

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA JADERNÁ A FYZIKÁLNĚ INŽENÝRSKÁ
KATEDRA MATEMATIKY

Obor: Matematické inženýrství
Zaměření: Aplikované matematicko-stochastické metody



Základní aspekty metody Markov Chain Monte Carlo

Bakalářská práce

Autor: Adam Šumník
Školitel: doc. RNDr. Jan Vybíral, Ph.D.
Školní rok: 2020/2021

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student:	Adam Šumník
Studijní program:	Aplikace přírodních věd
Studijní obor:	Matematické inženýrství
Studijní zaměření:	Aplikované matematicko-stochastické metody
Název práce (česky):	Základní aspekty metody Markov Chain Monte Carlo
Název práce (anglicky):	Basic aspects of Markov Chain Monte Carlo

Pokyny pro vypracování:

- 1) Student se seznámí se základním konceptem metody Markov Chain Monte Carlo (Metropolisův algoritmus, Gibbsovo vzorkování, Glauberova dynamika).
- 2) Student zadefinuje čas mixingů a zrekapituluje jeho základní vlastnosti, včetně pojmu silně stacionárního času a horních a dolních odhadů na čas mixingů.
- 3) Student aplikuje zavedené pojmy na různé příklady (hardcore model, Isingův model, míchání karet) včetně vhodně zvolených praktických simulací.

Doporučená literatura:

- 1) D. A. Levin, Y. Peres, E. L. Wilmer, Markov chains and mixing times. American Mathematical Society, Providence, R.I., 2009.
- 2) S. Brooks, Markov chain Monte Carlo method and its application. Journal of the Royal Statistical Society: series D (the Statistician) 47(1), 1998, 69-100.
- 3) S. Brooks, A. Gelman, G. Jones, X.-L. Meng, Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC, 2011.
- 4) M. Gasparini, Markov Chain Monte Carlo in Practice. The American Society for Quality Control and The American Statistical Association, 1997.

Jméno a pracoviště vedoucího bakalářské práce:

doc. RNDr. Jan Vybíral, Ph.D.

Katedra matematiky, FJFI ČVUT v Praze, Trojanova 13, 120 00 Praha

Jméno a pracoviště konzultanta:

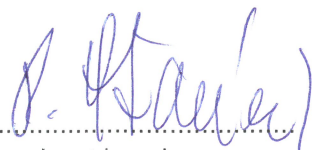
Datum zadání bakalářské práce: 31.10.2020

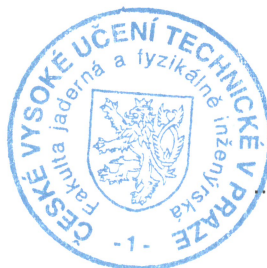
Datum odevzdání bakalářské práce: 7.7.2021

Doba platnosti zadání je dva roky od data zadání.

V Praze dne 30.10.2020


.....
garant oboru


.....
vedoucí katedry




.....
děkan

Poděkování:

Chtěl bych zde poděkovat především svému školiteli doc. RNDr. Janu Vybíralovi, Ph.D. za pečlivost, ochotu, vstřícnost a odborné i lidské zázemí při vedení mé diplomové práce.

Čestné prohlášení:

Prohlašuji, že jsem tuto práci vypracoval samostatně a uvedl jsem všechnu použitou literaturu.

V Praze dne 7. července 2021

Adam Šumník

Název práce:

Základní aspekty metody Markov Chain Monte Carlo

Autor: Adam Šumník

Druh práce: Bakalářská práce

Vedoucí práce: doc. RNDr. Jan Vybíral, Ph.D.
Katedra matematiky FJFI ČVUT v Praze

Abstrakt:

Tato bakalářská práce se zaměřuje na metodu Markov Chain Monte Carlo. Má za cíl představit základní myšlenky této metody a aplikovat ji na příklady. První část práce se věnuje konečným markovským řetězcům, které jsou základním stavebním kamenem metody Markov Chain Monte Carlo. Dále představuje Metropolisův algoritmus a Glauberovu dynamiku, které jsou ilustrovány praktickými simulacemi. Jelikož je žádoucí předem stanovit počet kroků pro sestavení kvalitních vzorků, které vzejdou ze simulace, v další kapitole je zaveden čas mixingů. V poslední části této práce pak je čas mixingů odhadován.

Klíčová slova: Markovské řetězce, Markov Chain Monte Carlo, Metropolisův algoritmus, Glauberova dynamika, čas mixingů

Title:

Basic aspects of Markov Chain Monte Carlo

Author: Adam Šumník

Abstract:

This bachelor thesis focuses on the Markov Chain Monte Carlo method. It aims to present the basic ideas of this method and apply it to examples. The first part of the work is devoted to finite Markov chains, which are the basic building blocks of the Markov Chain Monte Carlo method. It also presents the Metropolis algorithm and Glauber dynamics, which are illustrated by practical simulations. Since it is desirable to determine in advance the number of steps for assembling quality samples that will emerge from the simulations, it introduces mixing time in the next chapter. In the last part of this work, the mixing time is estimated.

Key words: Markov Chain, Markov Chain Monte Carlo, Metropolis algorithm, Glauber Dynamics, mixing time

Obsah

Úvod	1
1 Konečné Markovské řetězce	2
1.1 Markovský řetězec a matice přechodu	2
1.2 Neredukovatelnost a perioda markovských řetězců	3
1.3 Stacionární distribuce	4
1.4 Čas prvního příchodu	6
1.5 Reversibilita řetězce	6
1.6 Příklady	7
2 Markov Chain Monte Carlo: Metropolisovy a Glauberovy řetězce	11
2.1 Trocha historie	11
2.2 Úvod do metody Markov Chain Monte Carlo	11
2.2.1 Motivační příklad - správné q -barvení	12
2.3 Metropolisovy řetězce a Metropolisův algoritmus	13
2.3.1 Symetrický původní řetězec Ψ	13
2.3.2 Obecný původní řetězec Ψ	15
2.4 Glauberova dynamika	16
2.4.1 Reversibilita Glauberovy dynamiky a stacionarita π	17
2.4.2 Porovnání Glauberovy dynamiky a Metropolisova řetězce	18
2.4.3 Hardcore model s prchavostí	19
2.4.4 Isingův model	20
2.5 Praktické simulace	22
2.5.1 Správné q -barvení	22
2.5.2 Isingův model	25
3 Čas mixingů	28
3.1 Celková variační vzdálenost (TVD)	28
3.2 Párování	30
3.3 Čas mixingů	31
4 Horní a dolní odhady času mixingů	32
4.1 Horní odhad	32
4.1.1 Silně stacionární časy	33
4.1.2 Horní odhad pro míchání karet	35
4.2 Dolní odhad pro míchání karet	36
Závěr	38

Úvod

Markov Chain Monte Carlo (MCMC) jsou hojně využívané simulační metody, které nabízejí sestavení vzorků, jejichž kvalita roste se stoupajícím počtem kroků. Těto metody můžeme využít při jakékoli simulaci složitých stochastických systémů, typicky velkých dimenzí. Rovněž mají velký přínos při výpočtu vícerozměrných integrálů, kdy jsou vyčísleny numericky.

Počátky Metody Markov Chain Monte Carlo sahají do poloviny minulého století, kdy Nicholas Metropolis (1953) simuloval kapalinu v rovnováze s její plynnou fází. Algoritmus, který použil je pojmenován po něm, tedy Metropolisův algoritmus. Tato procedura byla roku 1970 zobecněna W. K. Hastingsem a výsledná podoba tohoto algoritmu nese jméno obou těchto osobností - Metropolis-Hastingsův algoritmus. Speciálním případem tohoto řetězce je Glauberova dynamika, kterou roku 1984 vymysleli bratři Donald a Stuart Gemanovi.

Téma jsem si vybral, neboť se rozhodně jedná o velmi zajímavou metodu se širokými možnostmi uplatnění.

Tato bakalářská práce je rozdělena do čtyř kapitol. První seznamuje čtenáře se základními pojmy a vlastnostmi markovských řetězců, kterých dále metoda Markov Chain Monte Carlo využívá. Druhá kapitola pak patří základním postupům metody Markov Chain Monte Carlo, zejména pak Metropolis-Hastingsovu algoritmu, Glauberově dynamice a závěr kapitoly nabízí některé zajímavé praktické simulace, ve kterých se této metody využívá. Jsou jimi například správné q -barvení nebo Isingův model. Protože je pro Markov Chain Monte Carlo stěžejní určit čas potřebný k běhu simulací, ve třetí kapitole je zaveden čas mixingů. Ten je poté odhadnut ve čtvrté kapitole pro příklad míchání karet, kde si způsob míchání předem stanovíme.

Cílem této práce je přiblížit čtenáři metodu MCMC od základů, vysvětlit ji co nejjednodušším a nejvýstižnějším způsobem a ukázat také její aplikaci na konkrétních příkladech. V simulacích se budeme pohybovat typicky na mřížce ve 2D, čímž si praktické užití metody v jistých ohledech zjednodušíme. Nebude to však na škodu, neboť nám to spíše poslouží pro lepší představu, jak metoda funguje, a navíc zobecnit metodu do více dimenzí nebo složitější struktury už není o moc složitějším úkolem.

Celá práce je doprovázena praktickými simulacemi, které byly sestaveny v programovacím jazyce MATLAB. Ilustrace v práci byly vytvořeny v editoru vektorové grafiky Inkscape.

Kapitola 1

Konečné Markovské řetězce

Ještě než se pustíme přímo do výstavby samotné metody Markov Chain Monte Carlo, uved' me nejprve několik velmi důležitých pojmů a vlastností týkajících se obecně Markovských řetězců, které budeme v celé této práci používat.

1.1 Markovský řetězec a matice přechodu

Zadefinujeme si na začátek nejzákladnější pojmy týkající se Markovských řetězců, kterými jsou Markovský řetězec samotný a matice přechodu řetězce. Informace z této sekce jsou čerpány z [12, sekce 1.1], [3, sekce 1.2], [5, sekce 1.3.2], [8, sekce 1.1] a [9, sekce 2.4.2].

Definice 1.1.1 (Markovský řetězec) *Posloupnost náhodných veličin $(X_n)_{n=1}^{\infty}$ je Markovský řetězec, pokud rozdělení X_{n+1} je určeno pouze hodnotou X_n . Množina hodnot, jichž nabývají stavy X_i , se nazývá stavovým prostorem \mathcal{X} Markovova řetězce. Matematickým zápisem tedy pro Markovské řetězce v každém čase t platí*

$$\mathbf{P}\{X_{t+1} = y | X_1 = x_1 \cap \dots \cap X_t = x_t\} = \mathbf{P}\{X_{t+1} = y | X_t = x_t\}, \quad (1.1)$$

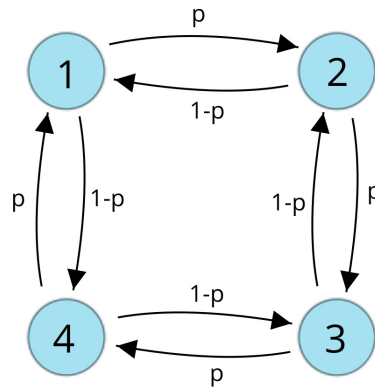
kde $x, y, x_1, x_2, \dots, x_t \in \mathcal{X}$.

Rovnice (1.1) se často nazývá **Markovova vlastnost** a znamená, že podmíněná pravděpodobnost stavu y trajektorií (posloupnosti stavů) před stavem y závisí jen a pouze na posledním členu posloupnosti před stavem y . Říká jinými slovy: Záleží pouze na tom, kde se nacházíme právě teď, nikoli na tom, kde jsme byli kdysi. Díky tomu, že přechod mezi libovolnými dvěma stavy nezávisí na čase t , ve kterém se přesouváme, můžeme tyto přechody mezi stavy zapsat do matice, která nám bude poskytovat veškeré informace o řetězci. Této matici se říká **matice přechodu** a je definována jako

$$P(x, y) = \mathbf{P}\{X_{t+1} = y | X_t = x\}, \quad (1.2)$$

přičemž tento vztah nezávisí na t . Díky této vlastnosti jsme schopni řetězec popsat pouze maticí přechodu P , která bude udávat s jakou pravděpodobností řetězec přechází mezi každými dvěma stavy. Velikost P je tedy $|\mathcal{X}| \times |\mathcal{X}|$.

Příklad 1 Ukážeme si Markovovu vlastnost na úplně jednoduchém příkladu. Uvažujme náhodnou procházku po čtverci. Házíme mincí, která má pravděpodobnost p , že padne panna a pravděpodobnost $1 - p$, že padne orel. Padne-li panna, posuneme se na čtverci na sousední vrchol čtverce po směru hodinových ručiček, zatímco padne-li orel, posuneme se na čtverci na sousední vrchol proti směru



Obrázek 1.1: Schéma náhodné procházky po čtverci.

hodinových ručiček (viz obr. 1.1). Necht' například začneme ve vrcholu 1 a máme posloupnost stavů $\{1, 4, 1, 2, 3, 2, 3, 4, 3, 4\}$. Jelikož jsou hody mincí nezávislé na průběhu trajektorie, můžeme jistě říct, že v následujícím hodu mincí je pravděpodobnost, že padne panna (a tedy se přesuneme do stavu 1), rovna p a pravděpodobnost, že padne orel (a pohyb bude směřovat do stavu 3) je $1 - p$. Přesně toto říká Markovova vlastnost. Ještě si nakonec sestavíme matici přechodu: V tomto případě je stavový prostor roven $X = \{1, 2, 3, 4\}$, tedy matice přechodu je velikosti 4×4 a má tvar

$$P = \begin{pmatrix} 0 & p & 0 & 1-p \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ p & 0 & 1-p & 0 \end{pmatrix}. \quad (1.3)$$

Takovým schématům jako je na obr. 1.1 se říká orientovaný váhový graf $G = (V, O)$. Množina V je množina vrcholů a O je množina dvojic stavů, které spolu komunikují. V tomto případě tedy $V = \{1, 2, 3, 4\}$ a $O = \{(1, 2), (2, 3), (3, 4), (4, 1)\}$.

1.2 Neredukovatelnost a perioda markovských řetězců

Markovské řetězce můžeme dělit do různých kategorií dle jejich vlastností. Dvě takové třídy dělí řetězce podle periody a redukovatelnosti/neredukovatelnosti (v některé literatuře také rozložitelnost/nerozložitelnost).

Definice 1.2.1 (Neredukovatelnost) Řetězec P se nazývá **neredukovatelný**, když pro libovolné dva stavy $x, y \in X$ existuje $t \in \mathbb{N}$ takové, že $P^t(x, y) > 0$. P^t zde značí matici přechodu P umocněnou na t .

To znamená, že je možné se pohybovat mezi libovolnými dvěma stavy v konečném čase [12, sekce 1.6], [8, sekce 1.3].

Náš příklad náhodné procházky po čtverci je rozhodně neredukovatelným řetězcem, neboť na sousední vrcholy se dostaneme v libovolném lichém čase a na protilehlý v libovolném sudém.

Definice 1.2.2 (Perioda) [12, sekce 1.6], [8, sekce 1.3]

Necht' $\tau(x) := \{t \geq 1 : P^t(x, x) > 0\}$ je množina časů, kdy je pro řetězec možné navrátit se zpět do stavu x . Perioda stavu x je definovaná jako největší společný dělitel $\tau(x)$. Periodu označíme jako $\gcd\{\tau(x)\}$ jako zkratku největší společný dělitel (the greatest common divisor).

Je-li $\tau(x) = \emptyset$, pak to znamená, že do tohoto stavu se nikdy nemůžeme vrátit, a tedy je řetězec redukovatelný. V takovém případě zavádíme periodu rovnou 0.

Vrátíme-li se opět k našemu příkladu náhodné procházky po čtverci, je jasné, že ať vyjdeme z jakéhokoli bodu, můžeme se do něho vrátit v čase $\{2, 4, 6, 8, \dots\}$. Z toho plyne, že pro periodu tohoto řetězce platí, že $\gcd\{\tau(x)\} = 2$ pro všechna $x \in \mathcal{X}$.

Lemma 1.2.3 [8, sekce 1.3]

Je-li P neredukovatelný, potom $\gcd\{\tau(x)\} = \gcd\{\tau(y)\}$ pro všechny stavy $x, y \in \mathcal{X}$.

Důkaz: Z neredukovatelnosti víme, že můžeme v konečném čase přecházet mezi libovolnými dvěma stavy. Zvolme tedy x, y libovolně a tedy existují $r, l > 0$ tak, že $P^r(x, y) > 0$ a $P^l(y, x) > 0$. Definujme nyní $m = r + l$, které nyní splňuje $P^m(x, x) > 0$ a zároveň $P^m(y, y) > 0$. Tím tedy máme zaručeno, že $m \in \tau(x) \cap \tau(y)$.

