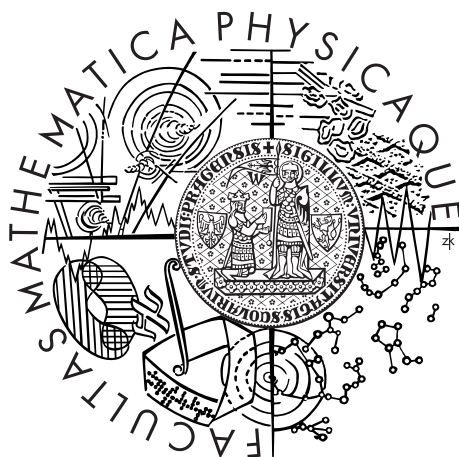


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Martin Kalaš

Coupling a rychlost konvergence diskrétních MCMC algoritmů

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Michaela Prokešová, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční matematika

Praha 2011

Na tomto místě bych rád poděkoval vedoucí bakalářské práce RNDr. Michaele Prokešové, Ph.D. za její trpělivost při konzultacích problému a podnětné připomínky při úpravě textu.

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne

Martin Kalaš

Název práce: Coupling a rychlost konvergence diskretních MCMC algoritmů

Autor: Martin Kalaš

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Michaela Prokešová, Ph.D.

Abstrakt: Konvergence marginálního rozdělení Markovova řetězce ke stacionárnímu rozdělení je důležitá vlastnost, která má v moderní matematice mnoho aplikací. Jednou z nich jsou např. Markov Chain Monte Carlo algoritmy, které slouží ke generování realizací ze složitých pravděpodobnostních rozdělení. Pro takové aplikace je klíčové správně odhadnout tzv. mixing time Markovova řetězce, tj. počet kroků nutný k tomu, aby se marginální rozdělení řetězce lišilo od stacionárního rozdělení jen s povolenou nepřesností. Cílem této práce je popsat metodu odhadu mixing time, která využívá obecnou pravděpodobnostní techniku zvanou coupling. V první části textu bude vybudován teoretický aparát, na jehož základě tuto metodu odvodíme. Ve druhé části předvedeme její použití na klasických příkladech Markovových řetězců, kterým je například náhodná procházka po grafu. V závěru ukážeme odhad rychlosti konvergence Metropolisova řetězce pro přípustná obarvení grafu, jakožto typického příkladu MCMC algoritmu.

Klíčová slova: Markovův řetězec, stacionární rozdělení, konvergence, coupling, MCMC algoritmus

Title: Coupling and speed of convergence of discrete MCMC algorithms

Author: Martin Kalaš

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Michaela Prokešová, Ph.D.

Abstract: Convergence of the marginal distribution of a Markov chain to its stationary distribution is an essential property of this model with many applications in different fields of modern mathematics. Such typical applications are for example the Markov Chain Monte Carlo algorithms, which are useful for sampling from complicated probability distributions. A crucial point for usefulness of such algorithms is the so called mixing time of corresponding Markov chain, i.e. the number of steps the chain has to make for the difference between its current marginal distribution and stationary distribution to be sufficiently small. The main goal of this thesis is to describe a method for estimation of the mixing time based on a probability technique called coupling. In the first part we collect some definitions and propositions to show how the method works. Later the method is demonstrated on several traditional examples of Markov chains including e.g. random walk on a graph. In the end we study Metropolis chain on the set of proper colorings of a graph as a specific example of MCMC algorithm and show how to estimate its mixing time.

Keywords: Markov chain, stationary distribution, convergence, coupling, MCMC algorithm

Obsah

Úvod	3
1 Markovovy řetězce s diskrétním časem	5
2 Totální variace a coupling rozdělení	7
2.1 Vzdálenost v totální variaci	7
2.2 Coupling pravděpodobnostních rozdělení	9
2.3 Vzdálenost od stacionárního rozdělení	11
2.4 Věta o konvergenci	16
2.5 Mixing time	17
3 Coupling Markovových řetězců	19
3.1 Markovský coupling	19
3.2 Odhad vzdálenosti pomocí času koalescence	19
4 Odhady rychlosti konvergence	23
4.1 Náhodná procházka po cyklu	23
4.2 Náhodná procházka po toru	24
4.3 Vítězná série	27
5 Grand coupling a rychlost konvergence Metropolisova řetězce	29
5.1 Grand coupling	29
5.2 Metropolisův řetězec	29
5.3 Problém přípustného obarvení grafu	31
5.4 Odhad rychlosti konvergence	31
Seznam použité literatury	34

Použité značení

$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	množina přirozených, celých, reálných čísel
\mathbf{a}	řádkový vektor
\mathbf{a}^T	sloupcový vektor
A	matice
$(a_{i,j})$	prvek matice A v i -tém řádku a j -tém sloupci
$\mathbf{P}\{\cdot\}$	pravděpodobnost
$\mathbf{E}(\cdot)$	střední hodnota
$\mathbf{E}_\mu(\cdot)$	střední hodnota za podmínky, že počátečním rozdělením Markovova řetězce je μ
(X_t)	Markovův řetězec
X_t	hodnota Markovova řetězce v čase t
P	matice pravděpodobností přechodu Markovova řetězce
$P^r(x, y)$	pravděpodobnost přechodu ze stavu x do stavu y po r krocích
$P^t(x, \cdot)$	marginální rozdělení řetězce s počátečním stavem x v čase t
μP^t	marginální rozdělení řetězce s počátečním rozdělením μ v čase t
Ω	stavový prostor Markovova řetězce
$ \Omega $	mohutnost množiny Ω
μ, ν	pravděpodobnostní rozdělení na množině Ω
\mathcal{P}	prostor pravděpodobnostních rozdělení na množině Ω
π	stacionární rozdělení Markovova řetězce
$\Pi = (\pi_{i,j})_{i,j \in \Omega}$	matice, jejíž řádky tvoří vektor π , tj. $\pi_{i,j} = \pi(j)$
$\ \mu - \nu\ _{TV}$	vzdálenost rozdělení μ a ν v totální variaci
$d(t)$	vzdálenost marginálního rozdělení řetězce v čase t od stacionárního rozdělení
$\bar{d}(t)$	max. vzájemná vzdálenost marginálních rozděl. řetězce v čase t
$t_{mix}(\varepsilon)$	mixing time Markovova řetězce
τ_{couple}	čas koalescence
$G = (V, E)$	graf s množinou vrcholů V a množinou hran E
$\mathbf{1}_{\{A\}}$	indikátorová funkce jevu A

Úvod

Markovovy řetězce, nesoucí jméno po ruském matematiku Andreji Markovovi (1856-1922), jsou klasickým příkladem stochastických procesů, který má mnoho aplikací napříč různými obory matematiky, fyziky a teoretické informatiky. Jednou z důležitých vlastností Markovových řetězců, využívanou zvláště v moderních algoritmických aplikacích řetězců s diskretním časem a konečným stavovým prostorem, je konvergence marginálního rozdělení řetězce k jeho stacionárnímu rozdělení. Tuto vlastnost využívají např. *Markov Chain Monte Carlo algoritmy* (dále jen MCMC algoritmy), které slouží ke generování realizací ze zvoleného pravděpodobnostního rozdělení definovaného na nějaké typicky velmi rozsáhlé množině. Princip takového algoritmu je poměrně jednoduchý. Stačí na oné množině zkonstruovat takový Markovův řetězec, aby jeho stacionární rozdělení bylo rozdělení, z něhož chceme generovat realizace. Díky výše zmíněné konvergenci může po dostatečně velkém počtu kroků aktuální stav řetězce sloužit jako (přibližná) realizace ze stacionárního rozdělení.

Vzhledem k výše uvedenému principu MCMC algoritmů je důležité znát, jak rychle marginální rozdělení daného Markovova řetězce ke stacionárnímu rozdělení konverguje (tzv. *rychlost mixingu*). Jinými slovy, hledáme odpověď na otázku, jak velký počet kroků je třeba nechat řetězec udělat, než bude rozdíl mezi aktuálním marginálním rozdělením řetězce a jeho stacionárním rozdělením dostatečně malý. Tento potřebný počet kroků nazýváme *mixing time* Markovova řetězce. Právě problematikou odvození mezí pro mixing time se budeme v této práci zabývat.

Metod pro odhadnutí mixing time daného Markovova řetězce je mnoho. Některé zkoumají vlastní čísla matice pravděpodobností přechodu, jiné například vodivost vhodného asociovaného grafu. V této práci se zaměříme na metodu využívající *coupling*, což je důkazová technika vycházející z teorie pravděpodobnosti. Coupling umožňuje srovnávat rozdělení dvou náhodných veličin tím, že z nich vytvoříme náhodný vektor a vhodně určíme jeho sdružené rozdělení. Pro účely odhadnutí rychlosti konvergence Markovových řetězců budeme pomocí tzv. *Markovského couplingu* sdružovat celé řetězce jakožto posloupnosti náhodných veličin.

