Dr. Jan Haskovec
P.O. Box 4148, 4700 KAUST
Thuwal 23955-6900, KSA
jan.haskovec@kaust.edu.sa

Posudek oponenta na bakalářskou práci
"Základní aspekty metody Markov Chain Monte Carlo"
studenta Adama Šumníka

Tématem předkládané bakalářské práce jsou vybrané teoretické aspekty metody Markov Chain Monte Carlo (MCMC), ilustrované na několika příkladech a numerických simulacích. Práce je rozdělena do čtyř kapitol. V první kapitole jsou zavedeny konečné markovské řetězce, které jsou základním konceptem pro definici metody MCMC. Druhá kapitola je věnována popisu metody MCMC, konkrétně Metropolis(-Hastingsova) algoritmu a Glauberově dynamice. Tyto teoretické koncepty jsou ilustrovány příklady a numerickými simulacemi (q-barvení a Ising model). Ve třetí kapitola je vysvětlen pojem času mixingu a ve čtvrté kapitole jsou prezentovány jeho horní a dolní odhady, aplikované pro příklad "dostatečně náhodného" zamíchání karet.

Práce je po formální stránce na dobré úrovni, až na sporadický výskyt gramatických chyb či překlepů a občasných stylistických neobratností. Po obsahové stránce jsou kapitoly 1-3 zpracovány čtivě a matematicky korektně. Čtvrtá kapitola nicméně působí dojmem uspěchané nedotaženosti a lehké chaotičnosti. Nejsou dostatečně vysvětleny souvislosti mezi zavedenými pojmy a příklady (např. souvislost mezi "sběratelnem kuponů" a mícháním karet), není patřičně diskutován význam definic 4.1.3 – 4.1.10, nejsou definovány některé pojmy ("populace") a symboly ($X_{\tau_{top}}$), důkazy jsou zkratkovité apod.

Několik konkrétních komentářů:

- S. 6: "Tato posloupnost tedy limitu nemá. Důležitou vlastností stacionární distribuce ..." Zde první věta, vztahujíc se ke konkrétnímu příkladu, nesouvisí s druhou, která je obecným výrokem o stacionárních distribucích. Toto může být pro problematiky neznalého čtenáře matoucí.
- Značení τ_x ve formuli (1.14) je v mírném konfliktu se značením $\tau(x)$ z Definice 1.2.2, což opět může být matoucí.
- S. 12: "Z konvergenční věty (věta 3.2.3) ..." - ovšem věta 3.2.3 je uvedena až v textu níže. Praxe citování vět či výroků dříve než jsou vysloveny obecně zhoršuje čitelnost a navozuje dojem neuspořádanosti textu (či, ještě hůře, myšlenek autora). Občas je nevyhnutelná, což zde ovšem není ten případ.
- S. 12: "Pro správné q-barvení je stacionární distribuce rovnoměrná". Proč je tomu tak? Je tento výrok zřejmý?
- Důkaz věty 3.1.2 by měl být lépe vysvětlen. Např., proč platí nerovnosti (3.7)?
- S. 33, Příklad 11: Z formulace "vložíme ji mezi ostatní karty" není zřejmé že karta může být vložena i jako nejspodnější. Např. pro 3 karty výraz "mezi" evokuje že třetí karta je vždy vložena mezi první a druhou.
- Ve Větě 4.1.2 není definován symbol $X_{\tau_{top}}$.
- V Definici 4.1.3 není definován pojem "populace".

- Dlouhá posloupnost `if-elseif` ve Zdrojovém kódu 1.1 je značně neobratná a jde proti principům programování v `matlabu`. Podstatného urychlení běhu programu se také dosáhne použitím funkce `randi` pro generování rovnoměrně rozdělených náhodných celých čísel. Můj návrh na vylepšení programu viz níže; v mé verzi `matlabu` jsem takto dosáhl zhruba čtyřnásobného zrychlení běhu (měřeno procedurou `tic - toc`).

Vzhledem k výše uvedenému doporučuji bakalářskou práci k obhajobě a navrhuji klasifikovat známkou **velmi dobře (B).**

Eindhoven, dne 10. 8. 2021 Jan Haškovec

Příloha - návrh na vylepšení Zdrojového kódu 1.1:

```
pocet_vzorku_na_prumer =150;
pocet_opakovani = 150;
mnozina_prumeru = zeros(1, pocet_opakovani);
maximum = 0;
minimum = 1000;

mv = [-2,1; -2,-1; -1,2; 1,2; 2,1; 2,-1; -1,-2; 1,-2];

for v = 1 : pocet_opakovani
    l=zeros(1,pocet_vzorku_na_prumer);
    for j = 1 : pocet_vzorku_na_prumer
        xy=[1 1];
        t = 0;
        while xy(1) ~= 1 || xy(2) ~= 1 || t == 0
            cislo = randi(8);
            ab = xy + mv(cislo,:);

            if ab(1) >= 1 && ab(2) >= 1 && ab(1) <= 8 && ab(2) <= 8
                xy=ab;
                t = t + 1;
            end
        end

        maximum = max(maximum,t);
        minimum = min(minimum,t);
        l(j) = t;
    end
    mnozina_prumeru(v) = mean(l);
end

...
```