Označme $O_x = \gcd\{\tau(x)\}$ a $O_y = \gcd\{\tau(y)\}$ a uvažujme libovolný čas k takový, že $k \in \tau(x)$. Pak $k + m \in \tau(y)$, neboť

$$P^{k+m}(y, y) \geq P^l(y, x)P^k(x, x)P^r(x, y) > 0.$$

Vidíme, že O_y dělí m i $k + m$, kde k je libovolný čas z množiny O_x . To implikuje, že O_y musí dělit i libovolné takové k a tedy musí platit, že $O_x \geq O_y$. Ze symetrie pak dostáváme zároveň $O_x \leq O_y$ a tedy celkem $O_x = O_y$.

□

Důsledek 1.2.4 [8, sekce 1.3]

Pro neredukovatelné řetězce je perioda řetězce definovaná jako perioda, která je stejná pro všechny stavy.

Definice 1.2.5 (Aperiodičnost řetězce) [12, sekce 1.6], [8, sekce 1.3]

*Řetězec nazveme **aperiodickým**, pokud všechny stavy mají periodu 1. Když řetězec není aperiodický, nazýváme ho periodický.*

Pro případ náhodné procházky na čtverci je tedy řetězec periodický, neboť jeho perioda je rovna 2.

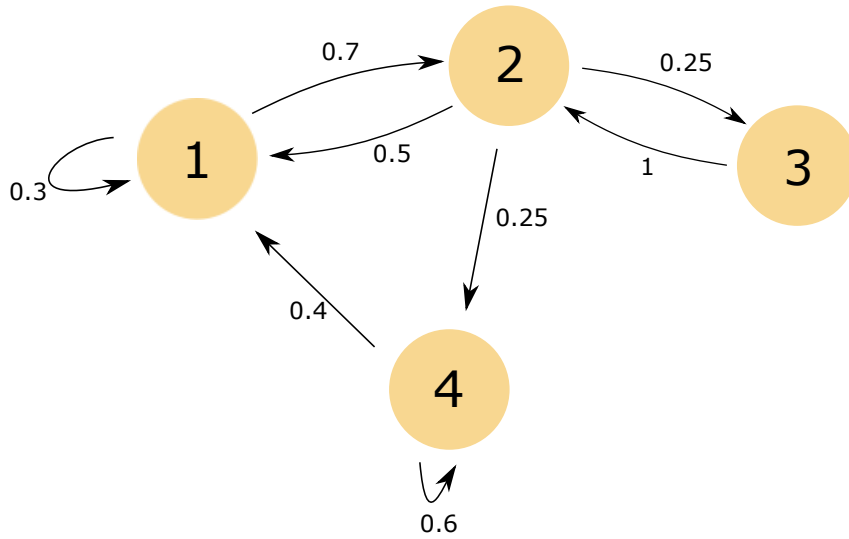
1.3 Stacionární distribuce

Nyní se podíváme na další důležitý pojem, kterým je stacionární distribuce. Ta charakterizuje limitní chování markovského řetězce. Nejprve se podívejme na následující příklad, abychom získali lepší představu o smyslu zavedení stacionární distribuce.

Příklad 2 Na obrázku 1.2 vidíme pravděpodobnostní schema přechodů mezi jednotlivými stavy. Zobrazený řetězec je aperiodický a neredukovatelný. Matice přechodu má v tomto případě tvar

$$P = \begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.5 & 0 & 0.25 & 0.25 \\ 0 & 1 & 0 & 0 \\ 0.4 & 0 & 0 & 0.6 \end{pmatrix}. \quad (1.4)$$

Uvažujme nyní, že vyjdeme ze stavu 1, tedy počáteční distribuce stavů bude vektor $\mu_0 = (1, 0, 0, 0)$. Dále distribuce v čase $t = 1$ bude ve tvaru



Obrázek 1.2: Pravděpodobnostní schéma

$$\mu_1 = (0.3, 0.7, 0, 0) = \mu_0 P. \quad (1.5)$$

Když si vyjádříme pravděpodobnost, s jakou budeme ve stavu 1 v čase $t = 2$, dostaneme

$$\mathbf{P}\{X_2 = 1 | X_0 = 1\} = 0.3 \times 0.3 + 0.7 \times 0.5 = \mu_1(1)P_{11} + \mu_1(2)P_{21}, \quad (1.6)$$

což znamená, že pro kompletní vektor μ_2 dostáváme

$$\mu_2 = \mu_1 P. \quad (1.7)$$

Pro obecné vyjádření μ_t a pro libovolnou počáteční distribuci pak platí

$$\mu_t = \mu_{t-1} P = \mu_0 P^t \quad \text{pro všechna } t \geq 1. \quad (1.8)$$

Sestavíme-li si několik členů této posloupnosti, zjistíme, že má tato posloupnost limitu. Tuto limitu označíme π . Použijeme-li limitu pro obě strany rovnice (1.8), dostáváme vztah

$$\pi = \pi P. \quad (1.9)$$

Tuto pravděpodobnostní distribuci π budeme nazývat stacionární distribuce řetězce a pro tento případ z (1.9) vychází

$$\pi = (0.36, \quad 0.34, \quad 0.085, \quad 0.215). \quad (1.10)$$

Definice 1.3.1 (Stacionární distribuce) [12, sekce 1.9], [3, sekce 1.4], [5, kapitola 15], [8, sekce 1.5]
 Mějme markovský řetězec s maticí přechodu P . **Stacionární distribucí** π rozumíme takový vektor, který splňuje vztah

$$\pi = \pi P. \quad (1.11)$$

Vztah (1.11) můžeme zapsat také prvkově

$$\pi(y) = \sum_{x \in \mathcal{X}} \pi(x) P(x, y) \quad \text{pro všechna } y \in \mathcal{X}. \quad (1.12)$$

Bude-li nás ale zajímat stacionární distribuce náhodné procházky po čtverci, což je případ neredukovatelného periodického řetězce, zjistíme, že vztah (1.11) jednoznačně splňuje stacionární distribuce

$$\pi = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right).$$

Sestavíme-li však prvních pár členů posloupnosti $(\mu_t)_{t=1}^{\infty}$ ze vztahu (1.8), kde $\mu_0 = (1, 0, 0, 0)$, zjistíme, že platí

$$\mu_t = \begin{cases} (0, 0.5, 0, 0.5) & \text{pro } t \text{ liché,} \\ (0.5, 0, 0.5, 0) & \text{pro } t \text{ sudé.} \end{cases} \quad (1.13)$$

Tato posloupnost tedy limitu nemá. Důležitou vlastností stacionární distribuce je, že nezávisí na počátečním stavu X_0 ani na čase t .

Na základě těchto dvou příkladů výše plyne jedno stěžejní tvrzení pro metodu Markov chain Monte Carlo.

Věta 1.3.2 *Necht' máme neredukovatelný a aperiodický řetězec. Potom pro něho existuje právě jedna stacionární distribuce.*

1.4 Čas prvního příchodu

Někdy nás může zajímat, za jak dlouho (po kolika krocích) přijdeme do určitého stavu $x \in \mathcal{X}$. První takový čas označujeme jako *čas prvního příchodu*.

Definice 1.4.1 (Čas prvního příchodu a čas prvního návratu) [8, sekce 1.5]

Předpokládejme, že Markovský řetězec (X_0, X_1, \dots) má konečný stavový prostor \mathcal{X} a matici přechodu P . Pro libovolné $x \in \mathcal{X}$ definujeme čas prvního příchodu jako

$$\tau_x := \min\{t \geq 1 : X_t = x\}. \quad (1.14)$$

Pokud $X_0 = x$, nazýváme τ_x čas prvního návratu.

Pro libovolné dva stavy x, y neredukovatelného řetězce platí $\mathbf{E}_x(\tau_y) = \mathbf{E}[\tau(y)|X_0 = x] < \infty$, kde znak \mathbf{E}_x značí střední hodnotu při počátku v x .

Uveď me zde také větu, která poskytuje spojitost mezi středním časem prvního příchodu a stacionární distribucí.

Věta 1.4.2 *Mějme markovský řetězec s maticí přechodu P a stacionární distribucí π . Pak střední počet kroků, aby se řetězec dostal do stavu $x \in \mathcal{X}$ je dán vztahem*

$$\mathbf{E}[\tau_x] = \frac{1}{\pi(x)}. \quad (1.15)$$

1.5 Reversibilita řetězce

Necht' je dán markovský řetězec se stavovým prostorem \mathcal{X} a maticí přechodu P . Předpokládejme, že existuje pravděpodobnostní distribuce na \mathcal{X} , která splňuje

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{pro všechna } x, y \in \mathcal{X}. \quad (1.16)$$

Rovnice (1.16) se nazývají rovnice detailní rovnováhy [8, sekce 1.6].

Věta 1.5.1 [8, sekce 1.6]

Necht' je dán markovský řetězec se stavovým prostorem X a maticí přechodu P . Libovolná pravděpodobnostní distribuce π splňující rovnice detailní rovnováhy (1.16) je stacionární pro P .

Důkaz: Pokud obě strany rovnice (1.16) vysčítáme přes všechny $y \in X$, dostaneme

$$\sum_{y \in X} \pi(y)P(y, x) = \sum_{y \in X} \pi(x)P(x, y) = \pi(x). \quad (1.17)$$

□

Tato vlastnost nám bude velice užitečná, neboť často je ověření rovnic detailní rovnováhy nejjednodušší cestou, jak najít stacionární distribuci řetězce.

Podíváme-li se blíže na to, co nám tato vlastnost říká, zjistíme, že pokud máme neredukovatelný řetězec a stav X_0 má stacionární rozdělení (tedy i každý následující stav má stacionární rozdělení), pak platí

$$\mathbf{P}_\pi\{X_0 = x_0, \dots, X_n = x_n\} = \mathbf{P}_\pi\{X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0\}. \quad (1.18)$$

Z tohoto důvodu řetězec, který splňuje (1.16) se nazývá *reversibilní* [8, sekce 1.6]. Symbol \mathbf{P}_π v (1.18) značí, že už počáteční stav pochází ze stacionární distribuce π .

Pokud bychom naopak chtěli odvodit (1.16) z (1.18), postup bude následující

$$\begin{aligned} P(i, j) &= \mathbf{P}(X_k = j | X_{k-1} = i, X_{k-2} = i_{k-2}, \dots, X_0 = i_0) = \\ &= \mathbf{P}(X_k = j | X_{k+1} = i, X_{k+2} = i_{k-2}, \dots, X_{2k+1} = i_0) = \\ &= \frac{\mathbf{P}(X_k = j, X_{k+1} = i, X_{k+2} = i_{k-2}, \dots, X_{2k+1} = i_0)}{\mathbf{P}(X_{k+1} = i, X_{k+2} = i_{k-2}, \dots, X_{2k+1} = i_0)} = \\ &= \frac{\pi(j)P(j, i)P(i, i_{k-2}) \dots P(i_1, i_0)}{\pi(i)P(i, i_{k-2}) \dots P(i_1, i_0)} = \frac{\pi(j)}{\pi(i)}P(j, i), \end{aligned} \quad (1.19)$$

kde jsme napřed využili Markovovu vlastnost, poté náš požadavek na reversibilitu a následně definici podmíněné pravděpodobnosti. Nakonec jsme využili definici stacionárního řetězce [6, kapitola 16].

Pro náhodnou procházku po čtverci plyne, že reversibilní obecně není. I když má rovnoměrnou stacionární distribuci, reversibilní je jen v jednom speciálním případě a to, pokud $p = 1/2$.

1.6 Příklady

Na závěr této kapitoly se podívejme na nějaké zajímavější příklady než byla jen náhodná procházka po čtverci. Budou to příklady, u kterých nelze sestavit matici přechodu nebo by to bylo velmi zdlouhavé kvůli početné množině stavového prostoru X .

Příklad 3 (Jezdec na šachovnici)

Mějme šachovnici o velikosti 8×8 polí a figurku jezdce, který se po ní pohybuje standardně. Uvažujme, že se vždy o následujícím poli rozhoduje rovnoměrně náhodně z možných polí.

- Jaká je pravděpodobnost výskytu jezdce v libovolném rohu šachovnice po dlouhé době?
- Pokud jezdec začne na poli 1A, kolik tahů průměrně udělá, aby se dostal zpět do tohoto pole?

8								
7								
6								
5								
4								
3								
2								
1	●							
	A	B	C	D	E	F	G	H

Obrázek 1.3: Šachovnice s koňem v rohu

8	2	3	4	4	4	4	3	2
7	3	4	6	6	6	6	4	3
6	4	6	8	8	8	8	6	4
5	4	6	8	8	8	8	6	4
4	4	6	8	8	8	8	6	4
3	4	6	8	8	8	8	6	4
2	3	4	6	6	6	6	4	3
1	2	3	4	4	4	4	3	2
	A	B	C	D	E	F	G	H

Obrázek 1.4: Možné tahy z jednotlivých polí

Řešení:

Na šachovnici je 64 políček, což znamená, že kdybychom se snažili sestavit matici přechodu a použít tak na tento příklad "hrubou sílu", dalo by nám toto řešení mnoho práce - přeci jenom bychom v tomto příkladu sestavili matici přechodu P o velikosti $|P| = 64 \times 64$. Bude tedy lepší použít nějakou jednodušší strategii.

Zapišme si do každého políčka na šachovnici, kolik možností má kůň z daného pole (obr. 1.4). Dále využijme rovnice detailní rovnováhy.

Pokud uvážíme dvojici políček, se stejnými počty možných tahů (např. $(3C, 5D)$ a $(7C, 5B)$), vyjde nám

$$\pi(3C)P(3C, 5D) = \pi(5D)P(5D, 3C)$$

$$\pi(3C)\frac{1}{8} = \pi(5D)\frac{1}{8} \Rightarrow \pi(3C) = \pi(5D)$$

a

$$\pi(7C)P(7C, 5B) = \pi(5B)P(5B, 7C)$$

$$\pi(7C)\frac{1}{6} = \pi(5B)\frac{1}{6} \Rightarrow \pi(7C) = \pi(5B).$$

Tedy políčka se stejným počtem možných přeskoků jsou stejně pravděpodobné. Jaký je vztah mezi políčky s různými počty možných přeskoků si ukážeme nyní např. na dvojicích $(1A, 3B)$ a $(6A, 5C)$:

$$\pi(1A)P(1A, 3B) = \pi(3B)P(3B, 1A)$$

$$\pi(1A)\frac{1}{2} = \pi(3B)\frac{1}{6} \Rightarrow 3\pi(1A) = \pi(3B)$$

a

$$\pi(6A)P(6A, 5C) = \pi(5C)P(5C, 6A)$$

$$\pi(6A)\frac{1}{4} = \pi(5C)\frac{1}{8} \Rightarrow 2\pi(6A) = \pi(5C).$$

Vyjádríme-li si nyní $\pi(1A)$, vyjde nám

$$\pi(1A) = \frac{\pi(1A)}{\sum_{i=1}^{64} \pi(i)} = \frac{\pi(1A)}{4\pi(1A) + 8\frac{3}{2}\pi(1A) + 20 * 2\pi(1A) + 16 * 3\pi(1A) + 16 * 4\pi(1A)} = \frac{1}{168}. \quad (1.20)$$

Označme jev $\omega = \{1A, 1H, 8A, 8H\}$ jako množinu stavů, kdy se jezdec nachází v rohu. Pak pro pravděpodobnost ω platí

$$\mathbf{P}(\omega) = 4\pi(1A) = \frac{4}{168} = \frac{1}{42}. \quad (1.21)$$

Známe-li stacionární distribuci pro pole 1A, je jednoduché odpovědět na druhou otázku, neboť pro střední čas návratu platí z věty 1.4.2

$$\mathbf{E}[t_{1A}] = \frac{1}{\pi(1A)} = 168. \quad (1.22)$$

Tuto střední dobu prvního návratu si můžeme snadno ověřit simulací. Necháme si vygenerovat vzorky doby prvního návratu z rohu 1A zpět. Pokud bychom za náhodnou veličinu volili X jako dobu prvního návratu, obdržíme průměrováním této náhodné veličiny sice blízký odhad, ovšem s obrovským rozptylem (nám vyšlo rozmezí délky trajektorie v intervalu (2, 2114)). Proto za náhodnou veličinu zvolíme $(\bar{X}_{M_i})_{i=1}^N$, tedy průměr M vzorků. Zvolíme např. $N = 150$, $M = 150$ a získáváme

$$\widehat{\mathbf{E}[t_{1A}]} = \frac{1}{N} \sum_{i=1}^N \bar{t}_{1A_i} = 165.98, \quad (1.23)$$

s rozptylem

$$\widehat{\mathbf{D}[t_{1A}]} = 293.99, \quad (1.24)$$

což je ve shodě s naší teoretickou hodnotou.