Pokud chceme zkoumat rychlost mixingu Markovova řetězce, musíme si také vybrat vhodný nástroj pro měření rozdílu mezi dvěma pravděpodobnostními rozděleními definovanými na stejné množině. K tomuto účelu budeme užívat *vzdálenost v totální variaci*. Tato volba je pro nás výhodná, protože (jak ukážeme) existuje úzká souvislost mezi vzdáleností v totální variaci dvou pravděpodobnostních rozdělení a tzv. *optimálním couplingem* náhodných veličin, pro něž jsou tato rozdělení marginálními. Důsledkem této souvislosti je, že maximální vzdálenost v totální variaci mezi dvěma pravděpodobnostními rozděleními lze odhadnout pomocí analýzy vhodného couplingu příslušně rozdělených náhodných veličin.

Jak uvidíme, analogický princip funguje i pro rozdělení celého Markovova řetězce. Zde zkoumáme Markovský coupling dvou řetězců a snažíme se získat odhad na *čas koalescence* tohoto couplingu, resp. snažíme se odhadnout pravděpodobnost jevu, že až do nějakého času t nedošlo k setkání obou řetězců. Právě pomocí této pravděpodobnosti totiž lze získat odhad na vzdálenost v totální variaci mezi marginálním rozdělením řetězce v čase t a stacionárním rozdělením, v důsledku

čehož dostaneme požadovanou horní mez pro mixing time příslušného řetězce.

Práce je organizována následovně: V první kapitole jsou připomenuty některé základní pojmy a důležité vlastnosti Markovových řetězců s diskrétním časem, které jsou ve zbytku textu často využívány. Ve druhé části jsou představeny pojmy totální variace, coupling pravděpodobnostních rozdělení a mixing time Markovova řetězce. Následně ukážeme některé jejich zajímavé vlastnosti a prozkoumáme souvislost mezi couplingem rozdělení a vzdáleností těchto rozdělení v totální variaci. Tato souvislost je společně s dalšími poznatky ve třetí kapitole užita k odvození metody odhadu pro mixing time Markovova řetězce. Ve čtvrté kapitole ilustrujeme užití této metody na některých klasických příkladech jednoduchých Markovových řetězců, jako je například náhodná procházka po cyklu. V poslední části práce pak ukážeme odhad mixing time Metropolisova řetězce na množině všech přípustných obarvení grafu, jakožto typického příkladu MCMC algoritmu.

1. Markovovy řetězce s diskretním časem

V této kapitole nejprve připomeneme některé základní pojmy a důležité vlastnosti Markovových řetězců s diskretním časem, které budeme v dalším textu využívat. Další podrobnosti o Markovových řetězcích včetně důkazů zde uvedených tvrzení lze nalézt v knihách [1] a [2].

Vzhledem k tomu, že se v této práci budeme věnovat pouze určitému typu Markovových řetězců, domluvme se na následujícím. Pokud neuvedeme jinak, budeme v dalším textu pod pojmem *Markovův řetězec s maticí P a stavovým prostorem Ω* rozumět homogenní Markovův řetězec s diskretním časem s maticí pravděpodobnostní přechodu P a konečnou množinou stavů Ω .

Definice. Markovův řetězec s maticí P a stavovým prostorem Ω se nazývá *nerozložitelný*, jestliže pro každou dvojici stavů $x, y \in \Omega$ existuje přirozené číslo t takové, že $P^t(x, y) > 0$.

Definice. Označme $\mathcal{T}(x) := \{t \geq 1 : P^t(x, x) > 0\}$ množinu všech možných časů návratu do stavu x . *Perioda* stavu x je definována jako největší společný dělitel čísel z $\mathcal{T}(x)$.

Řekneme, že Markovův řetězec je *aperiodický*, jestliže všechny jeho stavy mají periodu 1.

Věta 1. (Tvrzení 1.7 v [1]) *Nechť Markovův řetězec s maticí P je aperiodický a nerozložitelný. Potom existuje $r \in \mathbb{N}$ takové, že $P^r(x, y) > 0$ pro všechny $x, y \in \Omega$.*

Definice. Řekneme, že rozdělení π na Ω je *stacionární*, splňuje-li

$$\pi(y) = \sum_{x \in \Omega} \pi(x)P(x, y) \quad \text{pro všechna } y \in \Omega. \quad (1.1)$$

Maticový zápis tohoto vztahu je

$$\pi = \pi P. \quad (1.2)$$

Věta 2. (Věta 2.26 v [2]) *V nerozložitelném řetězci s konečně mnoha stavy stacionární rozdělení existuje.*

Definice. Nechť (X_t) je Markovův řetězec s maticí P a stavovým prostorem Ω . Říkáme, že rozdělení π na množině Ω splňuje *podmínky detailní rovnováhy* vzhledem k P , jestliže platí

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{pro všechna } x, y \in \Omega. \quad (1.3)$$

Věta 3. (Tvrzení 1.19 v [1]) *Nechť (X_t) je Markovův řetězec s maticí P a stavovým prostorem Ω . Jakékoli rozdělení π splňující podmínky detailní rovnováhy je stacionárním rozdělením řetězce (X_t) .*

Definice. *Reprezentací* matice pravděpodobností přechodu P na stavovém prostoru Ω pomocí náhodného přiřazení rozumíme dvojici (f, Z) , kde $f : \Omega \times \Lambda \rightarrow \Omega$ je *přechodová funkce*, Z je náhodná veličina nabývající hodnot z množiny Λ a pro všechna $x, y \in \Omega$ platí vztah

$$\mathbf{P}\{f(x, Z) = y\} = P(x, y). \quad (1.4)$$

Věta 4. (Tvrzení 1.5 v [1]) *Pro každou matici pravděpodobností přechodu na konečném stavovém prostoru existuje reprezentace pomocí náhodného přiřazení.*

Definice. *Jednoduchá náhodná procházka na množině $\{0, 1, \dots, n\}$ s absorpčními bariérami je Markovův řetězec se stavovým prostorem $\Omega = \{0, 1, \dots, n\}$ a maticí*

$$P(j, k) = \begin{cases} 1 & \text{pokud } k \equiv 0 \equiv j \text{ nebo } k \equiv n \equiv j \\ 1/2 & \text{pokud } k \equiv j + 1, \quad \text{kde } j = 1, 2, \dots, n - 1 \\ 1/2 & \text{pokud } k \equiv j - 1, \quad \text{kde } j = 1, 2, \dots, n - 1 \\ 0 & \text{jinak.} \end{cases} \quad (1.5)$$

Věta 5. (Tvrzení 2.1 v [1]) *Nechť (X_t) je jednoduchá náhodná procházka na množině $\{0, 1, \dots, n\}$ s absorpčními bariérami. Označme τ čas absorpce, tj. čas, kdy řetězec poprvé navštíví stav 0 nebo stav n . Předpokládejme, že platí $X_0 = k$, kde $0 \leq k \leq n$. Potom je*

$$\mathbf{E}_k(\tau) = k(n - k). \quad (1.6)$$

2. Totální variace a coupling rozdělení

V této kapitole nejprve představíme vzdálenost v totální variaci jakožto metriku, kterou budeme měřit vzdálenost mezi dvěma pravděpodobnostními rozděleními a uvedeme několik jejích užitečných charakterizací, které budeme užívat ve zbytku práce. V další sekci vysvětlíme pojem coupling pravděpodobnostních rozdělení a ukážeme jeho souvislost se vzdáleností v totální variaci. Dále budeme zkoumat vzdálenost marginálního rozdělení řetězce od stacionárního rozdělení, zejména pak vývoj této vzdálenosti v čase. V další části dokážeme Větu o konvergenci a nakonec představíme klíčový pojem mixing time Markovova řetězce.

Závěrem ještě poznamenejme, že většina materiálu prezentovaného v této části je k nalezení ve čtvrté kapitole knihy [1].

2.1 Vzdálenost v totální variaci

Abychom mohli zkoumat rychlost konvergence marginálního rozdělení Markovova řetězce k jeho stacionárnímu rozdělení, budeme potřebovat vhodný nástroj pro měření vzdálenosti mezi pravděpodobnostními rozděleními na stavovém prostoru řetězce. Tímto nástrojem pro nás bude vzdálenost v totální variaci.

