

Posudek oponenta k bakalářské práci **Bohuše Němčoviče** (2014):

Odhady v Markovských řetězcích se spojitým časem

Práce se zabývá odhadem matice intenzit homogenních Markovských řetězců se spojitým časem.

V prvních dvou částech je shrnuta potřebná teorie, v první ohledně matic intenzit a jejich odhadům metodou maximální věrohodnosti pro případy úplných a neúplných pozorování a ve druhé ohledně EM algoritmu a jeho aplikaci na odhady matic intenzit. Zejména v úvodu a první části jsou některé definice a značení, případně jejich interpretace, formulovány zmateně. U obhajoby může být objasněno, zda je student myslel správně a rozumí jim. Připouštím možnost, že jde o přehlédnutí a přepisy. Některé tyto nejasnosti a možné dotazy jsou rozepsány níže v posudku.

Ve třetí části je tato teorie aplikovaná na konkrétní poměrně šikovně vybrané příklady. Netriviální program, který provádí všechny potřebné výpočty, student sám naprogramoval v jazyce Wolfram Mathematica. Efektivita algoritmu, z pohledu vzdálenosti výsledku v eukleidovské normě po zastavení algoritmu od skutečné hodnoty, počtu iterací a času trvání algoritmu, je porovnána pro různé délky času pozorování a různé diskretizační kroky. Výsledky jsou srozumitelně a přehledně interpretovány a je vysvětleno, proč a pro které přístupy je algoritmus vhodný a efektivní.

V práci bych upozornil na:

1. Na str. 3 Definice 2 působí zmateně. Hlavně není jasné, co student myslí jako definici a co už jako důsledek zdefinovaného a kde má definice vlastně končit.
2. Na str. 5 za větou 2 mi není jasné, proč student používá $(t_k - t_{k-1})$ jako proměnnou v hustotě náhodné veličiny T (bez indexu), doby setrvání Markovského řetězce ve stavu i . Nikde předtím v práci není napsáno, co k znamená, ani není zmínka o ničem, co by mohlo být takto indexováno. V souvislosti s tím je dále matoucí, že student o dva řádky dál označuje $(t_k - t_{k-1})$ jako doby setrvání Markovského řetězce v uvažovaném stavě i (neměly by tyto doby setrvání na stavu i záviset?) a doslova o nich píše jako o nezávislých náhodných veličinách. U obhajoby by mě zajímalo, jak přesně toto bylo studentem myšleno.
3. Na str. 5, poslední odstavec. Co znamená v definici $G(t_k)$ index t u

procesu X ? Nemá místo t být t_k ? Nemělo by G záviset i na stavu i ? Zde jsou opět t_k myšleny jako náhodné veličiny?

4. Na straně 10 by bylo vhodné odlišit pozorování a proměnnou hustoty (neznačit je stejným písmenem) a bylo by vhodné odlišit hustoty pro úplná a neúplná pozorování (už jen proto, že mají různou dimenzi) vzhledem k tomu, že jde o bakalářskou práci.
5. Jak na str. 11 ve vzorci (2.3) závisí Q na θ ? Opravdu se (2.3) a (2.4) rovnají? Není zadefinováno, co přesně znamená $E_{X|y, \theta^{(m)}}$. Případně není někde ve vzorcích překlep?
6. Podobně není zadefinováno, co přesně znamená $\mathbb{E}_{\mathbf{Q}_0}$ ve vzorci (2.6) na str. 11.
7. Neměly by na str. 15 záviset T a t_k na j ? Přinejmenším v zápise $t_k - t_{k-1} = \Delta_j, j = 1, 2, \dots$, by nebylo na škodu to zdůraznit. Obecně je lepší zavádět, co jednotlivé indexy znamenají, zejména v bakalářské práci.

Celkově doporučuji předloženou práci uznat jako bakalářskou práci na MFF UK.