Zdrojový kód 1.1: Zdrojový kód koně na šachovnici

```

1 pocet_vzorku_na_prumer=150;
2 pocet_opakovani = 150;
3 mnozina_prumeru = zeros(1, pocet_opakovani);
4 maximum = 0;
5 minimum = 1000;
6 for v = 1 : pocet_opakovani
7     l=zeros(1,pocet_vzorku_na_prumer);
8     for j = 1 : pocet_vzorku_na_prumer
9         x = 1;
10        y = 1;
11        t = 0;
12        while x ~= 1 || y ~= 1 || t == 0
13            cislo = unifrnd(0, 8);
14            if cislo < 1
15                a = x - 2;
16                b = y + 1;
17            elseif 1 <= cislo && cislo < 2
18                a = x - 2;
19                b = y - 1;
20            elseif 2 <= cislo && cislo < 3
21                a = x - 1;
22                b = y + 2;
23            elseif 3 <= cislo && cislo < 4

```

```

24         a = x + 1;
25         b = y + 2;
26     elseif 4 <= cislo && cislo < 5
27         a = x + 2;
28         b = y + 1;
29     elseif 5 <= cislo && cislo < 6
30         a = x + 2;
31         b = y - 1;
32     elseif 6 <= cislo && cislo < 7
33         a = x - 1;
34         b = y - 2;
35     elseif 7 <= cislo && cislo < 8
36         a = x + 1;
37         b = y - 2;
38     end
39     if a >= 1 && b >= 1 && a <= 8 && b <= 8
40         x = a;
41         y = b;
42         t = t + 1;
43     end
44 end
45 if maximum < t
46     maximum = t;
47 end
48 if minimum > t
49     minimum = t;
50 end
51 l(j) = t;
52 end
53 mnozina_prumeru(v) = mean(l);
54 end
55 prumer = mean(mnozina_prumeru)
56 rozptyl = var(mnozina_prumeru)
57 odchylka = sqrt(rozptyl)
58 maximum
59 minimum

```


Kapitola 2

Markov Chain Monte Carlo: Metropolisovy a Glauberovy řetězce

2.1 Trocha historie

Informace v této sekci jsou čerpány z [3, sekce 1.1]. Historie simulačních metod sahá přibližně do poloviny 20. století. Již před příchodem počítačů sice bylo provedeno několik pozoruhodných simulací náhodných procesů, nicméně jejich širší praktické užití si muselo počkat až na vynález počítače, kterého bylo pro tento účel využito téměř okamžitě.

Původ názvu „Monte Carlo“ není nijak technický - kolem roku 1950 byl gambling na většině místech na světě zakázaný, avšak našlo se jisté město, kde gambling v podobě casina vzkvétal - Monte Carlo. Pojem *Monte Carlo* se však brzy stal pouhým technickým termínem simulační metody náhodných procesů.

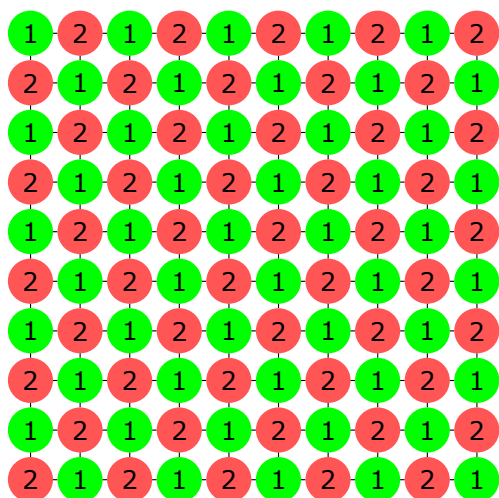
Počátky metody *Markov Chain Monte Carlo* na sebe nenechaly dlouho čekat a tato metoda přichází na svět krátce po obyčejné Monte Carlo metodě v Los Alamu, jednom z mála měst, ve kterých v té době byly počítače dostupné. Metropolis (1953) simuloval kapalinu v rovnováze s její plynnou fází [10]. Možný způsob, jak zjistit tuto termodynamickou rovnováhu, je simulovat dynamiku systému a nechat algoritmus běžet, dokud systém nedosáhne rovnováhy. Touto cestou nepotřeboval nasimulovat přesnou dynamiku; potřeboval jen nasimulovat nějaký Markovský řetězec, který by měl stejnou rovnovážnou distribuci. Simulace tohoto schématu, který použil Metropolis, se nazývá *Metropolisův algoritmus*. Jakmile se počítače stávaly dostupnějšími, Metropolisův algoritmus začal být stále více využíván ve fyzice a chemii, ale matematiky (statistiky) začal být využíván až po roce 1990. Hastings roku 1970 dále zobecnil Metropolisův algoritmus a toto schéma simulací nese název *Metropolis-Hastingsův algoritmus* [7]. Dalšími, kteří zkoumali simulační metody s využitím náhodných procesů byli bratři Donald Geman a Stuart Geman (1984). Ti přišli se speciálním případem Metropolis-Hastingsova algoritmu, zřejmě bez znalosti dřívějších zjištění. Simulace využívající tohoto schématu nese název *Gibbsovo vzorkování* [4].

2.2 Úvod do metody Markov Chain Monte Carlo

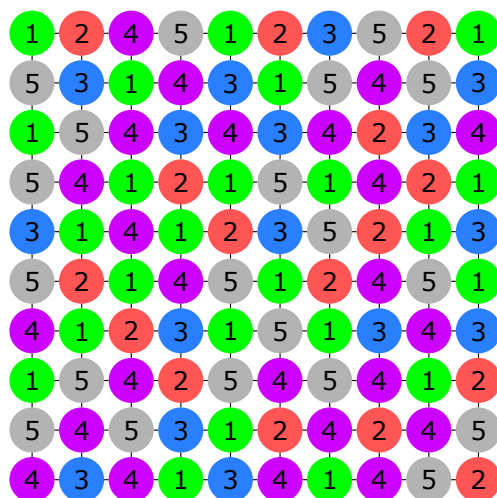
Navážeme nyní na kapitolu 1, kde jsme se seznámili se základními pojmy a vlastnostmi Markovských řetězců. Zabývali jsme se úlohami, kde stavový prostor nebyl příliš velký nebo se dalo vhodně využít vlastností daného řetězce.

2.2.1 Motivační příklad - správné q-barvení

Uvažme množinu barev $Q = \{1, 2, \dots, q\}$ a graf $G = (V, O)$, kde V je množina vrcholů grafu G a O je množina dvojic spolu komunikujících stavů (množina možných přechodů mezi stavy). Zdefinujme **správné q-barvení** grafu $G = (V, O)$ jako přiřazení barev množiny Q k jednotlivým vrcholům z V . Po správném q -barvení požadujeme, aby žádné dva sousední (spolu komunikující) vrcholy neměly stejnou barvu. Sestavení vzorků je pro nás velice přínosné, neboť skrze simulace můžeme zkoumat různé průměrné charakteristiky barvení, např. s jakou pravděpodobností mají všechny rohové vrcholy mřížky stejnou barvu [8, sekce 3.1].



Obrázek 2.1: Vzorek správného 2-barvení



Obrázek 2.2: Vzorek správného 5-barvení

Proto hledáme metodu pro vzorkování z \mathcal{X} , množiny všech správných q -barvení.

Pro některé typy grafů (např. stromy) existují jednoduché rekurzivní metody pro generování správného barvení, zatímco pro jiné může být náročnou překážkou přímo zkonstruovat náhodný vzorek.

Právě pro tento druhý případ se nabízí využít Markovské řetězce pro konstrukci vzorku. Předpokládejme, že (X_t) je řetězec se stavovým prostorem \mathcal{X} a s rovnoměrnou stacionární distribucí na \mathcal{X} . Z konvergenční věty (věta 3.2.3) dále plyne, že (X_t) je přibližně rovnoměrně rozdělen pro dostatečně velké t . Příklady správných q -barvení na mřížce 10×10 můžeme vidět na obr. 2.1 a 2.2. Zatímco pro dvoubarevné správné q -barvení lze jednoduše odhadnout, že stavový prostor \mathcal{X} bude pouze dvouprvkový, pro více barevné q -barvení už není jednoduché přesně určit velikost stavového prostoru.

Metodu, která k sestavení vzorků z dané pravděpodobnostní distribuce využívá Markovské řetězce, nazýváme metoda Markov Chain Monte Carlo. Pro správné q -barvení je stacionární distribuce rovnoměrná.

Shrneme-li tedy naše snažení, jde nám o to, abychom z dané pravděpodobnostní distribuce π na \mathcal{X} sestavili Markovský řetězec, který by měl svou stacionární distribuci rovnou π . Pokud takový řetězec můžeme sestavit, potom pro dostatečně velké t je distribuce X_t blízka π . V této kapitole se dále podíváme na metody, které nám pomohou najít Markovské řetězce ze zadané stacionární distribuce.

Praktické využití správného q-barvení

Správné q -barvení není samozřejmě jen učebnicovým příkladem, pro který by bylo snadné sestavit markovský řetězec, který by měl svou stacionární distribuci rovnoměrnou. Lze najít několik uplatnění z

praxe, kde se dá správného q-barvení využít. Jmenujme zde statistickou fyziku, která uplatňuje správné q-barvení v tzv. **Pottově modelu** [11, sekce 1.1].

Atomy/molekuly krystalu mohou být představovány vrcholy $V \subset \mathbb{Z}^d$. Každému vrcholu přiřazujeme magnetický spin - jednu z barev množiny Q . Ve feromagnetických materiálech mají spiny sousedních vrcholu tendenci se rovnat. Naopak u anti-feromagnetických materiálu mají tendenci se lišit. Síla těchto tendencí je závislá na teplotě a naplno jsou tyto tendence naplněny pro teplotu $T \rightarrow 0$ [11, sekce 1.1].

Z toho plyne, že naše q-barvení na rovnoměrné stacionární distribuci je příklad q-stavového Pottova modelu anti-feromagnetických materiálů při teplotě $T \rightarrow 0$.

Metoda Markov Chain Monte Carlo

Stojíme tedy před problémem, kdy známe pravděpodobnostní distribuci π a potřebujeme sestavit řetězec, který bude mít svou stacionární distribuci rovnou dané pravděpodobnostní distribuci π .

Metoda **Markov Chain Monte Carlo** (MCMC) umožňuje takovou úlohu řešit. Je jasné, že řetězců, které budou odpovídat dané stacionární distribuci bude celá řada, ba dokonce nekonečně mnoho. Kterou ale zvolit? Pro metodu MCMC je stěžejní věta 1.3.2, podle níž pro libovolný neredukovatelný a aperiodický řetězec existuje právě jedna stacionární distribuce. Proto budeme také požadovat splnění této podmínky neredukovatelnosti. Navíc budeme po řetězci také požadovat, aby byl reversibilní. Budeme tedy chtít sestavit matici přechodu P za požadavku splnění následujících podmínek [2, sekce 2]:

1. π je stacionární distribucí pro P , proto požadujeme splnění stacionarity $\pi = \pi P$.
2. Požadujeme také splnění reversibility, tedy s ní spojené rovnice detailní rovnováhy $\pi(x)P(x, y) = \pi(y)P(y, x)$ pro všechna $x, y \in \mathcal{X}$.

Metoda MCMC je simulační metoda, která je založena na výběru tzv. **náhodných vzorků**.

2.3 Metropolisovy řetězce a Metropolisův algoritmus

Nyní se už podívejme na konkrétní způsob sestavení markovského řetězce, který bude mít stacionární distribuci rovnou π . Princip **Metropolisova algoritmu**, resp. obecnějšího **Metropolis-Hastingsova algoritmu**, spočívá v tom, že při daném řetězci se stavovým prostorem \mathcal{X} a libovolnou stacionární distribucí π nám umožní tento řetězec modifikovat takovým způsobem, že má tento nově vzniklý řetězec stacionární distribuci rovnou π . Odvození si provedeme nejprve pro symetrický původní řetězec tak, jak to provedl roku 1953 Metropolis (Metropolisův algoritmus), a z nějž je také lépe patrné, jakou konečnou podobu tento modifikovaný řetězec má. Posléze toto odvození zobecníme na obecný základní řetězec, jak Metropolisův řetězec roku 1970 zobecnil Hastings (Metropolis-Hastingsův algoritmus).

2.3.1 Symetrický původní řetězec Ψ

Odvození Metropolisova algoritmu následuje postup, jaký je uveden v [8, sekce 3.2]. Mějme danou libovolnou pravděpodobnostní distribuci π na \mathcal{X} a symetrický řetězec Ψ se stavovým prostorem \mathcal{X} a maticí přechodu ψ . Pokud je Ψ symetrický, pak pro jeho matici přechodu platí

$$\psi(x, y) = \psi(y, x) \quad \text{pro všechny } x, y \in \mathcal{X}, x \neq y. \quad (2.1)$$

Z toho plyne, že Ψ je také reversibilní vzhledem k rovnoměrné distribuci μ na \mathcal{X} , což znamená platnost

$$\mu(x)\psi(x, y) = \mu(y)\psi(y, x) \quad \text{pro všechny } x, y \in \mathcal{X}, x \neq y. \quad (2.2)$$

Nyní ukážeme, jak modifikovat přechody vzhledem k Ψ tak, abychom získali řetězec, který bude mít stacionární distribuci π .

Modifikujme si tento stávající řetězec tímto způsobem:

- Pokud jsme ve stavu x , pak následující tah bude generován z distribuce $\Psi(x, \cdot)$.
- Pokud je navrhovaný stav y , pak je nový stav akceptován s pravděpodobností $a(x, y)$ a s pravděpodobností $1 - a(x, y)$ je zamítnut. To znamená, že s pravděpodobností $1 - a(x, y)$ zůstáváme ve stavu x .

Odmítnutí pohybu sice zpomaluje řetězec, ale může být nezbytné pro obdržení správné stacionární distribuce. Podívejme se, jak zvolit pravděpodobnost přijmutí stavu y $a(x, y)$. Matice přechodu P nového řetězce má následující tvar:

$$P(x, y) = \begin{cases} \psi(x, y)a(x, y), & \text{pokud } y \neq x \\ 1 - \sum_{z:z \neq x} \psi(x, z)a(x, z), & \text{pokud } y = x. \end{cases} \quad (2.3)$$

Podle věty 1.3.2 má matice přechodu P stacionární distribuci π , pokud

$$\pi(x)\psi(x, y)a(x, y) = \pi(y)\psi(y, x)a(y, x) \quad \text{pro všechny } x \neq y. \quad (2.4)$$

Zároveň platí (2.1) a odtud

$$b(x, y) = \pi(x)a(x, y) = \pi(y)a(y, x) = b(y, x) \quad \text{pro všechny } x \neq y. \quad (2.5)$$

Protože $a(x, y)$ je pravděpodobnost, musí platit $a(x, y) \leq 1$ a celkem pro $b(x, y)$ máme

$$b(x, y) \leq \pi(x). \quad (2.6)$$

$$b(x, y) = b(y, x) \leq \pi(y). \quad (2.7)$$

Jelikož s klesající hodnotou $a(x, y)$ klesá také rychlost řetězce, chtěli bychom, aby hodnoty $a(x, y)$ byly co největší. To znamená, že požadujeme také co nejvyšší hodnoty $b(x, y)$. Proto vrchní hranice b bude

$$b^*(x, y) = b^*(y, x) := \min\{\pi(x), \pi(y)\} \quad \text{pro všechny } x, y \in X. \quad (2.8)$$

Dosažením (2.8) do (2.5) a vyjádřením $a(x, y)$ dostáváme vztah pro pravděpodobnost akceptování pohybu ze stavu x do stavu y :

$$a(x, y) = \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}. \quad (2.9)$$

Z tohoto vztahu vidíme, že pro konstantní pravděpodobnostní distribuci π akceptujeme navrhovaný stav y s pravděpodobností $a(x, y) = 1$. Dosažením tohoto vztahu do (2.3) dostáváme

$$P(x, y) = \begin{cases} \psi(x, y) \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}, & \text{pokud } y \neq x, \\ 1 - \sum_{z:z \neq x} \psi(x, z) \min\left\{\frac{\pi(z)}{\pi(x)}, 1\right\}, & \text{pokud } y = x. \end{cases} \quad (2.10)$$

2.3.2 Obecný původní řetězec Ψ

Odvození obecného Metropolis-Hastingsova algoritmu následuje postup uvedený v [8, sekce 3.2]. Metropolisův řetězec samozřejmě nemusí být symetrický. Pro obecnou neredukovatelnou matici přechodu Ψ a libovolnou pravděpodobnostní distribuci π na \mathcal{X} je postup úplně stejný, jen musíme brát v potaz skutečnost, že nyní

$$\psi(x, y) \neq \psi(y, x). \quad (2.11)$$

Z toho důvodu po úpravách dostáváme pro $a(x, y)$ následující vztah:

$$a(x, y) = \min \left\{ \frac{\pi(y)\psi(y, x)}{\pi(x)\psi(x, y)}, 1 \right\}. \quad (2.12)$$

A konečně matice přechodu P má tvar

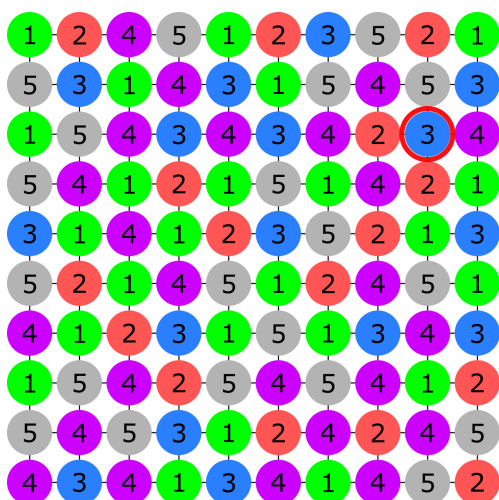
$$P(x, y) = \begin{cases} \psi(x, y) \min \left\{ \frac{\pi(y)\psi(y, x)}{\pi(x)\psi(x, y)}, 1 \right\}, & \text{pokud } y \neq x, \\ 1 - \sum_{z:z \neq x} \psi(x, z) \min \left\{ \frac{\pi(z)\psi(z, x)}{\pi(x)\psi(x, z)}, 1 \right\}, & \text{pokud } y = x. \end{cases} \quad (2.13)$$

Za počáteční řetězec se volí takový, který se dá jednoduše sestavit. Z něho lze tedy výše uvedeným způsobem získat modifikovaný, který bude mít stacionární distribuci rovnou π . Velmi přínosnou vlastností Metropolisova algoritmu je fakt, že k sestavení řetězce potřebujeme znát pouze poměr $\frac{\pi(x)}{\pi(y)}$.