Definice. *Vzdálenost v totální variaci* pravděpodobnostních rozdělení μ a ν na množině Ω je definována jako

$$\|\mu - \nu\|_{TV} = \max_{A \subset \Omega} |\mu(A) - \nu(A)|. \quad (2.1)$$

Vyjádření vzdálenosti v totální variaci jakožto maxima přes *všechny* podmnožiny Ω v (2.1) je sice názorné, ale není příliš pohodlné pro použití v odhadech této vzdálenosti. Uvedeme proto nyní dvě užitečné alternativní charakterizace vzdálenosti v totální variaci, a to pomocí součtu přes stavový prostor.

Věta 6. *Nechť μ a ν jsou dvě pravděpodobnostní rozdělení na Ω . Potom platí*

$$(i) \quad \|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| \quad (2.2)$$

$$(ii) \quad \|\mu - \nu\|_{TV} = \sum_{\substack{x \in \Omega \\ \mu(x) \geq \nu(x)}} [\mu(x) - \nu(x)]. \quad (2.3)$$

Důkaz. (i) Položme $B = \{x \in \Omega : \mu(x) \geq \nu(x)\}$. Nechť $A \subset \Omega$ je libovolný náhodný jev. Potom platí

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B) \quad (2.4)$$

a podobně

$$\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c). \quad (2.5)$$

Odečtením pravých stran (2.4) a (2.5) zjistíme, že oba horní odhady jsou stejné. Dále pokud vezmeme $A = B$ (nebo $A = B^c$), je výraz $|\mu(A) - \nu(A)|$ roven hornímu odhadu ve (2.4) (resp. ve (2.5)). Proto platí

$$\|\mu - \nu\|_{TV} = \frac{1}{2}[\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)] = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \quad (2.6)$$

(ii) Postupujeme zcela identicky jako v důkazu části (i). V závěru namísto (2.6) píšeme

$$\|\mu - \nu\|_{TV} = \mu(B) - \nu(B) = \sum_{x \in B} [\mu(x) - \nu(x)].$$

□

S pomocí *Věty 6* můžeme ukázat, že vzdálenost v totální variaci má skutečně vlastnosti metriky.

Věta 7. Označme \mathcal{P} prostor všech pravděpodobnostních rozdělení na Ω . Dvojice $(\mathcal{P}, \|\cdot, \cdot\|_{TV})$ je metrický prostor s metrikou $\|\cdot, \cdot\|_{TV} : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}$, definovanou vztahem $\|\cdot, \cdot\|_{TV} := \|\cdot - \cdot\|_{TV}$. To jest, pro každá $\mu, \nu, \eta \in \mathcal{P}$ platí

- (i) $\|\mu - \nu\|_{TV} \geq 0$
- (ii) $\|\mu - \nu\|_{TV} = 0 \iff \mu = \nu$
- (iii) $\|\mu - \nu\|_{TV} = \|\nu - \mu\|_{TV}$
- (iv) $\|\mu - \nu\|_{TV} \leq \|\mu - \eta\|_{TV} + \|\eta - \nu\|_{TV}$.

Důkaz. Vlastnosti (i),(ii) a (iii) plynou přímo z definice vzdálenosti v totální variaci. Stačí proto ukázat vlastnost (iv), tj. že vzdálenost v totální variaci splňuje trojúhelníkovou nerovnost. Podle *Věty 6* platí

$$\begin{aligned} \|\mu - \nu\|_{TV} &= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| \\ &= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \eta(x) + \eta(x) - \nu(x)| \\ &\leq \frac{1}{2} \sum_{x \in \Omega} [|\mu(x) - \eta(x)| + |\eta(x) - \nu(x)|] \\ &= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \eta(x)| + \frac{1}{2} \sum_{x \in \Omega} |\eta(x) - \nu(x)| \\ &= \|\mu - \eta\|_{TV} + \|\eta - \nu\|_{TV}, \end{aligned} \quad (2.7)$$

kde v (2.7) jsme užili trojúhelníkovou nerovnost pro absolutní hodnotu rozdílu reálných čísel. □

Nyní ukážeme, že vzdálenost dvou různých marginálních rozdělení řetězce se nevětší, posuneme-li řetězec v čase o krok vpřed.

Věta 8. Mějme Markovův řetězec s maticí P , množinou stavů Ω a necht' μ a ν jsou rozdělení na Ω . Potom platí

$$\|\mu P - \nu P\|_{TV} \leq \|\mu - \nu\|_{TV}. \quad (2.8)$$

Důkaz. Necht' μ, ν jsou jako ve znění věty. Pak podle Věty 6 máme

$$\begin{aligned} \|\mu P - \nu P\|_{TV} &= \frac{1}{2} \sum_{x \in \Omega} |\mu P(x) - \nu P(x)| \\ &= \frac{1}{2} \sum_{x \in \Omega} \left| \sum_{y \in \Omega} \mu(y) P(y, x) - \sum_{y \in \Omega} \nu(y) P(y, x) \right| \\ &= \frac{1}{2} \sum_{x \in \Omega} \left| \sum_{y \in \Omega} P(y, x) [\mu(y) - \nu(y)] \right| \\ &\leq \frac{1}{2} \sum_{x \in \Omega} \sum_{y \in \Omega} P(y, x) |\mu(y) - \nu(y)| \\ &= \frac{1}{2} \sum_{y \in \Omega} |\mu(y) - \nu(y)| \sum_{x \in \Omega} P(y, x) \\ &= \frac{1}{2} \sum_{y \in \Omega} |\mu(y) - \nu(y)| \\ &= \|\mu - \nu\|_{TV}. \end{aligned} \quad (2.9)$$

V nerovnosti (2.9) jsme použili trojúhelníkovou nerovnost a v závěru znovu Větu 6. \square

Všimněme si, že po dosazení $\mu := \mu P^t$ a $\nu := \pi$ do vztahu z předešlé věty dostáváme díky vlastnosti (1.2)

$$\|\mu P^{t+1} - \pi\|_{TV} \leq \|\mu P^t - \pi\|_{TV}. \quad (2.10)$$

To jest, posune-li se řetězec s počátečním rozdělením μ v čase o krok vpřed, vzdálenost jeho marginálního rozdělení od stacionárního rozdělení se neztvětší.

2.2 Coupling pravděpodobnostních rozdělení

Definice. Pojmeme *coupling* pravděpodobnostních rozdělení μ a ν budeme rozumět takový náhodný vektor (X, Y) , že marginální rozdělení X je μ a marginální rozdělení Y je ν , tj. coupling (X, Y) splňuje $\mathbf{P}\{X = x\} = \mu(x)$ a $\mathbf{P}\{Y = y\} = \nu(y)$.

Poznámka. Jak je vidět z definice, pro každou dvojici pravděpodobnostních rozdělení μ, ν není jejich coupling zdaleka určen jednoznačně. V dalším textu budeme konkrétní coupling rozdělení μ a ν specifikovat buď pomocí dvourozměrného rozdělení náhodného vektoru (X, Y) a nebo (pro usnadnění) jen slovním popisem vzájemné závislosti obou veličin X a Y .

Pro libovolná pravděpodobnostní rozdělení μ a ν zřejmě existuje nezávislý coupling těchto rozdělení, tj. takový coupling (X, Y) , že platí

$$\mathbf{P}\{X = x, Y = y\} = \mathbf{P}\{X = x\}\mathbf{P}\{Y = y\} = \mu(x)\nu(y)$$

pro všechna $x, y \in \Omega$. Naopak speciálně pro $\mu = \nu$, můžeme vytvořit takový coupling (X, Y) , že bude platit $\mathbf{P}\{X = Y\} = 1$. To ale nebude možné, pokud budou rozdělení μ a ν různá. Na otázku jak moc se můžeme k této vlastnosti přiblížit dává odpověď následující věta.

