Posudek oponenta bakalářské práce

Název: Základní stochastické epidemiologické modely
Autor: Karla Strachoňová

Shrnutí obsahu práce

Bakalářská práce Karly Strachoňové se zabývá modely pro šíření infekčních chorob založenými na Markovových řetězcích s diskrétním časem. Zkoumá dva typy modelů, které se liší konstrukcí pravděpodobností přechodu, a to Greenwoodův model a Reedův-Frostův model. Hlavní náplní práce je odvození rozdělení doby trvání epidemie a celkového počtu nakažených jedinců. Práce se dále zabývá odhadováním parametrů zmíněných modelů a testováním shody modelu s daty. Výsledky jsou ilustrovány na několika reálných příkladech.

Celkové hodnocení práce

Téma práce. Téma práce je pro bakalářskou práci velmi vhodné. Propojuje znalosti o diskrétních Markovových řetězcích se základními statistickými metodami. Předložená práce téma nepochybně naplňuje.

Vlastní příspěvek. Karla Strachoňová odvodila rozdělení doby trvání epidemie a celkového počtu nakažených jedinců pro Greenwoodův model a Reedův-Frostův model. Postup je převzatý z literatury, ale některé části jsou podrobněji vysvětleny. V práci je dále podrobně popsán postup pro odvození maximálně věrohodného odhadu parametru obou modelů a ukázán výpočet na reálných datech.

Matematická úroveň. Matematická úroveň práce je přijatelná. Na některých pasážích je patrná snaha o pečlivost. Jsou tu však ale i jisté podstatné nedostatky a některá odvození nejsou provedena důsledně ani přesvědčivě (viz připomínky níže).

Práce se zdroji. Zdroje jsou citovány hlavně v úvodu a závěru práce, jinde jen zřídka. V klíčové kapitole 3 není patrné, odkud byly předkládané výsledky převzaty. U některých výsledků (χ^2-test dobré shody) je uveden odkaz na celou knihu namísto na konkrétní větu, na kterou se autorka odvolává.

Formální úprava. Formální úprava je kvalitní, práce se dobře čte, překlepů je jen málo.

Připomínky

1. Hlavní připomínka: Nastavení Greenwoodova a Reedova-Frostova modelu není v souladu s teorií markovských řetězců. Není zde specifikována pevně daná množina stavů S (viz Def. 4). Celý výklad je podán tak, jakoby pozorovaná hodnota náhodného počátečního stavu $X_0 = x_0$ určovala množinu stavů $\{0, \dots, x_0\}$ pro X_t, $t > 0$. To ale nelze. Vyjádření pravděpodobnosti přechodu p_{ij} (str. 13) pak závisí na x_0 a tím porušuje markovskou vlastnost. Matice přechodu P má dimenzi $(x_0 + 1) \times (x_0 + 1)$ a její prvky závisejí na x_0. Z toho plynou i další nesrovnalosti, např. na str. 14 je vyjádřena $\mathsf{P}(T = t \mid X_0 = x_0)$ a poté vypočtena nepodmíněná $\mathsf{P}(T = t)$; ta ovšem ale stále závisí na x_0, což není možné. Tyto nesrovnalosti a rozpory naznačují, že autorka si svůj výklad nijak podrobně nerozmýšlela. Nakonec na straně 15 konečně padne předpoklad, že $\mathsf{P}(X_0 = x_0) = 1$, čili že počáteční stav je zafixován v x_0 a žádné jiné počáteční rozdělení se neuvažuje. Bez tohoto předpokladu by celý předchozí výklad byl zcela chybný. Ale zároveň je tento předpoklad fatálně omezující pro všechny případné aplikace.

2. Nerozumím výpočtu vytvořující funkce pro T a jejích derivací na str. 15. Pokud se autorka odvolává na *„vzorec pro maticový součet geometrické řady“*, měla by vysvětlit, o jaký vzorec jde, jaké jsou jeho

předpoklady, jak tyto předpoklady ověřuje a jak jej používá. Podobně nerozumím, jakým způsobem autorka derivuje vytvořující funkci $\Psi_T(\theta)$ a jak počítá $\mathsf{E}T$. Výpočet střední hodnoty (při $\theta = 1$) by implikoval, že $(I - R)^{-1}Q = I$, ale to není pravda.

3. Matice R definovaná na str. 14 není maticí pravděpodobností přechodu, protože nesplňuje podmínky definice 6. Nelze jí tak nazývat a nelze s ní tak pracovat. Zejména není jasné, proč R^{t-1} smí být interpretována jako „*matice pravděpodobností přechodu mezi různými stavy i a j v $t-1$ časových jednotkách*“.
4. U výrazů specifikujících pravděpodobnosti diskrétních veličin – $\mathsf{P}(X = x)$ – autorka nikdy neuvádí, pro jaká x toto platí. Skoro nikdy to není $\forall x \in \mathbb{N}$.
5. Str. 16, vzorec (3.7): Není jasné, jak volit $i_0, \dots, i_{t-1}$ a jak může jejich volba ovlivnit výsledek.
6. Str. 17, příklad: Nerozumím, odkud se vzalo vyjádření $\Psi_{(W,T)}(\varphi, \theta)$ v příkladu.
7. Str. 19, obr. 3.5 a 3.6: Proč v horním obrázku chybějí některé červené body a v dolním některé modré body?
8. Str. 24: Nerozumím tomu, jaký počáteční stav je předpokládán zde uvedeným nastavením $p(0)$.
9. Str. 27: Jaké je rozdělení dat uvedených v Tab. 4.1 a jak se získá věrohodnost L?
10. Str. 27: Chybí stanovení limitního rozdělení a kritických hodnot χ^2 statistiky.
11. Str. 29, (4.1): Neošetřená existence a jednoznačnost řešení věrohodnostní rovnice: Kolik řešení tato kvadratická rovnice má, které z nich je maximálně věrohodný odhad a jak víte, že právě jedno řešení bude v intervalu $(0, 1)$? Stejná připomínka lze vznést i ke všem dalším věrohodnostním rovnicím v kapitole 4. (Jsou tam i rovnice 6. stupně!!)
12. Kap. 5 – Aplikace: Dalo by se poznamenat, jak moc nerealistické a omezující jsou tyto modely, neboť předpokládají, že všechny rodiny jsou stejně velké a ve všech je na počátku tentýž počet nakažených jedinců. Ale také předpokládají, že nakazit se lze jen od nakažených jedinců uvnitř vlastní rodiny, nikde jinde. Za těchto okolností se člověk diví, odkud se v těchto rodinách ten první nakažený jedinec mohl vzít…
13. Str. 6: překlep „*fukce*“.
14. Str. 8: překlep „*věličiny*“.

Otázky a úkoly k obhajobě

(A) Předveďte korektní a přesvědčivý výpočet $\Psi_T(\theta)$, $\Psi'_T(\theta)$ a $\mathsf{E}T$ (viz str. 15 a připomínka 2).

(B) Odpovězte na připomínku 9.

(C) Předveďte korektní výpočet maximálně věrohodného odhadu z rovnice (4.1).

Závěr

Práce Karly Strachoňové navzdory dílčím nedostatkům splňuje požadavky kladené na bakalářskou práci, doporučuji ji tedy jako takovou uznat.

doc. Mgr. Michal Kulich, PhD.
KPMS MFF UK
20. června 2019