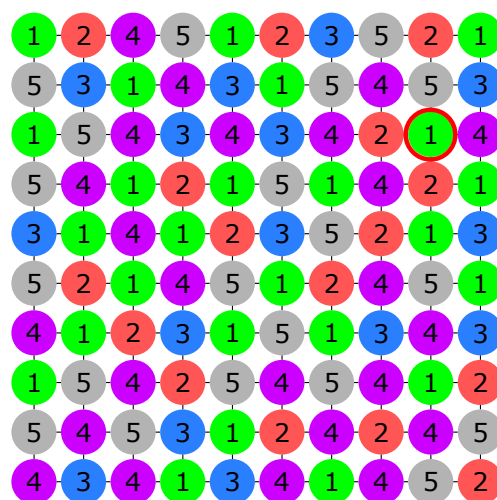
Ukážeme si na příkladě nejjednodušší možnost užití Metropolisova algoritmu, kdy požadujeme, aby stacionární distribuce byla rovnoměrná a již prvotně sestavený řetězec Ψ byl symetrický. Z toho plyne, že stav y zvolený z distribuce $\Psi(x, \cdot)$ bude schválen s pravděpodobností $a(x, y) = 1$.

Příklad 4 (Princip Metropolisova algoritmu na správném 5-barvení)

Nechť je dána mřížka 10×10 , množina barev $Q = \{1, \dots, 5\}$ a rovnoměrná pravděpodobnostní distribuce π . Nechť se nacházíme ve stavu x a byl vybrán bod v rovnoměrně s pravděpodobností $1/|V|$ (obr. 2.3).



Obrázek 2.3: Konfigurace x správného 5-barvení s vybraným vrcholem v



Obrázek 2.4: Konfigurace y správného 5-barvení

Vybranému vrcholu nyní můžeme přiřadit buď prvek $\{1\}$, tedy obarvit na zeleno a změnit stav x za stav y , nebo mu nechat přiřazený prvek $\{3\}$, tedy modrou barvu (a zůstat tak ve stavu x). Abychom rozhodli o tom, jaké obarvení vrcholu v bude upřednostněno, zvolíme rovnoměrně náhodně barvu z celé Q (každá barva má v tomto případě pravděpodobnost na výběr $p = \frac{1}{|Q|}$). Pokud bude zvolen prvek $\{1\}$, pak obarvíme tento vrchol na zeleno (obr. 2.4). Pokud bude zvolen jakýkoli jiný prvek, necháme řetězec ve stavu x . V tomto případě je zvolená aktualizace vrcholu akceptována vždy, neboť v tomto případě je π rovnoměrná a navíc sestrojený řetězec Ψ je symetrický.

Obecně tedy v Metropolisově algoritmu pro správné q -barvení s rovnoměrnou pravděpodobnostní distribucí postupujeme takto:

- Zvolíme vrchol $z \in V$ rovnoměrně náhodně, tedy každý vrchol má pravděpodobnost na výběr $\frac{1}{|V|}$.
- Zvolíme barvu q také rovnoměrně náhodně, tedy pravděpodobnost výběru každé barvy bude $\frac{1}{|Q|}$.
- Pokud zvolená barva q náleží do množiny přípustných barev pro správné q -barvení, potom přebarvíme vrchol touto barvou, jinak obarvení zamítáme a řetězec zůstává ve stavu x .

2.4 Glauberova dynamika

Glauberova dynamika je další užitečný způsob vzorkování, nicméně tento algoritmus se týká pouze těch řetězců, jejichž stavový prostor je množina tvaru S^V , kde V je vrcholová množina grafu a S je konečná množina. Prvky množiny S^V se nazývají konfigurace a jsou to funkce $f : V \rightarrow S$ (každému vrcholu V je přiřazen prvek množiny S).

Je-li dána pravděpodobnostní distribuce π na prostoru konfigurací, potom **Glauberova dynamika** je Markovský řetězec, který má stacionární distribuci π . Tento řetězec je často označován jako **Gibbsův vzorkovač**.

Gibbsovo vzorkování (Glauberova dynamika) je jen lehká modifikace Metropolis-Hastingsova algoritmu. Hlavním rozdílem mezi oběma algoritmy je skutečnost, že v Gibbsově vzorkovači budeme vybírat aktualizaci vrcholu pouze z přípustných hodnot, což znamená, že konvergence tohoto řetězce je značně rychlejší [3, sekce 1.12], [5, sekce 1.4], [8, sekce 3.3], [9, sekce 2.5.6].

Definice 2.4.1 (Obecná definice Glauberovy dynamiky) [8, sekce 3.3]

Nechť \mathcal{X} je podmnožina S^V a S i V jsou konečné množiny. Nechť π je pravděpodobnostní distribuce, jejíž nosič je \mathcal{X} . Glauberova dynamika pro π je reversibilní Markovský řetězec se stavovým prostorem \mathcal{X} , stacionární distribucí π a pravděpodobnostmi přechodů, které budou popsány níže.

Pohyb Glauberova řetězce ze stavu x se děje následovně:

- vrchol v je vybrán náhodně z rovnoměrného rozdělení z množiny V ,
- nový stav y je zvolen vzhledem k míře π za podmínky, že všechny vrcholy kromě v zůstávají se stejným přiřazeným prvkem z množiny S , jako tomu bylo ve stavu x (tuto míru označíme $\pi^{x,v}$).

Mějme pro $x \in \mathcal{X}$ a $v \in V$ množinu přípustných stavů

$$\mathcal{X}(x, v) = \{y \in \mathcal{X} : y(w) = x(w) \text{ pro všechna } w \neq v\}. \quad (2.14)$$

Definujme dále distribuci π podmíněnou množinou $\mathcal{X}(x, v)$

$$\pi^{x,v}(y) = \pi(y|\mathcal{X}(x, v)) = \begin{cases} \frac{\pi(y)}{\pi(\mathcal{X}(x, v))} & \text{pokud } y \in \mathcal{X}(x, v), \\ 0 & \text{pokud } y \notin \mathcal{X}(x, v). \end{cases} \quad (2.15)$$

Slovy to tedy znamená, že $\pi^{x,v}(y)$ je pravděpodobnost, že se přesuneme do konfigurace y za předpokladu, že se v dané chvíli nacházíme v konfiguraci x a byl vybrán vrchol v jako vrchol, ve kterém dojde ke změně. Například pokud množina možných konfigurací je $\mathcal{X}(x, v) = \{x, y, z\}$, pak $\pi(\mathcal{X}(x, v)) = \pi(x) + \pi(y) + \pi(z)$. Tedy nacházíme-li se ve stavu x a zvolíme vrchol v , pak tomuto vrcholu vybíráme prvek množiny S pouze z množiny přípustných prvků pro tento vrchol.

Dále platí, že Glauberova dynamika je vždy reversibilní a distribuce π stacionární. Toto tvrzení si dále ukážeme na dvou příkladech.

2.4.1 Reversibilita Glauberovy dynamiky a stacionarita π

Příklad 5 (Správné q -barvení)[8, sekce 3.3]

Uvažme znovu správné q -barvení grafu $G = (V, O)$. Množina S z obecné definice je v tomto případě množina barev $Q = \{1, 2, \dots, q\}$ a správné q -barvení je prvek x z množiny $\{1, 2, \dots, q\}^V$. Je to množina všech funkcí z V do $\{1, 2, \dots, q\}$, pro které platí $x(v) \neq x(w)$ pro všechny sousední vrcholy $\{v, w\}$. Sestavíme Markovský řetězec na množině správných q -barvení grafu G .

Mějme konfiguraci x a vrchol v . Řekneme, že barva $j \in Q$ je **přípustná** pro vrchol v , pokud žádný ze sousedních vrcholů v nemá přiřazenou barvu j .

Konfiguraci x (správné q -barvení) můžeme aktualizovat následovně:

- vrchol $v \in V$ vybereme náhodně z rovnoměrného rozdělení,
- barvu j vybereme také náhodně z rovnoměrného rozdělení, avšak jen z množiny přípustných barev pro v ,
- přebarvíme vrchol v barvou j .

Nyní ukážeme, že výsledný řetězec má rovnoměrnou stacionární distribuci π . Uvažujme, že přecházíme mezi konfiguracemi x a y . Tyto konfigurace jsou ale shodné a liší se pouze v obarvení vrcholu v . Pravděpodobnost, že přejdeme z konfigurace x do konfigurace y je $P(x, y) = |V|^{-1} |A_v(x)|^{-1}$, kde $A_v(x)$ je množina přípustných barev pro v za konfigurace x . Pokud přecházíme z konfigurace y do konfigurace x , množina $A_v(y)$ je stejně početná jako množina $A_v(x)$ (jak je vidět na obr. 2.3 a 2.4, ať obarvíme vybraný vrchol jakoukoli barvou, stále vybíráme ze stejné množiny přípustných barev). Z toho plyne, že díky platnosti $P(x, y) = P(y, x)$ jsou detailní bilanční rovnice (1.16) splněny právě rovnoměrnou distribucí π .

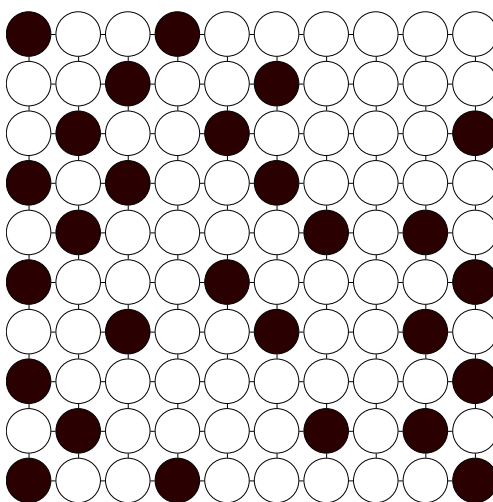
Takto sestavený řetězec se nazývá **Glauberova dynamika pro správné q -barvení**.

Příklad 6 (Hardcore model)[8, sekce 3.3]

Mějme znovu graf $G = (V, O)$. Nyní budeme chtít přiřazovat vrcholům částici za požadavku, že žádné dvě částice spolu nesousedí. Množina S z obecné definice tedy nyní bude množina $\{0, 1\}$, **hardcore konfigurace** nyní představuje prvek x z množiny $\{0, 1\}^V$, který splňuje podmínku $x(v)x(w) = 0$ pro všechna $v, w \in V$ takové, že $v \sim w$. $v \sim w$ zde značí, že v a w spolu sousedí. Toto označení budeme používat i nadále. Pokud $x(v) = 1$ řekneme, že vrchol v je **okupován**. V opačném případě vrchol v nazveme **prázdným**.

Mezi konfiguracemi můžeme přecházet následovně:

- Vrchol v je vybrán z V náhodně z rovnoměrného rozdělení a bez ohledu na současný stav v (opět je tedy pravděpodobnost výběru v rovna $|V|^{-1}$).
- Pokud je nějaký soused v okupován, v je opuštěn neobsazen a pokračujeme dále. Pokud žádný ze sousedů v okupován není, obsadíme v s pravděpodobností $1/2$ a se stejnou pravděpodobností necháme vrchol volný.



Obrázek 2.5: Příklad hardcore konfigurace

Příklad takové hardcore konfigurace vidíme na obr. 2.5, kde prázdná pole představují neobsazené vrcholy, černé pole představují částice.

Opět je výsledný řetězec reversibilní a má rovnoměrnou stacionární distribuci. Obdobným způsobem jako tomu bylo v příkladu 5 si ukážeme, proč tomu tak je. Necht' konfigurace x označuje situaci, kdy vrchol v obsazen není, naopak y označuje situaci, kdy vrchol v obsazen je (za předpokladu, že sousední vrcholy vrcholu v jsou neobsazeny, aby mohly nastat obě situace). Pravděpodobnost přechodu z x do y je rovna $|V|^{-1}2^{-1}$. Přechod z y do x je stejně pravděpodobný jako opačný přesun, tedy platí $P(x, y) = P(y, x)$, z čehož znovu plyne, že jsou rovnice detailní rovnováhy (1.16) splněny právě rovnoměrnou distribucí π .

2.4.2 Porovnání Glauberovy dynamiky a Metropolisova řetězce

Předpokládejme, že π je pravděpodobnostní distribuce na stavovém prostoru S^V , kde S je konečná množina a V vrcholová množina grafu $G = (V, O)$. Glauberův řetězec můžeme definovat vždy přesně takovým způsobem, jaký jsme si popsali výše. Na druhé straně předpokládejme, že máme řetězec, který vybírá vrchol v z V náhodně a má nějaký mechanismus na aktualizaci konfigurace pro vrchol v (například vybírá prvek z S také náhodně). Tento řetězec nemusí mít stacionární distribuci π , ale může být modifikován pomocí Metropolisova algoritmu tak, abychom získali řetězec se stacionární distribucí π . Takový řetězec může být podobný Glauberově, ale nemusí být přesně totožný.

Příklad 7 (správné q-barvení)

Napišme si schéma obou algoritmů na správném q-barvení. Metropolisův řetězec postupuje např. takovýmto způsobem:

- vrchol v zvolíme z V rovnoměrně náhodně,
- Vybereme barvu $j \in Q$ rovnoměrně náhodně (takto definujeme původní řetězec Ψ - v tomto případě se nejedná ani o správné q-barvení),
- tento řetězec upravíme Metropolisovým pravidlem, aby se jednalo o správné q-barvení s rovnoměrnou stac. distribucí na X :
barvu na aktualizaci připouštíme s pravděpodobností $p = 1$, pokud splňuje pravidlo pro správné q-barvení, jinak ji zamítáme.

Naproti tomu Glauberova dynamika pracuje na takovémto principu:

- vrchol v zvolíme z V rovnoměrně náhodně,
- barvu j zvolíme taktéž rovnoměrně náhodně, nicméně pouze z množiny přípustných barev,
- přebarvíme bod v zvolenou barvou j .