Věta 9. *Nechť μ a ν jsou pravděpodobnostní rozdělení na Ω . Potom*

$$\|\mu - \nu\|_{TV} = \inf\{\mathbf{P}\{X \neq Y\} : (X, Y) \text{ je coupling } \mu \text{ a } \nu\}. \quad (2.11)$$

Důkaz. Nejprve nechť (X, Y) je libovolný coupling μ a ν a $A \in \Omega$ je libovolný jev. Pak máme

$$\begin{aligned} \mu(A) - \nu(A) &= \mathbf{P}\{X \in A\} - \mathbf{P}\{Y \in A\} \\ &\leq \mathbf{P}\{X \in A, Y \notin A\} \end{aligned} \quad (2.12)$$

$$\leq \mathbf{P}\{X \neq Y\}. \quad (2.13)$$

Nerovnost (2.12) platí, protože jsme přičetli nezáporný výraz $\mathbf{P}\{X \notin A, Y \in A\}$. Z (2.13) potom plyne

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

K důkazu opačné nerovnosti postačí, zkonstruujeme-li coupling splňující

$$\|\mu - \nu\|_{TV} = \mathbf{P}\{X \neq Y\}.$$

Toho dosáhneme, pokud definujeme veličiny X a Y tak, aby nabývaly stejných hodnot co nejčastěji je to možné. Položme

$$p = \sum_{x \in \Omega} \min\{\mu(x), \nu(x)\}.$$

Máme

$$\sum_{x \in \Omega} \min\{\mu(x), \nu(x)\} = \sum_{\substack{x \in \Omega, \\ \mu(x) \leq \nu(x)}} \mu(x) + \sum_{\substack{x \in \Omega, \\ \mu(x) > \nu(x)}} \nu(x). \quad (2.14)$$

Přičtením a odečtením výrazu $\sum_{x: \mu(x) > \nu(x)} \mu(x)$ k pravé straně (2.14) dostáváme

$$\sum_{x \in \Omega} \min\{\mu(x), \nu(x)\} = 1 - \sum_{\substack{x \in \Omega, \\ \mu(x) > \nu(x)}} [\mu(x) - \nu(x)].$$

Užitím *Věty 6* dostáváme z předchozí rovnice

$$\sum_{x \in \Omega} \min\{\mu(x), \nu(x)\} = 1 - \|\mu - \nu\|_{TV} = p. \quad (2.15)$$

Nechť W je náhodná veličina s alternativním rozdělením s parametrem p . Podmíněné rozdělení náhodného vektoru (X, Y) pak definujeme následovně:

- (i) Jestliže W nabude hodnoty 1, položíme $X = Y = Z$, kde Z je náhodná veličina s rozdělením

$$\gamma(x) = \frac{\min\{\mu(x), \nu(x)\}}{p}.$$

- (ii) Jestliže W nabude hodnoty 0, bude mít veličina X rozdělení

$$\alpha(x) = \begin{cases} \frac{\mu(x) - \nu(x)}{\|\mu - \nu\|_{TV}} & \text{jestliže } \mu(x) > \nu(x), \\ 0 & \text{jinak} \end{cases}$$

a rozdělení veličiny Y bude

$$\beta(x) = \begin{cases} \frac{\nu(x) - \mu(x)}{\|\mu - \nu\|_{TV}} & \text{jestliže } \nu(x) > \mu(x), \\ 0 & \text{jinak,} \end{cases}$$

přičemž X a Y budou nezávislé.

Druhá část *Věty 6* zajišťuje, že α a β skutečně jsou pravděpodobnostní rozdělení. Dále vzhledem k tomu, že platí rovnosti

$$\begin{aligned} p\gamma + (1-p)\alpha &= \mu, \\ p\gamma + (1-p)\beta &= \nu, \end{aligned}$$

je μ marginálním rozdělením veličiny X a ν marginálním rozdělením veličiny Y , takže vektor (X, Y) je skutečně coupling rozdělení μ a ν . Dále si všimněme, že pokud $W = 0$, pak nutně $X \neq Y$, protože nosiče rozdělení α a β jsou disjunktní podmnožiny stavového prostoru Ω . Platí tedy $X = Y$, právě když $W = 1$ a proto podle (2.15) je

$$\mathbf{P}\{X \neq Y\} = \|\mu - \nu\|_{TV}.$$

□

Definice. Coupling (X, Y) rozdělení μ a ν splňující $\|\mu - \nu\|_{TV} = \mathbf{P}\{X \neq Y\}$ nazýváme *optimální coupling*.

Poznámka. V důkazu *Věty 9* jsme ukázali, že pro každá rozdělení μ a ν existuje optimální coupling.

Coupling je velmi užitečný pojem, protože umožňuje problém porovnání pravděpodobnostních rozdělení převést na problém porovnání náhodných veličin. Ve *Větě 9* jsme ukázali, jak lze pomocí couplingu vyjádřit vzdálenost v totální variaci. Tato věta bude v dalších kapitolách důležitým nástrojem pro hledání horního odhadu vzdálenosti mezi marginálním rozdělením řetězce v čase t a stacionárním rozdělením, což je důležitá charakteristika, které věnujeme následující odstavce.

2.3 Vzdálenost od stacionárního rozdělení

V dalším textu budeme užívat označení

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} \quad (2.16)$$

pro vzdálenost marginálního rozdělení řetězce v čase t od stacionárního rozdělení.

Dále bude užitečné zavést symbol

$$\bar{d}(t) := \max_{x,y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \quad (2.17)$$

pro maximální vzdálenost mezi marginálními rozděleními v čase t pro řetězce s různými počátečními stavy.

Nyní popíšeme vztah mezi těmito dvěma charakteristikami a některé jejich vlastnosti.

Lemma 10. *Pro každé $t \in \mathbb{N}$ platí*

$$d(t) \leq \bar{d}(t) \leq 2d(t). \quad (2.18)$$

Důkaz. Z trojúhelníkové nerovnosti pro vzdálenost v totální variaci plyne

$$\max_{x,y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} + \max_{y \in \Omega} \|\pi - P^t(y, \cdot)\|_{TV}.$$

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

S využitím vlastnosti (1.1) stacionárního rozdělení π dostáváme, že pro každou $A \subset \Omega$ platí

$$\begin{aligned} \pi(A) &= \sum_{x \in A} \sum_{y \in \Omega} \pi(y) P^t(y, x) \\ &= \sum_{y \in \Omega} \pi(y) \sum_{x \in A} P^t(y, x) \\ &= \sum_{y \in \Omega} \pi(y) P^t(y, A). \end{aligned}$$

Díky tomu můžeme psát

$$\begin{aligned} \|P^t(x, \cdot) - \pi\|_{TV} &= \max_{A \subset \Omega} |P^t(x, A) - \pi(A)| \\ &= \max_{A \subset \Omega} \left| \sum_{y \in \Omega} \pi(y) [P^t(x, A) - P^t(y, A)] \right| \\ &\leq \sum_{y \in \Omega} \pi(y) \max_{A \subset \Omega} |P^t(x, A) - P^t(y, A)| \quad (2.19) \end{aligned}$$

$$\begin{aligned} &= \sum_{y \in \Omega} \pi(y) \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \\ &\leq \sum_{y \in \Omega} \pi(y) \max_{z \in \Omega} \|P^t(x, \cdot) - P^t(z, \cdot)\|_{TV} \\ &= \max_{y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}, \quad (2.20) \end{aligned}$$

kde ve (2.19) jsme využili trojúhelníkové nerovnosti a faktu, že maximum součtu není větší než součet maxim. Výsledek $d(t) \leq \bar{d}(t)$ dostaneme maximalizací (2.20) přes $x \in \Omega$. \square

Lemma 11. *Funkce $\bar{d}(t)$ je submultiplikativní, tj. pro každé $s, t \in \mathbb{N}$ platí*

$$\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t).$$

Důkaz. Zvolme pevně nějakou dvojici stavů $x, y \in \Omega$. Necht' (X_s, Y_s) je optimální coupling rozdělení $P^s(x, \cdot)$ a $P^s(y, \cdot)$, tj. necht' platí

$$\|P^s(x, \cdot) - P^s(y, \cdot)\|_{TV} = \mathbf{P}\{X_s \neq Y_s\}. \quad (2.21)$$

Jelikož P^{s+t} je maticovým součinem P^s a P^t a rozdělení X_s je $P^s(x, \cdot)$, máme

$$P^{s+t}(x, w) = \sum_{z \in \Omega} P^s(x, z)P^t(z, w) = \sum_{z \in \Omega} \mathbf{P}\{X_s = z\}P^t(z, w) = \mathbf{E}(P^t(X_s, w)).$$

Analogicky obdržíme vztah $P^{s+t}(y, w) = \mathbf{E}(P^t(Y_s, w))$ a jejich odečtením pak

$$\begin{aligned} P^{s+t}(x, w) - P^{s+t}(y, w) &= \mathbf{E}(P^t(X_s, w)) - \mathbf{E}(P^t(Y_s, w)) \\ &= \mathbf{E}(P^t(X_s, w) - P^t(Y_s, w)). \end{aligned} \quad (2.22)$$

Sečtení absolutních hodnot obou stran (2.22) přes $w \in \Omega$, vydělení dvěma a následná aplikace *Věty 6* dávají

$$\|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\|_{TV} = \frac{1}{2} \sum_{w \in \Omega} |\mathbf{E}(P^t(X_s, w) - P^t(Y_s, w))|.$$

