. komentáře v seznamu výše, položky číslo 7 a 8.

2. s. 18, dole: nezávislost výše ztráty na čase události na takto dlouhém intervalu (10 let) je kvůli inflaci nerealistický. Například kumulovaná inflace v USA mezi lety 2007 a 2016 přesáhla 15 %. Jak se s tím v praxi vypořádat?

3. s. 19, rozhodování o homogenitě či nehomogenitě Poissonova procesu: dovedete formálně otestovat, zda homogenní model dobře popisuje data? Toto v práci citelně chybí, přestože odhadnuté rozdělení výše ztrát kontrolujete.

4. s. 19, dole: odhad parametru $\lambda$ (intenzity Poissonova procesu) není nijak vysvětlen. Prosím zdůvodněte ho.

5. s. 20, odhady parametrů v sekci 3.3: člověka první napadne metoda maximální věrohodnosti. Vysvětlete prosím, proč zde není vhodná, respektive proč nebyla použita, alespoň pro srovnání.

Závěr

Práci považuji za podprůměrnou, přesto podle mého názoru překonává potřebnou laťku a doporučuji ji uznat jako bakalářskou práci.

V Nesuchyni, dne 7. 6. 2022

RNDr. Jiří Dvořák, Ph.D.