Při porovnání těchto dvou řetězců zjistíme, že se liší při výběru obarvení vrcholu v . Pokud jsme tedy zvolili vrchol v a a je počet možných barev pro tento zvolený vrchol, potom je pravděpodobnost na setrvání ve stejném stavu pro Glauberův řetězec rovna $\frac{1}{a}$ a pro Metropolisův je tato pravděpodobnost rovna $\frac{1+q-a}{q}$. Pro Glauberovu dynamiku je tento fakt zřejmý; pro Metropolisův řetězec je to tak proto, že vybíráme ze všech barev, tedy máme q možností. Z nich a je přípustných a zároveň jedna barva z množiny a je současný stav. Proto stav měníme s pravděpodobností $\frac{a-1}{q}$.

Příklad 8 (Hardcore řetězce)

Uvažujme znovu příklad, kdy máme graf $G = (V, O)$ a vrcholy obsazujeme částicemi. Stejně jako v příkladě 7 si porovnejme oba algoritmy. Metropolisův řetězec bude mít takovou podobu:

- vrchol v volíme rovnoměrně náhodně z množiny V ,
- zvolíme náhodně konfiguraci, jestli vrchol zaplníme nebo nikoliv,
- pokud zaplnění proběhlo v rámci pravidel hardcore konfigurace, pak tuto aktualizaci provedeme.

Glauberova dynamika funguje tímto způsobem:

- vrchol v volíme rovnoměrně náhodně z množiny V ,
- konfiguraci volíme také náhodně, ale pouze z množiny možných konfigurací,
- konfiguraci potvrzujeme.

V tomto případě jsou Glauberova dynamika i Metropolisův řetězec shodné.

Doteď jsme si představili pouze příklady, ve kterých byla zadaná pravděpodobnostní distribuce π rovnoměrná. Jak jsme zmínili už zpočátku kapitoly, pravděpodobnostní distribuce být rovnoměrná nemusí a ve většině případů ani není. Nyní si tedy představíme dva další příklady, v nichž požadovaná pravděpodobnostní distribuce rovnoměrně zadaná není - **Hardcore model s prchavostí** a **Isingův model**. Stále však nebudeme schopni určit ani velikost stavového prostoru \mathcal{X} , a tak je třeba π zadat nějakým vhodnějším způsobem než vektorem hodnot. Výhodou pro nás je, že jak Metropolisův algoritmus tak i Glauberova dynamika jsou zadány pouze pomocí poměru $\frac{\pi(x)}{\pi(y)}$, což znamená, že musíme být pouze schopni určit, kolikrát je pravděpodobnější stav x než stav y . Na oba naše příklady použijeme Glauberovu dynamiku, neboť její konstrukce je jednodušší.

2.4.3 Hardcore model s prchavostí

Informace k tomuto příkladu pocházejí z [8, sekce 3.3]. Mějme graf $G = (V, O)$ a množinu \mathcal{X} jako množinu hardcore konfigurací na G . Potom definujeme **hardcore model s prchavostí** λ , kde $\lambda > 0$, jako pravděpodobnostní distribuci π na hardcore konfiguracích $x \in \{0, 1\}^V$

$$\pi(x) = \begin{cases} \frac{\lambda^{\sum_{v \in V} x(v)}}{Z(\lambda)} & \text{pokud } x(v)x(w) = 0 \text{ pro všechna } \{v, w\} \text{ takové, že } v \sim w, \\ 0 & \text{jinak.} \end{cases} \quad (2.16)$$

Faktor $Z(\lambda) = \sum_{x \in \mathcal{X}} \lambda^{\sum_{v \in V} x(v)}$ je normalizační konstantou π . Budeme-li si chtít odvodit řetězec pro tento model pomocí Glauberovy dynamiky, potom pro vybraný vrchol v , který není obsazen částicí platí, že pokud již nějaký ze sousedních vrcholů obsazen částicí je, pak tento vrchol zůstává prázdný. Pokud žádný ze sousedních vrcholů částicí obsazen není, pak obsadíme vybraný vrchol s pravděpodobností

$$p = \frac{\pi(y)}{\pi(x) + \pi(y)} = \frac{\lambda\pi(x)}{\pi(x) + \lambda\pi(x)} = \frac{\lambda}{1 + \lambda}. \quad (2.17)$$

Algoritmus pro tento případ tedy probíhá následujícím způsobem:

- vrchol w vybereme z V rovnoměrně náhodně,
- pokud existuje $w' \sim w$ takové, že $X_t(w') = 1$, potom $X_{t+1}(w) := 0$;
pokud pro každé $w' \sim w$ platí $X_t(w') = 0$, potom

$$X_{t+1}(w) := \begin{cases} 1 & \text{s pravděpodobností } \frac{\lambda}{1+\lambda}, \\ 0 & \text{s pravděpodobností } \frac{1}{1+\lambda}. \end{cases} \quad (2.18)$$

Navíc platí $X_{t+1}(v) = X_t(v)$ pro všechny $v \neq w$.

Symbol X_t zde značí konfiguraci v čase t , tedy $X_t(v)$ značí hodnotu vrcholu v v čase t . Při limitě $\lambda \rightarrow 0$ neobsazujeme žádný vrchol částicí, naopak všechny vrcholy postupně zůstávají neobsazeny. Pro $\lambda = 1$ tento model přechází na klasický hardcore model a při limitě $\lambda \rightarrow \infty$ naopak každý vhodně zvolený vrchol obsazujeme, ovšem natrvalo a nedáváme tak žádnou možnost jakýmkoli změnám. Znamená to, že po určitém počtu kroků se řetězec zastaví v jednom vzorku a nehne se z místa [8, sekce 3.3].

2.4.4 Isingův model

Informace k Isingovu modelu pocházejí z [8, sekce 3.3]. Uvažujme nyní graf $G = (V, O)$, kde každému vrcholu $v \in V$ přiřadíme *spin*, který pro nás bude reprezentován prvkem množiny $\{-1, 1\}$. Tento spin pro všechny vrcholy $v \in V$ označíme $\sigma(v)$. Každá konfigurace tedy bude představovat prvek z množiny $\mathcal{X} = \{-1, 1\}^V$. Pravděpodobnostní distribuce na \mathcal{X} se nazývá *spinový systém*.

Isingův model je model, který popisujeme pouze prostřednictvím dvojic sousedních vrcholů. Může být reprezentován např. umístěním magnetických dipólů na vrcholy grafu. V tomto systému definujeme *energii* konfigurace σ jako

$$H(\sigma) = - \sum_{\substack{v, w \in V \\ v \sim w}} \sigma(v)\sigma(w). \quad (2.19)$$

Je zřejmé, že čím více se v systému nachází sousedních vrcholů s nesouhlasným spinem, tím je energie tohoto systému větší.

Gibbsova distribuce odpovídající energii H je pravděpodobnostní distribuce μ na \mathcal{X} definovaná vztahem

$$\mu(\sigma) = \frac{1}{Z(\beta)} \exp(-\beta H(\sigma)), \quad (2.20)$$

kde $Z(\beta)$ je normalizační konstanta nazývaná *partiční funkce* a je dána vztahem

$$Z(\beta) := \sum_{\sigma \in \mathcal{X}} \exp(-\beta H(\sigma)). \quad (2.21)$$

Parametr $\beta \geq 0$ určuje vliv energie na Gibbsovu distribuci. Ve fyzikální interpretaci je β závislá na teplotě a říká se mu inverzní teplota. Při $T \rightarrow \infty$ ($\beta = 0$) nehraje žádnou roli a μ je rovnoměrnou distribucí

na \mathcal{X} . Zároveň nedochází k žádným interakcím mezi spiny vrcholů a náhodné veličiny $\{\sigma(v)\}_{v \in V}$ jsou nezávislé. S rostoucí $\beta > 0$ také roste vliv nízkenergetických konfigurací na μ . Aktualizace konfigurace σ probíhá následovně:

- vybereme $w \in V$ rovnoměrně náhodně,
- generujeme novou konfiguraci vzhledem k μ z množiny konfigurací, které jsou totožné se σ ve všech vrcholech odlišných od w . Využijeme glauberovy dynamiky a tedy aktualizujeme pouze spin ve vrcholu w a to s pravděpodobností že zvolíme kladný spin

$$p^+(\sigma, w) = \frac{\mu(\sigma_w^+)}{\mu(\sigma_w^+) + \mu(\sigma_w^-)} = \frac{e^{\beta S(\sigma, w)}}{e^{\beta S(\sigma, w)} + e^{-\beta S(\sigma, w)}} = \frac{1 + \tanh(\beta S(\sigma, w))}{2}, \quad (2.22)$$

kde $S(\sigma, w) := \sum_{u: u \sim w} \sigma(u)$ a $\mu(\sigma_w^+)$ a $\mu(\sigma_w^-)$ značí pravděpodobnostní distribuce, kdy dané konfiguraci σ aktualizujeme daný vrchol w příslušným spinem. Pravděpodobnost, že zvolíme záporný spin je pak samozřejmě dána jako

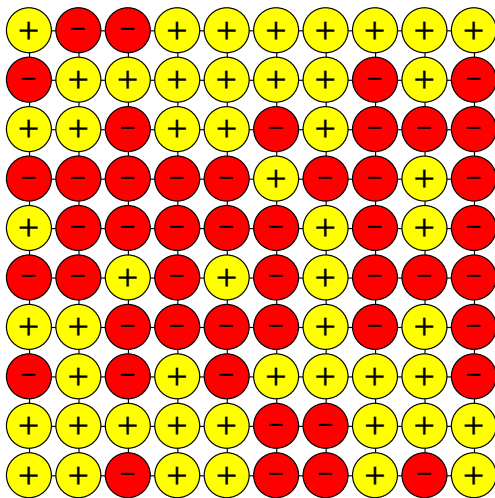
$$p^-(\sigma, w) = \frac{e^{-\beta S(\sigma, w)}}{e^{\beta S(\sigma, w)} + e^{-\beta S(\sigma, w)}} = \frac{1 - \tanh(\beta S(\sigma, w))}{2}. \quad (2.23)$$

Poznamenejme, že rozhodnutí, jaký spin přiřadíme nové konfiguraci závisí pouze na spinech sousedních vrcholů zvoleného vrcholu w . Proto je matice přechodu na \mathcal{X} definována jako

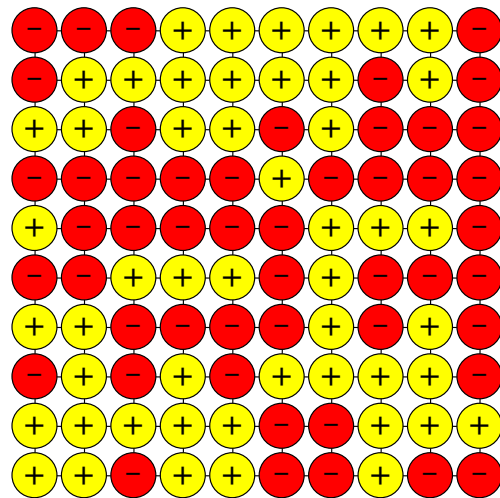
$$P(\sigma, \sigma') = \frac{1}{|V|} \sum_{w \in V} \frac{e^{\beta \sigma'(w) S(\sigma, w)}}{e^{\beta \sigma'(w) S(\sigma, w)} + e^{-\beta \sigma'(w) S(\sigma, w)}} \cdot \mathbf{1}_{\{\sigma(v) = \sigma'(v) \text{ pro } v \neq w\}}. \quad (2.24)$$

Gibbsova distribuce μ je zároveň i stacionární distribucí řetězce [8, sekce 3.3].

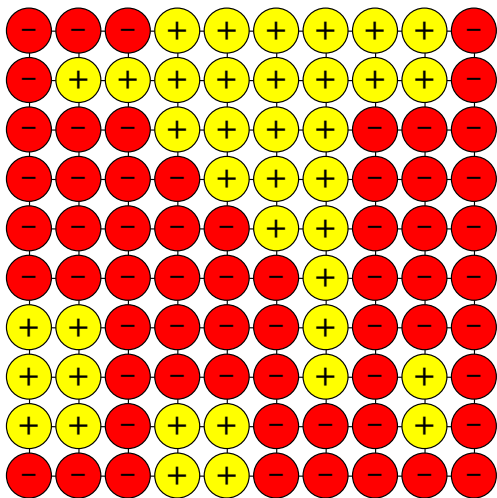
Tento algoritmus si ukážeme opět na simulaci na mřížce 10×10 pro $\beta = 1$. Pokud začneme v situaci, kdy máme na mřížce přibližně stejný počet kladně i záporně nabitých částic, budeme sledovat, jak se tato mřížka vyvíjí v čase. Na obr. 2.6-2.9 vidíme, jak situace vypadá v čase $t = 0$, $t = 10$, $t = 100$ a $t = 1000$. Zdrojový kód je pak stejný jako v případě zdrojového kódu 2.3 (při zvolení parametrů: velikost site = 10; pocet iteraci = 1000).



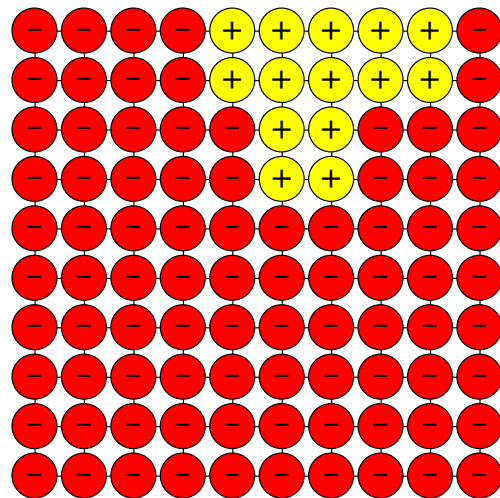
Obrázek 2.6: Počáteční stav Isingova modelu



Obrázek 2.7: Isingův model po 10-ti krocích



Obrázek 2.8: Isingův model po 100 krocích



Obrázek 2.9: Isingův model po 1000 krocích

2.5 Praktické simulace

Ukažme si nyní nějaké praktické využití metody MCMC. Budeme se věnovat dvěma příkladům, ve kterých budeme zkoumat různé charakteristiky zadaných systémů. Nejprve se budeme věnovat správnému q -barvení a poté si ukážeme simulaci přes Isingův model.

2.5.1 Správné q -barvení

Uvažujme opět správné q -barvení na dvoudimenzionální mřížce 10×10 , nyní nám to bohatě postačí. Položme si následující otázky, které nás budou zajímat:

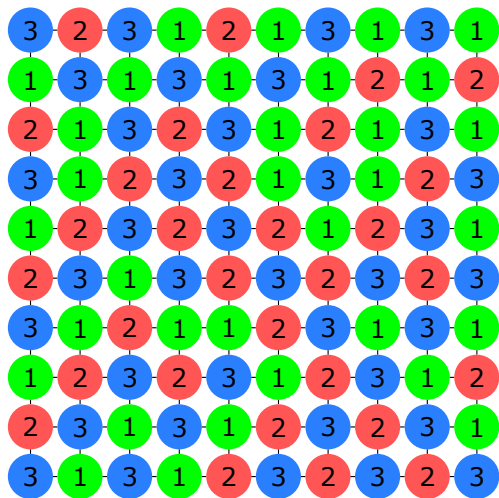
1. Pro jaké nejmenší q můžeme správné q -barvení sestavit za pomoci Metropolisova algoritmu nebo Glauberovy dynamiky?
2. Jak často nastane pro správné 5-barvení situace, že všechny rohové vrcholy obarvíme stejnou barvou?

Řešení:

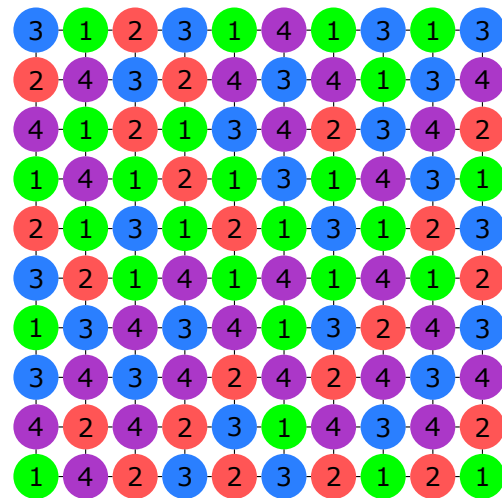
Zkusme nechat odpovědět simulace. Předpokládejme, že vyjdeme z konfigurace, kdy jsou všechny vrcholy obarveny stejnou barvou. Je možné z tohoto stavu dosáhnout konfigurace na správném q -barvení? Jak jsme řekli na začátku kapitoly, pro $q = 2$ existují jen 2 konfigurace správného barvení a mezi nimi nelze přejít volbou pouze jednoho vrcholu. Změní se ale tento výsledek nějak, když budeme uvažovat $q = 3$ nebo $q = 4$? Volme např. $t = 100000$ kroků a počáteční stav, kdy jsou všechny vrcholy obarveny zelenou barvou.