Jelikož pro náhodnou veličinu Z s konečnou střední hodnotou platí $|\mathbf{E}Z| \leq \mathbf{E}|Z|$, je pravá strana poslední rovnosti určitě menší nebo rovna výrazu

$$\mathbf{E} \left(\frac{1}{2} \sum_{w \in \Omega} |P^t(X_s, w) - P^t(Y_s, w)| \right). \quad (2.23)$$

Použijeme-li na (2.23) znovu *Větu 6* dostáváme celkově

$$\|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\|_{TV} \leq \mathbf{E} (\|P^t(X_s, \cdot) - P^t(Y_s, \cdot)\|_{TV}).$$

Vzdálenost $\|P^t(X_s, \cdot) - P^t(Y_s, \cdot)\|_{TV}$ je určitě nulová pokud $X_s = Y_s$ a navíc je vždy shora ohraničená funkcí $\bar{d}(t)$, takže můžeme psát

$$\|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\|_{TV} \leq \bar{d}(t) \mathbf{E}(\mathbf{1}\{X_s \neq Y_s\}) = \bar{d}(t) \mathbf{P}\{X_s \neq Y_s\}. \quad (2.24)$$

Díky (2.21) je ale pravděpodobnost vpravo rovna $\|P^s(x, \cdot) - P^s(y, \cdot)\|_{TV}$. K dokončení důkazu proto stačí maximalizovat obě strany (2.24) přes $x, y \in \Omega$. \square

Důsledek 12. *Pro každá $c, t \in \mathbb{N}$ platí*

$$d(ct) \leq \bar{d}(ct) \leq (\bar{d}(t))^c. \quad (2.25)$$

Důkaz. První nerovnost platí díky *Lemmatu 10*, druhá díky *Lemmatu 11*. \square

Následující věta ukazuje, že vzdálenost marginálního rozdělení řetězce od stationárního rozdělení se s rostoucím časem nezvětšuje, tj. $d(t)$ je neklesající funkcí t .

Věta 13. *Buď $d(t)$ definováno jako v (2.16). Potom pro každé $t \in \mathbb{N}$ platí*

$$d(t+1) \leq d(t). \quad (2.26)$$

Důkaz. Podle Věty 8 máme

$$d(t+1) = \max_{x \in \Omega} \|P^{t+1}(x, \cdot) - \pi\|_{TV} \leq \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} = d(t).$$

□

Podívejme se ještě, co se stane, dosadíme-li do vztahu z Věty 8 $\mu := P^t(x, \cdot)$ a $\nu := P^t(y, \cdot)$. Dostáváme

$$\|P^{t+1}(x, \cdot) - P^{t+1}(y, \cdot)\|_{TV} \leq \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}.$$

Odkud však maximalizací přes $x, y \in \Omega$ je

$$\bar{d}(t+1) \leq \bar{d}(t), \quad (2.27)$$

tedy i $\bar{d}(t)$ je neklesající funkcí času.

Na závěr této sekce si ještě ukážeme alternativní vyjádření $d(t)$ a $\bar{d}(t)$, a to pomocí suprema přes množinu všech počátečních rozdělání (resp. všech dvojic počátečních rozdělání) na Ω . Tato charakterizace je obecnější, než vyjádření (2.16) a (2.17), která berou v úvahu jen degenerovaná počáteční rozdělání, a ukazuje tak smysluplnost $d(t)$ a $\bar{d}(t)$ jako charakteristik vzdálenosti marginálního rozdělání řetězce v čase t od stacionárního rozdělání.

Věta 14. *Nechť \mathcal{P} značí prostor všech pravděpodobnostních rozdělání na množině Ω a buď $d(t), \bar{d}(t)$ definovány jako v (2.16), resp. (2.17). Potom platí*

$$(i) \quad d(t) = \sup_{\mu \in \mathcal{P}} \|\mu P^t - \pi\|_{TV} \quad (2.28)$$

$$(ii) \quad \bar{d}(t) = \sup_{\mu, \nu \in \mathcal{P}} \|\mu P^t - \nu P^t\|_{TV}. \quad (2.29)$$

Důkaz. (i) Vzhledem k tomu, že množina stavů Ω je konečná a že degenerovaná rozdělání jsou obsažena v \mathcal{P} , můžeme psát

$$\begin{aligned} d(t) &= \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} = \sup_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} \\ &\leq \sup_{\mu \in \mathcal{P}} \|\mu P^t - \pi\|_{TV}. \end{aligned}$$

Naopak pro každé $\mu \in \mathcal{P}$ s využitím Věty 6 a trojúhelníkové nerovnosti dostáváme

$$\begin{aligned} \|\mu P^t - \pi\|_{TV} &= \frac{1}{2} \sum_{x \in \Omega} |\mu P^t(x) - \pi(x)| \\ &= \frac{1}{2} \sum_{x \in \Omega} \left| \sum_{y \in \Omega} \mu(y) P^t(y, x) - \sum_{y \in \Omega} \mu(y) \pi(x) \right| \\ &\leq \frac{1}{2} \sum_{x \in \Omega} \sum_{y \in \Omega} \mu(y) |P^t(y, x) - \pi(x)| \\ &= \sum_{y \in \Omega} \mu(y) \frac{1}{2} \sum_{x \in \Omega} |P^t(y, x) - \pi(x)| \\ &= \sum_{y \in \Omega} \mu(y) \|P^t(y, \cdot) - \pi\|_{TV} \\ &\leq \sum_{y \in \Omega} \mu(y) \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} \\ &= \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV}. \end{aligned}$$

Odtud plyne

$$\sup_{\mu \in \mathcal{P}} \|\mu P^t - \pi\|_{TV} \leq d(t).$$

(ii) Díky konečnosti stavového prostoru a faktu, že degenerovaná počáteční rozdělení jsou obsažena v \mathcal{P} opět snadno dostáváme

$$\bar{d}(t) \leq \sup_{\mu, \nu \in \mathcal{P}} \|\mu P^t - \nu P^t\|_{TV}.$$

Naopak pro každá počáteční rozdělení $\mu, \nu \in \Omega$ podobně jako v (i) máme

$$\begin{aligned} \|\mu P^t - \nu P^t\|_{TV} &= \frac{1}{2} \sum_{x \in \Omega} |\mu P^t(x) - \nu P^t(x)| \\ &= \frac{1}{2} \sum_{x \in \Omega} \left| \sum_{y \in \Omega} \mu(y) P^t(y, x) - \nu P^t(x) \right| \\ &\leq \frac{1}{2} \sum_{x \in \Omega} \sum_{y \in \Omega} \mu(y) |P^t(y, x) - \nu P^t(x)| \\ &= \sum_{y \in \Omega} \mu(y) \|P^t(y, \cdot) - \nu P^t\|_{TV} \\ &\leq \sum_{y \in \Omega} \mu(y) \max_{x \in \Omega} \|P^t(x, \cdot) - \nu P^t\|_{TV} \\ &= \max_{x \in \Omega} \|P^t(x, \cdot) - \nu P^t\|_{TV}. \end{aligned} \tag{2.30}$$

Výraz $\|P^t(x, \cdot) - \nu P^t\|_{TV}$ můžeme ale opět odhadnout shora

$$\begin{aligned} \|P^t(x, \cdot) - \nu P^t\|_{TV} &= \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \nu P^t(y)| \\ &= \frac{1}{2} \sum_{y \in \Omega} \left| P^t(x, y) - \sum_{z \in \Omega} \nu(z) P^t(z, y) \right| \\ &\leq \frac{1}{2} \sum_{y \in \Omega} \sum_{z \in \Omega} \nu(z) |P^t(x, y) - P^t(z, y)| \\ &= \sum_{z \in \Omega} \nu(z) \|P^t(x, \cdot) - P^t(z, \cdot)\|_{TV} \\ &\leq \sum_{z \in \Omega} \nu(z) \max_{y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \\ &= \max_{y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}. \end{aligned} \tag{2.31}$$

Zkombinováním (2.30) a (2.31) tudíž dostáváme

$$\|\mu P^t - \nu P^t\|_{TV} \leq \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}$$

pro každá $\mu, \nu \in \mathcal{P}$. Odtud plyne

$$\sup_{\mu, \nu \in \mathcal{P}} \|\mu P^t - \nu P^t\|_{TV} \leq \bar{d}(t)$$

a důkaz věty je tak dokončen. □

2.4 Věta o konvergenci

Jelikož cílem tohoto textu je popsat metodu pro měření rychlosti konvergence marginálního rozdělení Markovova řetězce k jeho stacionárnímu rozdělení, bude vhodné podrobněji ukázat, zda — a případně za jakých předpokladů — k této konvergenci vůbec dochází. Následující věta ukazuje, že pro nerozložitelný a aperiodický Markovův řetězec skutečně marginální rozdělení řetězce konverguje ke stacionárnímu, a to v tom smyslu, že jejich vzdálenost v totální variaci jde limitně k 0.