Jak je vidět na obr. 2.10, když obarvujeme jen třemi barvami, nejedná se o správné barvení. I když jsme volili velké množství kroků, nepodařilo se nám podmínku správného q -barvení splnit. Pokud se podíváme na oblasti, které tuto podmínku splňují, vidíme však, že málokterý vrchol má více než jednu možnost (barvu, jíž je aktuálně obarven) na přebarvení.

Naproti tomu z obr. 2.11 je patrné, že jsme nejen splnili podmínku správného barvení, ale navíc mají vrcholy i více možností na přebarvení a tedy je umožněn přechod mezi stavy. Nejmenší počet barev pro použití metody MCMC je tedy $q = 4$.



Obrázek 2.10: 3-barvení



Obrázek 2.11: správné 4-barvení

Zdrojový kód 2.1: Zdrojový kód q -barvení 1

```

1 function [B] = q_barveni(pocet_iteraci, pocet_barev)
2 A=zeros(12);
3 q=1:pocet_barev;
4 p=zeros(1,pocet_barev);
5 A(2:11,2:11) = 1;
6 for i = 1 : pocet_iteraci
7     a=unifrnd(1, 11, 1, 2);
8     x=ceil(a(1));
9     y=ceil(a(2));
10    pocet_moznosti=0;
11    k=1;
12    for j = 1 : pocet_barev
13        if ((q(j)==A(x+1,y)) || (q(j)==A(x-1, y)) || (q(j) == A(x, y
14            -1)) || (q(j) == A(x,y+1))) && q(j) ~= A(x,y)
15            p(j)=0;
16        else
17            p(j)=1;
18        end
19        if p(j)==1
20            mozne_barvy(k)=q(j);
21            k=k+1;
22        end
23        pocet_moznosti=pocet_moznosti + p(j);
24    end
25    b=ceil(unifrnd(0,pocet_moznosti));
26    A(x,y)=mozne_barvy(b);
27 end
28 B = A(2:11,2:11);

```

Budeme-li se zabývat druhou otázkou, spustíme simulaci několikrát a podělíme počet pokusů, které splnili podmínku celkovým počtem provedených simulací. Vychází nám, že tento jev nastane přibližně v 1% případů.

Zdrojový kód 2.2: Zdrojový kód q -barvení 2

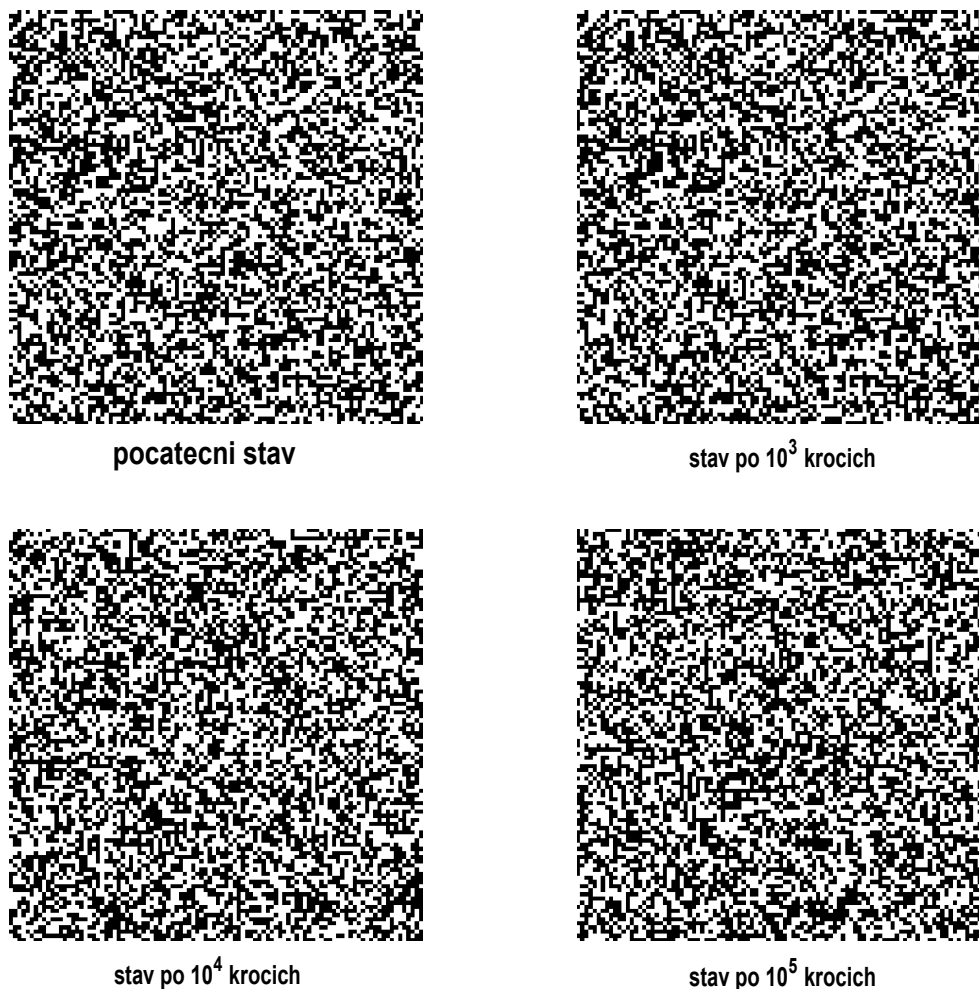
```

1 function [pravdepodobnost] = q_barveni_pravdepodobnost(
    pocet_opakovani, pocet_kroku_v_opakovani, pocet_barev)
2 pocet_uspesnych_barveni = 0;
3 for i = 1 : pocet_opakovani
4     A=zeros(12);
5     q=1:pocet_barev;
6     p=zeros(1,pocet_barev);
7     A(2:11,2:11) = 1;
8     for krok = 1 : pocet_kroku_v_opakovani
9         a=unifrnd(1, 11, 1, 2);
10        x=ceil(a(1));
11        y=ceil(a(2));
12        pocet_moznosti=0;
13        k=1;
14        for j = 1 : pocet_barev
15            if ((q(j)==A(x+1,y)) || (q(j)==A(x-1, y)) || (q(j) == A(x
16                , y-1)) || (q(j) == A(x,y+1))) && q(j) ~= A(x,y)
17                p(j)=0;
18            else
19                p(j)=1;
20                mozne_barvy(k)=q(j);
21                k=k+1;
22            end
23            pocet_moznosti=pocet_moznosti + p(j);
24        end
25        b=ceil(unifrnd(0,pocet_moznosti));
26        A(x,y)=mozne_barvy(b);
27
28
29    end
30    B = A(2:11,2:11);
31    if B(1, 1) == B(1, 10) && B(1, 10) == B(10, 1) && B(10, 1) == B
32        (10, 10)
33        pocet_uspesnych_barveni = pocet_uspesnych_barveni + 1;
34    end
35    pravdepodobnost = pocet_uspesnych_barveni / pocet_opakovani;

```

2.5.2 Isingův model

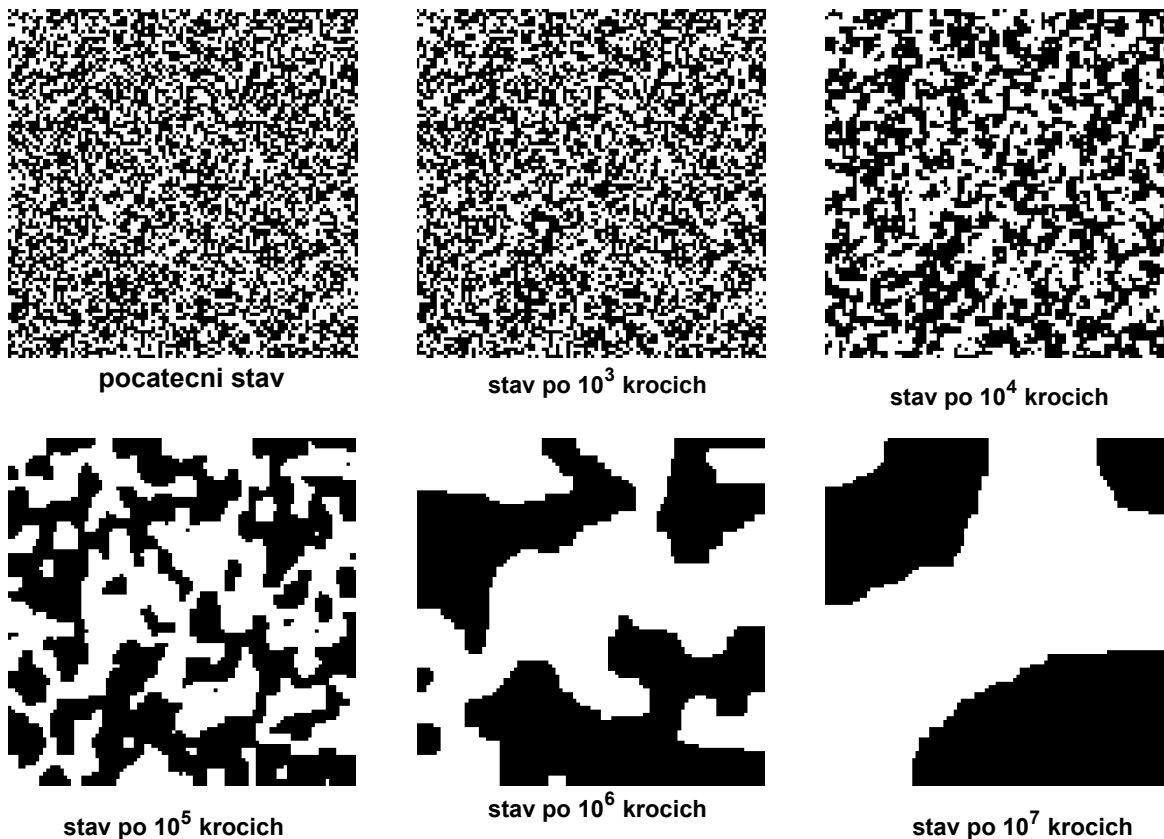
Zkoumejme nyní chování částic v Isingově modelu v závislosti na β prostřednictvím simulace. Mějme mřížku 100×100 a počáteční stav volený náhodně tak, aby v mřížce byl přibližně stejný počet kladných i záporných spinů. Tzn., že projedeme každý vrchol mřížky a zvolíme u něho kladný spin s pravděpodobností $p = 1/2$ a se stejnou pravděpodobností mu přiřadíme záporný spin.



Obrázek 2.12: Simulace Isingova modelu pro $\beta = 0$

Pro $\beta = 0$ se systém chová naprosto náhodně (obr. 2.12). Je vidět, že pro velmi vysoké teploty se struktura systému příliš nemění, což je dané tím, že je spin vybrané částici aktualizován libovolně pro oba spiny se stejnou pravděpodobností $1/2$.

Pro $\beta = 10$ nastává následující situace. Počáteční stav jsme zvolili naprosto shodným postupem jako pro $\beta = 0$. Přibližně po 10 000 krocích začínají být zřetelné oblasti se stejným spinem.



Obrázek 2.13: Simulace Isingova modelu pro $\beta = 10$

V systému se tedy postupně tvoří oblasti, ve které mají částice totožný spin.

Zdrojový kód 2.3: Zdrojový kód Isingův model

```

1 velikost_site = 100;
2 X = zeros(velikost_site, velikost_site, 3);
3 A = zeros(velikost_site + 2, velikost_site + 2);
4 pocet_iteraci = 10000000;
5 for i = 2 : velikost_site + 1
6     for j = 2 : velikost_site + 1
7         c = unifrnd(0,1);
8         if c < 0.5
9             X(i-1,j-1,:) = 0;
10            A(i,j) = -1;
11        else
12            X(i-1,j-1,:) = 1;
13            A(i,j) = 1;
14        end
15    end
16 end

```



```

17 figure;
18 subplot(2,3,1)
19 imshow(X)
20 title('pocatecni stav');
21 beta = 0;
22 for i = 1 : pocet_iteraci
23     a = unifrnd(1, velikost_site + 1, 1, 2);
24     x=ceil(a(1));
25     y=ceil(a(2));
26     suma = A(x-1,y) + A(x+1,y) + A(x,y-1) + A(x,y+1);
27     b = unifrnd(0, exp(beta*suma) + exp(-beta*suma));
28     if b < exp(beta*suma)
29         A(x,y) = 1;
30         X(x-1,y-1,:) = 1;
31     else
32         A(x,y) = -1;
33         X(x-1,y-1,:) = 0;
34     end
35
36     if i == 1000
37         subplot(2,3,2)
38         imshow(X)
39         title('stav po 10^3 krocich')
40     end
41     if i == 10000
42         subplot(2,3,3)
43         imshow(X)
44         title('stav po 10^4 krocich')
45     end
46     if i == 100000
47         subplot(2,3,4)
48         imshow(X)
49         title('stav po 10^5 krocich')
50     end
51     if i == 1000000
52         subplot(2,3,5)
53         imshow(X)
54         title('stav po 10^6 krocich')
55     end
56     if i == 10000000
57         subplot(2,3,6)
58         imshow(X)
59         title('stav po 10^7 krocich')
60     end
61
62 end

```

Kapitola 3

Čas mixingu

V předchozí kapitole jsme si ukázali, jakým způsobem můžeme sestavit vzorky ze zadané pravděpodobnostní distribuce. Při sestavení vzorku ale potřebujeme vědět, jaký čas stanovit, aby byl náš vzorek kvalitní z požadované distribuce. Např. uvažujme, že potřebujeme zamíchat karty, které jsou v balíčku seřazeny podle velikosti. Jak dlouho je třeba nějakým vhodně zvoleným způsobem tento balíček míchat, aby byly karty seřazeny opravdu náhodně?

Cílem této kapitoly bude zadefinovat čas mixingu a poté ve čtvrté kapitole se ho budeme snažit odhadnout právě pro případ míchání karet. Obecně půjde o to, že máme Markovský řetězec se stavovým prostorem \mathcal{X} a maticí přechodu P . Známe také jeho stacionární distribuci π a počáteční distribuci μ_0 . Otázkou je, kolik kroků musíme v řetězci udělat, aby naše pravděpodobnostní distribuce μ_t byla dostatečně blízká té stacionární. První takové $t > 0$, ve kterém jsme se přiblížili stacionární distribuci na požadovanou vzdálenost nazveme *časem mixingu*. Ještě předtím, než budeme schopni zadefinovat čas mixingu sám o sobě, si musíme zvolit míru, která by nám umožnila měřit vzdálenost mezi dvěma pravděpodobnostními distribucemi. Tuto vzdálenost nazveme celkovou variační vzdáleností.

3.1 Celková variační vzdálenost (TVD)

Definice 3.1.1 (Celková variační vzdálenost) [8, sekce 4.1]

Mějme dvě pravděpodobnostní distribuce μ a ν a stavový prostor \mathcal{X} . Potom definujeme *celkovou variační vzdálenost* jako

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)|. \quad (3.1)$$

Ukážeme si nyní použití této obecné definice na jednoduchém příkladu ze života.

Příklad 9 (Výběr automobilu)

Uvažme, že máme k dispozici dva automobily a každý den se jedním z nich potřebujeme dostat do práce. V každém automobilu máme k dispozici minci, s níž na konci každého dne házíme o to, jakým autem pojedeme následující den. Pokud padne panna, do práce následující den jedeme opět stejným autem a padne-li orel, následující den auta prohodíme. Mince auta A má pravděpodobnost, že padne orel p a auta B pravděpodobnost q . Matice přechodu má v tomto případě tedy tvar

$$P = \begin{pmatrix} P(A, A) & P(A, B) \\ P(B, A) & P(B, B) \end{pmatrix} = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}. \quad (3.2)$$

Stacionární distribuci získáme ze vztahu $\pi = \pi P$ (viz (1.3.1)), čímž dostáváme

$$\pi = \left(\frac{q}{p + q}, \frac{p}{p + q} \right). \quad (3.3)$$

Zvolme první den auto A, tedy $\mu_0 = (1, 0)$, a definujeme rozdíl pravděpodobnostních distribucí Δ_t jako

$$\Delta_t = \mu_t(A) - \pi(A) \quad (3.4)$$

a odtud si vyjádříme celkovou variační vzdálenost

$$\|\mu_t - \pi\|_{TV} = |\Delta_t| = |P^t(A, A) - \pi(A)| = |(1 - \pi(A)) + (P^t(A, A) - 1)| = |\pi(B) - P^t(A, B)|.$$

Rekurentní vztah pro Δ_t si vyjádříme jednoduše, protože platí následující 2 rovnice:

$$\mu_{t+1}(A) = (1 - p)\mu_t(A) + q\mu_t(B) = (1 - p)\mu_t(A) + q(1 - \mu_t(A)),$$

$$\Delta_{t+1} = \mu_{t+1}(A) - \frac{q}{p+q},$$

A z toho dostáváme dosazením první rovnice do druhé vztah

$$\Delta_{t+1} = (1 - p)\mu_t(A) + q(1 - \mu_t(A)) - \frac{q}{p+q} = (1 - p - q)\left[\mu_t(A) - \frac{q}{p+q}\right] = (1 - p - q)\Delta_t$$

a konečně rozvinutím Δ_{t+1} dostáváme výsledný tvar

$$\Delta_t = (1 - p - q)^t \Delta_0. \quad (3.5)$$

Pro hodnoty p a q tohoto dvou-stavového řetězce takové, že $|1 - p - q| \in (0, 1)$ plyne, že celková variační vzdálenost klesá k nule exponenciálně.

Použití výpočtu celkové variační vzdálenosti přímo z definice není příliš komfortní cestou, poněvadž nalezení maxima přes *všechny* podmnožiny \mathcal{X} mnohdy není krátkou ani snadnou úlohou. Proto si teď ukážeme velice užitečné alternativní možnosti, jak tuto celkovou variační vzdálenost nalézt.

Věta 3.1.2 (1. alternativa TVD) [8, sekce 4.1]

Nechť μ a ν jsou dvě pravděpodobnostní distribuce na \mathcal{X} . Potom

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|. \quad (3.6)$$

Důkaz: Důkaz provedeme stejným způsobem jako je tomu v [8, sekce 4.1]. Zadefinujeme si množinu $A = \{x : \mu(x) \geq \nu(x)\}$ a libovolnou množinu $B \subset \mathcal{X}$. Platí

$$\mu(B) - \nu(B) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(A) - \nu(A). \quad (3.7)$$

Z toho plyne, že největší taková množina $A \subseteq \mathcal{X}$ z (3.1.1) je buď $A_1 = \{x : \mu(x) \geq \nu(x)\}$ nebo naopak $A_2 = \{x : \mu(x) < \nu(x)\}$ (z inverzní analogie). Zároveň také platí, že $A_2 = A_1^c$. A máme tedy

$$\mu(A_1) - \nu(A_1) = (1 - \nu(A_1)) + (\mu(A_1) - 1) = \nu(A_1^c) - \mu(A_1^c), \quad (3.8)$$

což znamená, celková variační vzdálenost se pro obě tyto množiny rovná. Z toho důvodu máme celkem

$$\sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| = \sum_{i=1}^2 |\mu(A_i) - \nu(A_i)| = 2\|\mu - \nu\|_{TV}, \quad (3.9)$$

čímž je důkaz dovršen. □

3.2 Párování

Další možností, jak vypočítat celkovou variační vzdálenost je použití tzv. **párování**. Napřed si tedy párování zdefinujeme a posléze vyslovíme 2. alternativu pro TVD.

Definice 3.2.1 (Párování) [8, sekce 4.2]

Mějme pravděpodobnostní distribuce μ a ν a stavový prostor \mathcal{X} . Potom **párováním** μ a ν nazveme dvojici náhodných veličin (X, Y) definované na \mathcal{X} tak, že μ je marginální distribuce náh. veličiny X a ν je marginální distribucí náh. veličiny Y . Párování tedy splňuje $\mathbf{P}\{X = x\} = \mu(x)$ a $\mathbf{P}\{Y = y\} = \nu(y)$.

Párování je mocný nástroj a má mnoho různých využití. Úzce také souvisí s celkovou variační vzdáleností, na což se teď podíváme [8, sekce 4.2].

Věta 3.2.2 (2. alternativa TVD) [8, sekce 4.2]

Nechť μ a ν jsou dvě pravděpodobnostní distribuce na stavovém prostoru \mathcal{X} . Potom

$$\|\mu - \nu\|_{TV} = \inf\{\mathbf{P}\{X \neq Y\} : (X, Y) \text{ je párování } \mu \text{ a } \nu\}. \quad (3.10)$$

Důkaz: Ukážeme si pouze první část důkazu a to, že platí $\|\mu - \nu\|_{TV} \leq \mathbf{P}\{X \neq Y\}$.

Pro libovolné párování (X, Y) μ a ν a libovolnou událost $A \subset \mathcal{X}$ platí

$$\mu(A) - \nu(A) = \mathbf{P}\{X \in A\} - \mathbf{P}\{Y \in A\} \quad (3.11)$$

$$\leq \mathbf{P}\{X \in A, Y \notin A\} \quad (3.12)$$

$$\leq \mathbf{P}\{X \in A, Y \notin A\} + \mathbf{P}\{X \notin A, Y \in A\} = \mathbf{P}\{X \neq Y\}. \quad (3.13)$$

První nerovnost je splněna, neboť pro $(a, b) \in \langle 0, 1 \rangle$ platí

$$a(1 - b) \geq a - b.$$

Jelikož (3.11) platí pro libovolnou množinu A , pak platí i pro $A^* = \{x : \mu(x) \geq \nu(x)\}$, která vystupuje v důkazu věty (3.1.2) a přímo pro ni platí

$$\mu(A^*) - \nu(A^*) = \|\mu - \nu\|_{TV}. \quad (3.14)$$

Ve druhé fázi je třeba dokázat rovnost

$$\mathbf{P}\{X \neq Y\} = \|\mu - \nu\|_{TV}, \quad (3.15)$$

kterou však dokazovat nebudeme. Dá se nalézt v [8, sekce 4.2].

□

Jak vidíme, nejmenší variační vzdálenost se dá pomocí párování převést na porovnání dvou náhodných veličin namísto dvou distribucí. Přesněji celková variační vzdálenost je definována jako infimum přes všechna párování. To nám nabízí jednoduchý vrchní odhad celkové variační vzdálenosti, protože pro každé párování (X, Y) platí:

$$\|\mu - \nu\|_{TV} \leq \mathbf{P}\{X \neq Y\} : (X, Y) \text{ je párování } \mu \text{ a } \nu. \quad (3.16)$$

Doteď jsme si v naší cestě za definováním času mixingů stačili zdefinovat celkovou variační vzdálenost včetně dvou modifikací obecné definice. V tuto chvíli by bylo příhodné alespoň zmínit konvergenční větu, která říká, že za určitých předpokladů Markovský řetězec konverguje ke své stacionární distribuci.

Věta 3.2.3 (Konvergenční věta) [8, sekce 4.3]

Necht' P je neredukovatelný a aperiodický řetězec se stacionární distribucí π . Potom existují konstanty $\alpha \in (0, 1)$ a $C > 0$ takové, že

$$\max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{TV} \leq C\alpha^t. \quad (3.17)$$

Už tedy víme, že za předpokladu neredukovatelnosti a aperiodicity Markovský řetězec konverguje ke své stacionární distribuci. Dále nás bude zajímat právě odhadnutí této maximální vzdálenosti mezi $P^t(x_0, \cdot)$ a π pro konečné $t \geq 0$. Pro další lepší orientaci si tedy označíme tuto vzdálenost

$$d(t) := \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{TV}. \quad (3.18)$$

Také pro nás bude důležitý následující výraz [8, sekce 4.4]

$$\bar{d}(t) := \max_{x, y \in \mathcal{X}} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}. \quad (3.19)$$

Vztah mezi $d(t)$ a $\bar{d}(t)$ je dán následujícím lemmatem.

Lemma 3.2.4 [8, sekce 4.4]

Jestliže jsou dány $d(t)$ a $\bar{d}(t)$ takového tvaru, jak jsme si je definovali výše, pak vztah mezi nimi je

$$d(t) \leq \bar{d}(t) \leq 2d(t). \quad (3.20)$$

Důkaz: Druhá nerovnost plyne z trojúhelníkové nerovnosti, první si dokážeme.

$$\begin{aligned} |P^t(x, A) - \pi(A)| &= \left| \sum_{y \in \mathcal{X}} \pi(y) [P^t(x, A) - P^t(y, A)] \right| \leq \\ &\leq \sum_{y \in \mathcal{X}} \pi(y) \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \bar{d}(t), \end{aligned} \quad (3.21)$$

kde rovnost plyne ze stacionarity.

□

Užitečné pro nás bude také následující lemma.

Lemma 3.2.5 Pro funkci \bar{d} platí: $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$, kde s, t jsou časy, tedy $s, t \in \mathbb{N}$. Z toho také přímo plyne $\bar{d}(ct) \leq \bar{d}(t)^c$, kde $c, t \in \mathbb{N}$.

3.3 Čas mixingů

Nyní máme k dispozici veškerý aparát na zavedení času mixingů. Jak jsme uvedli na počátku kapitoly, je to nejmenší takový počet kroků $t > 0$, který Markovský řetězec potřebuje na to, aby se přiblížil ke své stacionární distribuci na požadovanou vzdálenost [8, sekce 4.5]

$$t_{\text{mix}}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}. \quad (3.22)$$

Za požadovanou vzdálenost se obvykle volí $\varepsilon = 1/4$, tedy

$$t_{\text{mix}} := t_{\text{mix}}(1/4). \quad (3.23)$$

Ve čtvrté kapitole se tomuto času mixingů budeme snažit najít horní i dolní odhady.

Kapitola 4

Horní a dolní odhady času mixingu

Při odhadu času mixingu nelze postupovat komplexně pro všechny typy úloh dohromady, proto se zde omezíme na příklad míchání karet, pro nějž čas mixingu odhadneme. Budeme uvažovat tento postup míchání: Mějme balíček n karet, z nějž bereme vždy vrchní kartu a vmícháváme ji náhodně do balíčku.

4.1 Horní odhad

Ještě než se pustíme přímo do horního odhadu tohoto příkladu, spočítejme zde jeden příklad, který nám bude pro tento úkol užitečný.

Příklad 10 [Sběratel kuponů]

Společnost vydává n odlišných typů kuponů, přičemž sběratel chce získat kompletní sbírku. Kolik kuponů musí v průměru získat, aby jeho sbírka obsahovala všechny typy? Pravděpodobnost získání jednotlivých kuponů je pro všechny kupony stejná.

Řešení:

Nechť t označuje počet kuponů celkem a X_t označuje počet získaných kuponů. Jednoduše sestavíme podmíněné pravděpodobnosti:

- $\mathbf{P}\{X_{t+1} = k + 1 | X_t = k\} = \frac{n-k}{n}$,
- $\mathbf{P}\{X_{t+1} = k | X_t = k\} = \frac{k}{n}$.

Nechť τ je označení doby, kdy má sběratel kompletní sbírku. τ je náh. veličina, tedy nás bude zajímat její střední hodnota.

τ_k si označíme počet všech kuponů ve chvíli, kdy máme ve své sbírce k různých. Pak

$$\tau = \tau_n = \tau_1 + (\tau_2 - \tau_1) + \dots + (\tau_n - \tau_{n-1}) \quad (4.1)$$

Také $\tau_k - \tau_{k-1}$ je náhodná veličina s pravděpodobností úspěchu $\frac{n-k+1}{n}$ a tedy $\mathbf{E}(\tau_k - \tau_{k-1}) = \frac{n}{n-k+1}$. Vyjádříme-li si tedy střední hodnotu pro τ , dostaneme

$$\mathbf{E}(\tau) = \sum_{k=1}^n \mathbf{E}(\tau_k - \tau_{k-1}) = n \sum_{k=1}^n \frac{1}{n-k+1} = n \sum_{k=1}^n \frac{1}{k}. \quad (4.2)$$

Připomeňme také, že protože platí $|\sum_{k=1}^n \frac{1}{k} - \log(n)| \leq 1$, máme, že $|\mathbf{E}(\tau) - n \log(n)| \leq n$.

Dále dokažme následující větu, která pro nás bude dále důležitá:

Věta 4.1.1 *Necht' τ je náhodná veličina označující čas nasbírání všech kuponů. Pak pro libovolné $c > 0$,*

$$\mathbf{P}\{\tau > \lceil n \log(n) + cn \rceil\} \leq e^{-c}. \quad (4.3)$$

Důkaz:

Označme A_i událost, že i -tý typ kuponu nemáme mezi prvními $\lceil n \log(n) + cn \rceil$ kupony.

Potom

$$\mathbf{P}\{\tau > \lceil n \log n + cn \rceil\} = \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbf{P}(A_i). \quad (4.4)$$

Každý kupon má pak v každém pokusu pravděpodobnost, že nebude zařazen $1 - \frac{1}{n}$ a nakonec tedy máme

$$\sum_{i=1}^n \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} \leq n \exp\left(-\frac{n \log n + cn}{n}\right) = e^{-c}. \quad (4.5)$$

□

Příklad 11 [Míchání karet]

Mějme balíček n karet a volíme následující postup míchání. Z balíčku vždy vezmeme vrchní kartu a vložíme ji mezi ostatní karty rovnoměrně náhodně. Jak dlouho musíme míchat balíček tímto způsobem, aby bylo seřazení karet blízké náhodnému? Tento proces je náhodnou procházkou na grupě \mathcal{S}_n s $n!$ možnými permutacemi. Necht' τ_{top} je čas, ve kterém je původní spodní karta poprvé umístěna do balíčku. To znamená, že v čase $\tau_{top} - 1$ se původní spodní karta poprvé dostává na vrch balíčku.

Na začátek si ukážeme, že v čase τ_{top} budou karty zamíchány náhodně.

Věta 4.1.2 *Uvažme stejný způsob míchání karet jako v příkladu výše. Náhodná veličina $X_{\tau_{top}}$ má rovnoměrnou distribuci a čas τ_{top} je nezávislý na $X_{\tau_{top}}$.*

Důkaz: Chceme ukázat, že v čase τ_{top} jsme měli šanci dosáhnout všech $n!$ permutací. V čase $t = 0$ máme pod původní spodní kartou 0 karet. Označme t_k situaci, kdy vložíme k -tou kartu pod původní spodní kartu. Uvažujme, že v čase $t_1 \neq 0$ poprvé dáme kartu pod původně spodní kartu a ta tak musí být umístěna jedinec bezprostředně pod spodní kartou ($1!$ permutací). V čase t_k umístíme k -tou kartu pod původně spodní, a ta má k možností, kam může být vložena a tedy pod původně spodní kartou se už nabízí $k!$ permutací. Navíc jsme vybírali pozici vždy náhodně, tedy každá z těchto permutací má stejnou pravděpodobnost. Pro zamíchání původní spodní karty zpět do balíčku je n možností a tedy už můžeme zajistit $n!$ permutací s rovnoměrnou distribucí.

□

4.1.1 Silně stacionární časy

Abychom byli schopni teoreticky omezit čas mixingu a nějakým vhodným způsobem zadefinovat obecně čas, jehož příkladem je i τ_{top} , musíme nejdříve získat potřebný aparát.

Definice 4.1.3 (Dělení Π) [8, apendix A.2]

Počitatelné dělení Π populace Ω je posloupnost neslučitelných událostí $\{A_i\}$ taková, že $\bigcup_i A_i = \Omega$.

Předpokládejme, že takové dělení vždy splňuje $\mathbf{P}(A_i) > 0$ pro všechna i . Mějme Y jako diskrétní náhodnou veličinu s hodnotami $\{y_i\}$. Potom například události $A_i = \{Y = y_i\}$ tvoří takové dělení.

Definice 4.1.4 (σ -algebra) [8, appendix A.2]

Mějme dělení $\Pi = (A_1, \dots, A_n)$. Potom definujeme σ -algebru $\mathcal{G} = \mathcal{G}(\Pi)$ jako množinu všech spočetných sjednocení množin z Π . Tedy

$$\mathcal{G} = \left\{ \bigcup_j A_{ij} : A_{ij} \in \Pi \right\}. \quad (4.6)$$

Pokud pozorovatel ví, které prvky z dělení Π nastaly a nemá žádné další informace, pak množiny v \mathcal{G} jsou jen ty množiny, pro které víme, zda nastaly nebo nenastaly. Množiny A_{ij} z definice 4.1.4 jsou tedy události, o kterých víme, že nastaly (ty jsme označili A_i) a z nich tvoříme spočetná sjednocení. Hlubší teoretický vzhled lze nalézt v [8, appendix A].