Jak bývá u takto důležitých poznatků často běžné, existuje více možností jak *Větu o konvergenci* dokázat. Zde uvedená verze důkazu využívá rozkladu Markovova řetězce na opakované generování realizací ze stacionárního rozdělení a jiného řetězce.

Věta 15. *Nechť (X_t) je nerozložitelný aperiodický Markovův řetězec s maticí P a stacionárním rozdělením π . Potom existují konstanty $\alpha \in (0, 1)$ a $C \geq 0$ takové, že platí*

$$d(t) \leq C\alpha^t. \quad (2.32)$$

Důkaz. Vzhledem k nerozložitelnosti a aperiodicitě řetězce existuje podle *Věty 1* takové $r \in \mathbb{N}$, že matice P^r má všechny prvky kladné nenulové. Dále nechť $\Pi_{(|\Omega| \times |\Omega|)}$ je matice, jejíž každý řádek je vektor π . Pro dostatečně malé $\delta > 0$ máme

$$P^r(x, y) \geq \delta\pi(y)$$

pro všechny $x, y \in \Omega$. Položíme-li proto $\theta := 1 - \delta$, můžeme rovnici

$$P^r = (1 - \theta)\Pi + \theta Q \quad (2.33)$$

definovat stochastickou matici Q .

Není těžké ověřit, že pro každou stochastickou matici $M_{(|\Omega| \times |\Omega|)}$ platí

$$M\Pi = \Pi. \quad (2.34)$$

Označíme-li $(m\pi)_{ij}$ prvky matice $M\Pi$, je totiž

$$(m\pi)_{ij} = \sum_{k=1}^{|\Omega|} m_{ik}\pi_{kj} = \sum_{k=1}^{|\Omega|} m_{ik}\pi_j = \pi_j.$$

Podobně pro každou matici $M_{(|\Omega| \times |\Omega|)}$ splňující $\pi M = \pi$ platí

$$\Pi M = \Pi. \quad (2.35)$$

Předpokládanou rovnost $\pi M = \pi$ totiž můžeme rozepsat jako $\pi_j = \sum_{k=1}^{|\Omega|} \pi_k m_{kj}$ a díky tomu je potom

$$(\pi m)_{ij} = \sum_{k=1}^{|\Omega|} \pi_k m_{kj} = \sum_{k=1}^{|\Omega|} \pi_k m_{kj} = \pi_j.$$

Nyní indukci ukážeme, že pro $k \geq 1$ platí vztah

$$P^{rk} = (1 - \theta^k)\Pi + \theta^k Q^k. \quad (2.36)$$

Pro $k = 1$ rovnost platí díky (2.33). Předpokládejme nyní, že (2.36) platí pro $k = n$. Máme

$$\begin{aligned} P^{r(n+1)} &= P^{rn} P^r = [(1 - \theta^n)\Pi + \theta^n Q^n] P^r \\ &= (1 - \theta^n)\Pi P^r + \theta^n Q^n P^r \\ &= (1 - \theta^n)\Pi P^r + (1 - \theta)\theta^n Q^n \Pi + \theta^{n+1} Q^{n+1}, \end{aligned} \quad (2.37)$$

kde v poslední rovnosti jsme použili (2.33). Dále vzhledem k tomu, že Q^n je stochastická a $\pi P^r = \pi$, můžeme použít (2.34) a (2.35). Z rovnosti (2.37) tedy dostáváme

$$\begin{aligned} P^{r(n+1)} &= (1 - \theta^n)\Pi + (1 - \theta)\theta^n \Pi + \theta^{n+1} Q^{n+1} \\ &= (1 - \theta^{n+1})\Pi + \theta^{n+1} Q^{n+1}, \end{aligned}$$

čímž je indukce dokončena.

Po vynásobení právě dokázané rovnosti (2.36) zprava výrazem P^j máme

$$P^{r^{k+j}} = (1 - \theta^k)\Pi P^j + \theta^k Q^k P^j = (1 - \theta^k)\Pi + \theta^k Q^k P^j$$

a odtud

$$P^{r^{k+j}} - \Pi = \theta^k (Q^k P^j - \Pi). \quad (2.38)$$

Zvolme nyní nějaký stav $x_0 \in \Omega$. Sečtením absolutních hodnot prvků v řádku odpovídajícímu x_0 na obou stranách maticové rovnosti (2.38) a následným vydělením obou stran dvěma získáváme

$$\frac{1}{2} \sum_{y \in \Omega} |P^{r^{k+j}}(x_0, y) - \pi(y)| = \theta^k \frac{1}{2} \sum_{y \in \Omega} |Q^k P^j(x_0, y) - \pi(y)|. \quad (2.39)$$

Uvědomíme-li si, že vzdálenost v totální variaci dvou pravděpodobnostních rozdělení je vždy menší nebo rovna jedné, stačí k dokončení celého důkazu použít na obě strany poslední rovnosti *Větu 6*. Pak totiž dostaneme

$$\|P^{r^{k+j}}(x_0, \cdot) - \pi\|_{TV} = \theta^k \|Q^k P^j(x_0, \cdot) - \pi\|_{TV} \leq \theta^k,$$

odkud již plyne tvrzení věty vzhledem k tomu, že stav x_0 byl zvolen libovolně. \square

2.5 Mixing time

V předešlém textu jsme ukázali, že vzdálenost marginálního rozdělení řetězce od stacionárního rozdělení je neklesající funkcí času, a že tato vzdálenost konverguje k 0. Nyní začneme zkoumat rychlost této konvergence. Náš problém se dá zformulovat následovně: Pro daný Markovův řetězec a danou přesnost ε nás zajímá, jak velké musí být t_0 , aby pro všechna $t \geq t_0$ platilo $d(t) \leq \varepsilon$. Díky tomu, že zmiňovaná konvergence je monotónní, stačí nalézt nejmenší t_0 , které splňuje $d(t_0) \leq \varepsilon$. Toto t_0 budeme nazývat mixing time daného řetězce.

Definice. *Mixing time* Markovova řetězce definujeme jako

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\} \quad (2.40)$$

a speciálně budeme značit

$$t_{mix} := t_{mix}(1/4). \quad (2.41)$$

Lemma 16. *Pro každé $\ell \in \mathbb{N}$ platí*

$$d(\ell t_{mix}(\varepsilon)) \leq (2\varepsilon)^\ell. \quad (2.42)$$

Důkaz. Postupným užitím *Lemmatu 10*, *Důsledku 12* a znovu *Lemmatu 10* dostáváme

$$d(\ell t_{mix}(\varepsilon)) \leq \bar{d}(\ell t_{mix}(\varepsilon)) \leq (\bar{d}(t_{mix}(\varepsilon)))^\ell \leq (2\varepsilon)^\ell.$$

□

Všimněme si, že pokud vezmeme $\varepsilon = 1/4$, máme podle *Lemmatu 13*

$$d(\ell t_{mix}) \leq 2^{-\ell} \quad (2.43)$$

a odtud pro $\ell = \log_2 \varepsilon^{-1}$ vzhledem k (2.40) dostáváme

$$t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}, \quad (2.44)$$

což je užitečná nerovnost, umožňující shora odhadnout $t_{mix}(\varepsilon)$, pokud již máme nějaký odhad na t_{mix} .

3. Coupling Markovových řetězců

V této kapitole nejprve představíme klíčový pojem coupling Markovových řetězců. Dále odvodíme efektivní metodu, jak lze pomocí času koalescence vhodně zvoleného couplingu řetězců s různými počátečními stavy získat odhad na vzdálenost marginálního rozdělení řetězce od stacionárního rozdělení a v důsledku tak shora odhadnout mixing time daného Markovova řetězce.

3.1 Markovský coupling

Definice. *Coupling Markovových řetězců* s maticí P definujeme jako náhodný proces $(X_t, Y_t)_{t=0}^\infty$ s vlastností, že (X_t) a (Y_t) jsou Markovovy řetězce s maticí P , jejichž počáteční rozdělení může být různé.

Každý coupling Markovových řetězců (X_t) a (Y_t) s maticí P můžeme upravit tak, aby platilo

$$\text{je-li } X_s = Y_s, \text{ potom } X_t = Y_t \text{ pro každé } t \geq s. \quad (3.1)$$

Zkonstruování takového couplingu je snadné. Stačí ponechat oba řetězce vyvíjet se podle původního couplingu, a to až do času, kdy se poprvé setkají. Od té doby coupling predefinujeme tak, aby trajektorie obou řetězců byly stejné. To provedeme tak, že necháme např. řetězec (X_t) vyvíjet se dál podle matice P a přitom klademe $Y_t := X_t$.

Označení. Pro coupling řetězců (X_t) a (Y_t) s počátečními stavy $X_0 = x$ a $Y_0 = y$ budeme pravděpodobnost na prostoru, kde je proces (X_t, Y_t) definován, značit $\mathbf{P}_{x,y}$.

Definice. Mějme Markovův řetězec s maticí P a množinou stavů Ω . *Markovský coupling* odpovídající matici P je Markovův řetězec s množinou stavů $\Omega \times \Omega$, jehož matice pravděpodobností přechodu Q splňuje

$$(i) \text{ pro všechny } x, y, x' \in \Omega \text{ platí } \sum_{y'} Q((x, y), (x', y')) = P(x, x')$$

$$(ii) \text{ pro všechny } x, y, y' \in \Omega \text{ platí } \sum_{x'} Q((x, y), (x', y')) = P(y, y').$$

Poznámka. Zřejmě každý Markovský coupling je coupling Markovových řetězců ve smyslu definice na začátku této kapitoly. Všechny couplingy použité v této práci budou Markovské.

3.2 Odhad vzdálenosti pomocí času koalescence

Definice. Nechtě (X_t) a (Y_t) jsou Markovovy řetězce. Náhodnou veličinu zvanou *čas koalescence* definujeme jako

$$\tau_{couple} := \min \{t : X_t = Y_t\}. \quad (3.2)$$

Všimněme si, že pro coupling Markovových řetězců splňující podmínku (3.1) je τ_{couple} čas, od kterého oba řetězce splynou.

Nyní uvedeme důležitou větu, jejíž důsledek bude klíčovým nástrojem k odhadnutí mixing time Markovova řetězce, který budeme využívat v následujících kapitolách.

Věta 17. *Nechť $\{(X_t, Y_t)\}$ je coupling splňující (3.1), pro který je $X_0 = x$ a $Y_0 = y$. Potom pro každé $t \in \mathbb{N}$ je*

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \mathbf{P}_{x,y}\{\tau_{couple} > t\}. \quad (3.3)$$

Důkaz. Všimněme si, že $P^t(x, z) = \mathbf{P}_{x,y}\{X_t = z\}$ a $P^t(y, z) = \mathbf{P}_{x,y}\{Y_t = z\}$, tudíž (X_t, Y_t) je coupling rozdělení $P^t(x, \cdot)$ a $P^t(y, \cdot)$. Aplikací Věty 9 tak dostáváme

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \mathbf{P}_{x,y}\{X_t \neq Y_t\}.$$

Díky předpokladu (3.1) ale platí $\mathbf{P}_{x,y}\{X_t \neq Y_t\} = \mathbf{P}_{x,y}\{\tau_{couple} > t\}$, což spolu s předchozí nerovností dává závěr. \square

Důsledek 18. *Předpokládejme, že pro každou dvojici stavů $x, y \in \Omega$ existuje coupling (X_t, Y_t) s počátečními stavy $X_0 = x$ a $Y_0 = y$. Potom*

$$d(t) \leq \max_{x,y \in \Omega} \mathbf{P}_{x,y}\{\tau_{couple} > t\}.$$

Důkaz. Podle Lemmatu 10 a předchozí věty můžeme psát

$$d(t) \leq \bar{d}(t) = \max_{x,y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \max_{x,y \in \Omega} \mathbf{P}_{x,y}\{\tau_{couple} > t\}.$$

\square

Naši obecnou strategií pro odhad mixing time Markovova řetězce tedy bude zkonstruovat nějaký coupling dvou verzí tohoto řetězce $(X_t, Y_t)_{t=0}^{\infty}$ s různými počátečními stavy, který bude splňovat (3.1). Poté se budeme snažit shora odhadnout čas koalescence tohoto couplingu, resp. jeho maximum přes všechny možné dvojice počátečních stavů. Často pro nás bude pohodlné využít verzi *Markovovy nerovnosti*. Ta říká, že pro náhodnou veličinu X s konečnou střední hodnotou a libovolné $\varepsilon > 0$ platí

$$\mathbf{P}\{|X| > \varepsilon\} \leq \frac{\mathbf{E}|X|}{\varepsilon}. \quad (3.4)$$

Platnost tohoto vztahu lze ověřit následovně:

$$\begin{aligned} \mathbf{E}|X| &= \mathbf{E}[|X|(\mathbf{1}_{\{|X|>\varepsilon\}} + \mathbf{1}_{\{|X|\leq\varepsilon\}})] \geq \mathbf{E}(|X| \cdot \mathbf{1}_{\{|X|>\varepsilon\}}) \\ &\geq \mathbf{E}(\varepsilon \cdot \mathbf{1}_{\{|X|>\varepsilon\}}) = \varepsilon \cdot \mathbf{P}\{|X| > \varepsilon\}. \end{aligned}$$

Aplikací nerovnosti (3.4) na čas koalescence a maximalizací přes všechny dvojice počátečních stavů potom získáváme užitečnou nerovnost

$$\max_{x,y \in \Omega} \mathbf{P}_{x,y}\{\tau > t\} \leq \frac{\max_{x,y} \mathbf{E}_{x,y}(\tau)}{t}, \quad (3.5)$$

jejíž pravá strana se dá díky Důsledku 18 užít pro odhadnutí $d(t)$ shora, čímž zároveň získáme odhad na mixing time našeho řetězce.

Vzhledem k výše popsané metodě je přirozené se ptát, zda pro každou matici pravděpodobností přechodu existuje takový coupling řetězců s touto maticí, že čas, kdy se oba řetězce poprvé setkají je konečný. Odpověď na tuto otázku dává následující věta.

Věta 19. Pro každou dvojici Markovových řetězců $(X_t), (Y_t)$ s nerozložitelnou a aperiodickou maticí P existuje takový coupling, že čas, kdy se řetězce poprvé setkají je konečný skoro jistě. To jest označíme-li $\tau := \min\{t : X_t = Y_t\}$, platí $\mathbf{P}\{\tau < \infty\} = 1$.

Důkaz. Nechtě (X_t) a (Y_t) jsou Markovovy řetězce s maticí P a množinou stavů Ω . Vzhledem k nerozložitelnosti a aperiodicitě matice P zaručuje Věta 1 existenci přirozeného čísla r splňujícího

$$P^r(u, v) > 0 \quad \text{pro všechny } u, v \in \Omega. \quad (3.6)$$

Zvolme nyní libovolně nějaký stav $z \in \Omega$ a položme

$$p := \min_{w \in \Omega} P^r(w, z).$$

Díky vlastnosti (3.6) platí $p \in (0, 1)$. Zvolme dále libovolně nějaké dva stavy $x, y \in \Omega$ a definujme coupling řetězců (X_t, Y_t) tak, že $X_0 = x, Y_0 = y$ a oba řetězce budou vzájemně nezávislé. Zřejmě pak platí

$$\mathbf{P}_{x,y}\{X_r = z\} \geq p \quad \text{a} \quad \mathbf{P}_{x,y}\{Y_r = z\} \geq p. \quad (3.7)$$

Nyní indukci ukážeme, že pro všechna $k \in \mathbb{N}$ platí vztah

$$\mathbf{P}_{x,y} \left\{ \bigcap_{j=1}^k \{X_{jr} = z = Y_{jr}\}^c \right\} \leq (1 - p^2)^k. \quad (3.8)$$

Pro $k = 1$ můžeme díky nezávislosti řetězců $(X_t), (Y_t)$ a s využitím odhadů (3.7) psát

$$\begin{aligned} \mathbf{P}_{x,y} \left\{ \{X_r = z = Y_r\}^c \right\} &= 1 - \mathbf{P}_{x,y}\{X_r = z = Y_r\} \\ &= 1 - \mathbf{P}_{x,y}\{X_r = z\} \mathbf{P}_{x,y}\{Y_r = z\} \\ &\leq 1 - p^2. \end{aligned} \quad (3.9)$$

Nyní předpokládejme, že vztah (3.8) platí pro $k \in \mathbb{N}$. Máme

$$\begin{aligned} &\mathbf{P}_{x,y} \left\{ \bigcap_{j=1}^{k+1} \{X_{jr} = z = Y_{jr}\}^c \right\} \\ &= \mathbf{P}_{x,y} \left\{ \bigcap_{j=1}^k \{X_{jr} = z = Y_{jr}\}^c, \{X_{(j+1)r} = z = Y_{(j+1)r}\}^c \right\} \\ &= \mathbf{P}_{x,y} \left\{ \{X_{(j+1)r} = z = Y_{(j+1)r}\}^c \mid \bigcap_{j=1}^k \{X_{jr} = z = Y_{jr}\}^c \right\} \\ &\quad \cdot \mathbf{P}_{x,y} \left\{ \bigcap_{j=1}^k \{X_{jr} = z = Y_{jr}\}^c \right\} \\ &\leq (1 - p^2)(1 - p^2)^k, \end{aligned}$$

kde v poslední nerovnosti jsme užili indukční předpoklad, homogenitu řetězců a fakt, že (3.9) platí pro libovolnou dvojici počátečních stavů $x, y \in \Omega$.