Definice 4.1.5 (Filtrace) [8, sekce 6.2]

Filtrací $\{\mathcal{F}_t\}$ rozumíme posloupnost σ -algeber takovou, že $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ pro všechna t . Pokud je náhodná veličina X_t \mathcal{F}_t -měřitelná pro všechna t , potom říkáme, že je markovský řetězec $\{X_t\}$ **přízpůsoben** filtraci $\{\mathcal{F}_t\}$. Pokud máme na mysli filtraci $\mathcal{H}_t = \sigma(X_0, X_1, \dots, X_t)$, potom \mathcal{H}_t nazýváme **přirozená filtrace**.

Přirozená filtrace \mathcal{H}_t tedy vytváří σ -algebru nad množinou stavů X_0, \dots, X_t . Obecně ale nemusí být filtrace sestrojena nad touto množinou, ale nad nějakou jinou pomocnou množinou.

Pod pojmem filtrace si můžeme představit množinu, díky níž můžeme sledovat systém a také určit, kdy je systém z požadované stacionární distribuce. Někdy nám postačí zadefinovat takovou množinu přímo nad trajektoriemi (X_0, \dots, X_t) , ale občas se nevyhneme stanovení jiné směřodatné náhodné veličiny $Y_t \dots$

Definice 4.1.6 (Markovský řetězec s ohledem na $\{\mathcal{F}_t\}$) [8, sekce 6.2]

Necht' $\{X_t\}$ je přízpůsoben $\{\mathcal{F}_t\}$. Potom řekneme, že $\{X_t\}$ je **markovský řetězec s ohledem na $\{\mathcal{F}_t\}$** , pokud platí

$$\mathbf{P}_x\{X_{t+1} = y | \mathcal{F}_t\} = P(X_t, y), \quad (4.7)$$

kde P je matice přechodu řetězce $\{X_t\}$. Symbolem \mathbf{P}_x rozumíme pravděpodobnost při počátečním stavu $X_0 = x$ a $P(X_t, y)$ značí $P(z, y)$, kde $X_t = z$.

Pokud $\{\mathcal{F}_t\}$ je přirozená filtrace, pak markovský řetězec definovaný v definici 1.1.1 splňuje zároveň (4.7).

Definice 4.1.7 (Čas zastavení) [8, sekce 6.2]

Čas zastavení pro filtraci $\{\mathcal{F}_t\}$ je náhodná veličina τ s hodnotami $\{0, 1, \dots\}$ taková, že $\{\tau = t\} \in \mathcal{F}_t$.

Uvažme opět míchání karet. Chtěli bychom přestat míchat karty v čase τ_{top} , tedy, aby byl τ_{top} časem zastavení. Nemůžeme ale zadefinovat tento čas zastavení přímo, protože je to čas, ve kterém jsou karty zamíchány úplně náhodně. Nicméně víme o tomto čase, že nastává tah po situaci, kdy je původně spodní karta navrchu balíčku. Zadefinujme tedy čas prvního dotyku tak, aby odpovídal času, kdy je původně spodní karta navrchu balíčku.

Definice 4.1.8 (Čas prvního dotyku) [8, sekce 6.2]

Necht' $A \subseteq \mathcal{X}$ a navíc mějme vektor stavů (X_0, X_1, \dots, X_t) , přičemž množinu A jsme navštívili poprvé v čase t . Tedy

$$\tau_A = \min\{t \geq 0 : X_t \in A\}. \quad (4.8)$$

Pro míchání karet tedy uvažujme událostí A z definice času prvního dotyku množinu stavů, kdy je původní spodní karta navrchu balíčku. Čas prvního dotyku τ_A je tedy čas, kdy se původní spodní karta dostane navrch balíčku a čas zastavení je čas, kdy původní spodní kartu vmícháme do balíčku, tedy $\tau_{top} = \tau_A + 1$

Definice 4.1.9 (Stacionární čas) [8, sekce 6.3]

Necht' (X_t) je neredukovatelný markovský řetězec se stacionární distribucí π . Necht' $\{\mathcal{F}_t\}$ je filtrace a $\{X_t\}$ je přizpůsobený $\{\mathcal{F}_t\}$. **Stacionární čas** τ pro (X_t) je čas zastavení pro filtraci $\{\mathcal{F}_t\}$ takový, že distribuce X_τ je π . Tzn., že platí

$$\mathbf{P}_x\{X_\tau = y\} = \pi(y) \quad \text{pro všechna } y. \quad (4.9)$$

τ_{top} je tedy stacionární čas pro míchání karet. Může se stát, že už počáteční stav je zvolen z rovnoměrné distribuce, tedy $\tau = 0$. To ale znamená, že τ a X_τ nejsou nezávislé. Pro míchání karet chceme odhadnout čas mixingů t_{mix} právě časem τ_{top} a k tomu potřebujeme silnější vlastnost než je (4.9). Potřebujeme zajistit, aby τ a X_τ byly nezávislé a přesně proto definujeme silně stacionární čas.

Definice 4.1.10 (Silně stacionární čas) [8, sekce 6.4]

Necht' (X_t) je markovský řetězec přizpůsobený filtraci $\{\mathcal{F}_t\}$ se stacionární distribucí π . **Silně stacionárním časem** pro řetězec (X_t) a počáteční stav x rozumíme $\{\mathcal{F}_t\}$ -čas zastavení τ takový, že X_τ má distribuci π a je nezávislý na τ . Tedy

$$\mathbf{P}_x\{\tau = t, X_\tau = y\} = \mathbf{P}_x\{\tau = t\}\pi(y). \quad (4.10)$$

Věta 4.1.11 [8, sekce 6.4] *Pokud je τ silně stacionární čas pro počáteční stav x , pak*

$$\|P^t(x, \cdot) - \pi\|_{TV} \leq \mathbf{P}_x\{\tau > t\}. \quad (4.11)$$

Důkaz:

Důkaz provedeme podle stejného schématu jako je tomu v [1, sekce 2].

Pro libovolnou množinu $A \in \mathcal{X}$ platí

$$\begin{aligned} P^t(x, A) &= \mathbf{P}\{X_t \in A\} = \\ &= \sum_{j \leq t} \mathbf{P}\{X_t \in A, \tau = j\} + \mathbf{P}\{X_t \in A, \tau > t\} = \\ &= \sum_{j \leq t} \pi(A) \mathbf{P}\{\tau = j\} + \mathbf{P}\{X_t \in A | \tau > t\} \mathbf{P}\{\tau > t\} \\ &= \pi(A)(1 - \mathbf{P}\{\tau > t\}) + \mathbf{P}\{X_t \in A | \tau > t\} \mathbf{P}\{\tau > t\} = \\ &= \pi(A) + \mathbf{P}\{\tau > t\}(\mathbf{P}\{X_t \in A | \tau > t\} - \pi(A)) \end{aligned} \quad (4.12)$$

a z toho plyne

$$|P^t(x, \cdot) - \pi| \leq \mathbf{P}_x\{\tau > t\}. \quad (4.13)$$

□

4.1.2 Horní odhad pro míchání karet

Dejme to tedy vše dohromady. Z věty 4.1.11 máme

$$d(t) \leq \mathbf{P}_x\{\tau_{top} > t\}. \quad (4.14)$$

Necht' t označuje počet tahů, X_t označuje počet karet pod původní spodní kartou. Pak podmíněné pravděpodobnosti mají tvar:

- $\mathbf{P}\{X_{t+1} = k + 1 | X_t = k\} = \frac{k+1}{n}$,

- $\mathbf{P}\{X_{t+1} = k | X_t = k\} = \frac{n-k-1}{n}$.

Budeme chtít získat střední hodnotu času τ_{top} , kdy vmícháme i původní spodní kartu zpět do balíčku. Ozn. $\tau_{top} - 1$ (čas, kdy se dostane původní spodní karta navrch balíčku) jako τ a τ_k nám bude značit počet tahů ve chvíli, kdy je pod původní spodní kartou k karet.

$$\tau_{top} = 1 + \tau = 1 + \tau_{n-1} = 1 + \tau_1 + (\tau_2 - \tau_1) + \dots + (\tau_{n-1} - \tau_{n-2}) \quad (4.15)$$

Rozdíl $(\tau_k - \tau_{k-1})$ je také náhodná veličina s pravděpodobností úspěchu $\frac{k}{n}$ a tedy $\mathbf{E}(\tau_k - \tau_{k-1}) = \frac{k}{n}$. Vyjádříme-li si střední hodnotu pro τ_{top} , dostáváme

$$\mathbf{E}(\tau_{top}) = 1 + \sum_{k=1}^{n-1} \mathbf{E}(\tau_k - \tau_{k-1}) = \frac{n}{n} + n \sum_{k=1}^{n-1} \frac{1}{k} = n \sum_{k=1}^n \frac{1}{k}, \quad (4.16)$$

což je stejný výsledek jako v případě sběratele kuponů a dále bychom postupovali úplně stejně. Tedy máme dále

$$d(\lceil n \log(n) + cn \rceil) \leq \mathbf{P}_x\{\tau > \lceil n \log(n) + cn \rceil\} \leq e^{-c}. \quad (4.17)$$

Z definice času mixingu požadujeme, aby $d(t)$ bylo menší nebo rovno ε , tedy celkem

$$d(\lceil n \log(n) + cn \rceil) \leq e^{-c} = \varepsilon, \quad (4.18)$$

a z toho

$$c = \log(\varepsilon^{-1}). \quad (4.19)$$

Nakonec odhad času mixingu pro míchání karet je tedy

$$t_{mix}(\varepsilon) \leq \lceil n \log(n) + \log(\varepsilon^{-1})n \rceil. \quad (4.20)$$

4.2 Dolní odhad pro míchání karet

Podobným způsobem, jakým jsme odhadli čas mixingu shora, zkusíme nyní taktéž učinit pro odhad zdola. Postup provedeme obdobně jako je uvedeno v [8, sekce 7.4]. Nejprve si vyjádříme čas, za jaký jsme schopni původně j -tou kartu odspoda vmíchat zpět do balíčku. Teoretická střední hodnota a rozptyl vycházejí

$$\mathbf{E}(\tau_j) = \sum_{i=j}^{n-1} \frac{n}{i} \geq n \int_j^n \frac{dx}{x} = n \log(n) - n \log(j) \quad (4.21)$$

a

$$\mathbf{D}(\tau_j) \leq n^2 \sum_{i=j}^{\infty} \frac{1}{i(i+1)} \leq \frac{n^2}{j-1}. \quad (4.22)$$

Střední hodnotu odhadneme vztahem

$$\widehat{\mathbf{E}(\tau_j)} = n \log(n) - \alpha n. \quad (4.23)$$

Naším cílem bude nyní najít vhodné $\alpha = \alpha(\varepsilon)$ v závislosti na ε tak, aby

$$d_n(n \log(n) - \alpha n) \geq 1 - \varepsilon, \quad (4.24)$$

a tedy

$$t_{mix}(1 - \varepsilon) \geq n \log(n) - \alpha n. \quad (4.25)$$

Využijeme nyní nerovnosti (4.21) a (4.22) společně s Čebyševovou nerovností, čímž dostaneme

$$\mathbf{P}\{\tau_j < n \log(n) - \alpha n\} \leq \mathbf{P}\{\tau_j - \mathbf{E}(\tau_j) < -n(\alpha - \log(j))\} \leq \frac{1}{j-1}, \quad (4.26)$$

pod podmínkou, že $\alpha \geq \log(j) + 1$. Pokud platí, že $\tau_j \geq n \log(n) - \alpha n$, potom je původních j karet v čase $n \log(n) - \alpha n$ v tomtéž relativním pořadí jako na začátku. Množinu, kdy je původních j karet ve stejném relativním pořadí označíme A_j . Vyjádříme-li si pravděpodobnost $P^{n \log(n) - \alpha n}(id, A_j)$, dostáváme

$$P^{n \log(n) - \alpha n}(id, A_j) \geq \mathbf{P}\{\tau_j \geq n \log(n) - \alpha n\} \geq 1 - \frac{1}{j-1}, \quad (4.27)$$

pro $\alpha \geq \log(j) + 1$. Zároveň víme, že pro stacionární distribuci pro stav A_j platí

$$\pi(A_j) = \frac{1}{j!} \leq \frac{1}{j-1}, \quad (4.28)$$

odkud, pro $\alpha \geq \log(j) + 1$

$$d_n(n \log(n) - \alpha n) \geq \|P^{n \log(n) - \alpha n}(id, \cdot) - \pi\|_{TV} \geq P^{n \log(n) - \alpha n}(id, A_j) - \pi(A_j) > 1 - \frac{2}{j-1}. \quad (4.29)$$

Nyní už dosazením $j = \lceil e^{\alpha-1} \rceil$ za předpokladu, že $n \geq j$, tedy $n \geq \lceil e^{\alpha-1} \rceil$, dostáváme konečných

$$d_n(n \log(n) - \alpha n) > 1 - \frac{2}{\lceil e^{\alpha-1} \rceil - 1} = 1 - \varepsilon. \quad (4.30)$$

Nalezli jsme tedy dolní odhad pro míchání karet pro obecné ε .

Závěr

Hlavním záměrem této bakalářské práce bylo stručně a výstižně popsat základní aspekty metody Markov Chain Monte Carlo.

Celá metoda stojí na konečných markovských řetězcích, kterými se zabývá celá první kapitola. Zdefinovali jsme markovský řetězec, zabývali jsme se jejich důležitými vlastnostmi pro metodu Markov Chain Monte Carlo a také vše názorně ukázali na několika příkladech.

Druhá kapitola se pak zaměřuje na metodu Markov Chain Monte Carlo samotnou. Jsou zde podrobně zpracovány Metropolis-Hastingsův algoritmus a Glauberova dynamika a oba algoritmy jsou názorně demonstrovány na praktických simulacích. Zaměřili jsme se v nich zejména na model správné q -barvení a Isingův model. Ačkoli jsme systémy simulovali jen na mřížkách relativně malých rozměrů, přesvědčili jsme se, že simulace může být časově poměrně náročná, navíc pouhým "tipnutím" rozumné doby nemáme žádnou jistotu, jestli je náš konečný vzorek dostatečně kvalitní z požadované distribuce.

Pro tyto účely jsme zdefinovali ve třetí kapitole čas mixingů a na příkladu jednoduchého míchání karet jsme se ve čtvrté kapitole pokusili jeho hodnotu shora i zespodu teoreticky odhadnout. Pro horní odhad jsme využili postup přes silně stacionární časy.

Tato práce může sloužit jako studijní materiál nejen pro studenty matematického inženýrství, ale také pro nadšence, kterým může poskytnout rychlý a přehledný vhled do metody Markov Chain Monte Carlo.

Protože je použití metody Markov Chain Monte Carlo rozsáhlým tématem, nabízí se zde hned několik možností, jak na tuto práci navázat. Můžeme se dále věnovat výpočtu vícerozměrných integrálů, využít více do hloubky párování, či simulovat složitější systémy.

Literatura

- [1] D. Aldous, P. Diaconis: *Shuffling cards and stopping times*. Amer. Math. Monthly 93, no. 5, 1986, 333-348.
- [2] S. Brooks: *Markov chain Monte Carlo method and its application*. Journal of the Royal Statistical Society: series D (the Statistician) 47(1), 1998, 69-100.
- [3] S. Brooks et al.: *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011.
- [4] S. Geman, D. Geman: *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, 1984, 721-741.
- [5] W. R. Gilks, S. Richardson, D. J. Spiegelhalter: *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1995.
- [6] J. Gravner: *Lecture Notes for Introductory Probability* [online]. Mathematics Department, University of California, Davis, CA 95616, 2020 [cit. 2021-06-20]. Dostupné z <<https://www.math.ucdavis.edu/gravner/MAT135A/resources/lecturenotes.pdf>>.
- [7] W. K. Hastings: *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. Biometrika, vol. 57, no. 1, 97-109.
- [8] D. A. Levin, Y. Peres, E. L. Wilmer: *Markov chains and mixing times*. American Mathematical Society, Providence, R.I., 2009.
- [9] M. Mathew: *The Application of Markov Chain Monte Carlo Techniques in Non-Linear Parameter Estimation for Chemical Engineering Models* [online]. University of Waterloo, Ontario, Canada, 2013 [cit. 2021-06-20]. Dostupné z <<https://core.ac.uk/download/pdf/144146881.pdf>>.
- [10] N. Metropolis et al.: *Equation of State Calculations by Fast Computing Machines*. The Journal of Chemical Physics: vol. 21, no. 6, 1953, 1087-1092.
- [11] R. Peled, Y. Spinka: *Three lectures on random proper colorings of \mathbb{Z}^d* [online]. 2020 [cit. 2021-06-20]. Dostupné z <<https://arxiv.org/pdf/2001.11566.pdf>>.
- [12] R. Serfozo: *Basics of Applied Stochastic Processes*. Springer-Verlag Berlin Heidelberg, 2009.