Dále vzhledem k tomu, že pro libovolnou dvojici počátečních stavů a pro každé $k \in \mathbb{N}$ platí inkluze náhodných jevů

$$\{\tau > kr\} \subset \left\{ \bigcap_{j=1}^k \{X_{jr} = z = Y_{jr}\}^c \right\},$$

můžeme s využitím (3.8) psát

$$\mathbf{P}_{x,y} \{\tau > kr\} \leq \mathbf{P}_{x,y} \left\{ \bigcap_{j=1}^k \{X_{jr} = z = Y_{jr}\}^c \right\} \leq (1 - p^2)^k. \quad (3.10)$$

Odtud potom limitním přechodem pro $k \rightarrow \infty$ dostáváme $\mathbf{P}_{x,y} \{\tau = \infty\} = 0$, tudíž musí být $\mathbf{P}_{x,y} \{\tau < \infty\} = 1$.

□

4. Odhady rychlosti konvergence

V této kapitole ukážeme na několika klasických modelech Markovových řetězců odhad rychlosti konvergence jejich marginálního rozdělení ke stacionárnímu rozdělení metodou couplingu řetězců, která byla vyložena v předchozí kapitole. Níže uvedené příklady jsou spolu s dalšími ukázkami k nalezení v páté kapitole knihy [1].

4.1 Náhodná procházka po cyklu

Definice. Položme $\Omega = \mathbb{Z}_n = \{0, 1, \dots, n-1\}$ a uvažujme následující matici pravděpodobností přechodu:

$$P(j, k) = \begin{cases} 1/2 & \text{pokud } k \equiv j+1 \pmod{n}, \\ 1/2 & \text{pokud } k \equiv j-1 \pmod{n}, \\ 0 & \text{jinak.} \end{cases} \quad (4.1)$$

Markovův řetězec (X_t) s maticí P a množinou stavů Ω nazýváme *náhodná procházka po n -cyklu*.

Dříve než začneme zkoumat rychlost konvergence náhodné procházky po n -cyklu, podívejme se, co se stane, pokud bude n sudé. Stavový prostor Ω lze v tomto případě rozdělit na dvě disjunktní podmnožiny — stavy *sudé* a stavy *liché* — takovým způsobem, že řetězec může přecházet s kladnou pravděpodobností jen mezi stavy z různých podmnožin, tj. graf incidence této procházky je bipartitní. Takový řetězec je ale zřejmě periodický s periodou dvě, tudíž nejsou splněny předpoklady *Věty o konvergenci*. Skutečně pro každé $t \in \mathbb{N}$ a nějaký lichý počáteční stav $x_0 \in \Omega$ je rozdělení řetězce $P^{2t}(x_0, \cdot)$ nulové na všech sudých stavech, zatímco rozdělení $P^{2t+1}(x_0, \cdot)$ je nulové na všech stavech lichých. Z toho je jasné vidět, že rozdělení $P^t(x_0, \cdot)$ nekonverguje pro $t \rightarrow \infty$.

Abychom se vyhnuli podobným problémům budeme v tomto a i některých dalších případech zkoumat následovně upravenou verzi řetězce.

Definice. Nechť P je libovolná matice pravděpodobností přechodu. Položme

$$Q := \frac{I + P}{2}, \quad (4.2)$$

kde I je jednotková matice o rozměrech $|\Omega| \times |\Omega|$. Markovův řetězec s maticí Q pak nazýváme *línou verzí* řetězce s maticí P .

Je vidět, že líná verze řetězce musí vždy nutně být aperiodická, jelikož platí $Q(x, x) > 0$ pro všechna $x \in \Omega$.

Matice pravděpodobností přechodu líné verze procházky po n -cyklu tedy vypadá následovně:

$$P(j, k) = \begin{cases} 1/4 & \text{pokud } k \equiv j+1 \pmod{n}, \\ 1/4 & \text{pokud } k \equiv j-1 \pmod{n}, \\ 1/2 & \text{pokud } k \equiv j, \\ 0 & \text{jinak.} \end{cases} \quad (4.3)$$

Nyní konstrukcí vhodného couplingu shora odhadneme mixing time tohoto řetězce. Necht' (X_t, Y_t) jsou dvě líné náhodné procházky po \mathbb{Z}_n s počátečními stavy x a y . Zavedme pomocný náhodný proces (W_t) , kde pro každé $t \in \mathbb{N}$ bude W_t náhodná veličina s alternativním rozdělením s parametrem $1/2$, nezávislá na všech předchozích hodnotách (W_t) . Náš coupling se bude v každém kroku řídit hodnotou procesu (W_t) . Ta bude určovat, který z řetězců změní stav, a který setrvá ve své pozici. Je-li $W_t = 1$, řetězec (X_t) se posune o krok dopředu tak, že nový stav X_{t+1} bude vybrán z rozdělení z matice *klasické verze* procházky po n -cyklu (4.1). V tomto případě položíme $Y_{t+1} = Y_t$. Je-li naopak $W_t = 0$, posune se podle matice (4.1) o krok dopředu řetězec (Y_t) , zatímco (X_t) setrvá ve své pozici. Jakmile však dojde ke koalescenci obou řetězců, jejich trajektorie se navždy spojí a budou se dál vyvíjet v čase jako jedna *líná verze* náhodné procházky po n -cyklu.

Uvědomme si, že náhodný proces (X_t, Y_t) , který jsme právě popsali je skutečně coupling dvou líných verzí náhodné procházky po n -cyklu dle definice v sekci 3.1. Dále si všimněme, že (X_t, Y_t) má vlastnost (3.1).

Necht' D_t je vzdálenost mezi stavy řetězců (X_t) a (Y_t) v čase t měřená po směru hodinových ručiček. Všimněme si, že (D_t) je jednoduchá náhodná procházka na množině $\{0, 1, \dots, n\}$ s absorpčními bariérami dle definice v (1.3). Označíme-li $\tau = \min\{t \geq 0 : D_t \in \{0, n\}\}$, máme podle *Věty 5*

$$\mathbf{E}_{x,y}(\tau) = k(n - k),$$

kde k je vzdálenost mezi x a y měřená podle směru hodinových ručiček. Jelikož pro náš coupling platí $\tau = \tau_{couple}$, dostáváme postupným užitím *Důsledku 18* a Markovovy nerovnosti

$$d(t) \leq \max_{x,y \in \mathbb{Z}_n} \mathbf{P}_{x,y}\{\tau > t\} \leq \frac{\max_{x,y} \mathbf{E}_{x,y}(\tau)}{t}. \quad (4.4)$$

Zřejmě je

$$\max_{k \in \{0, \dots, n\}} k(n - k) = \frac{n^2}{4}$$

a proto z (4.4) dostáváme

$$d(t) \leq \frac{n^2}{4t}. \quad (4.5)$$

Pravá strana předchozí nerovnosti je rovna $1/4$ pro $t = n^2$, tudíž máme $t_{mix} \leq n^2$.

4.2 Náhodná procházka po toru

Definice. d -rozměrný torus je graf, jehož množina vrcholů je tvořena kartézským součinem

$$\mathbb{Z}_n^d = \underbrace{\mathbb{Z}_n \times \dots \times \mathbb{Z}_n}_{d\text{-krát}}.$$

Vrcholy $\mathbf{x} = (x^1, \dots, x^d)$ a $\mathbf{y} = (y^1, \dots, y^d)$ jsou v \mathbb{Z}_n^d spojeny hranou, jestliže pro nějaké $j \in \{1, \dots, d\}$ platí $x^i = y^i$ pro všechna $i \neq j$ a $x^j \equiv (y^j \pm 1) \pmod{n}$.

Všimněme si, že pro n sudé je graf incidence náhodné procházky po \mathbb{Z}_n^d bipartitní a tato procházka je tedy opět periodická. Abychom se vyhnuli této komplikaci, budeme uvažovat línou verzi náhodné procházky po \mathbb{Z}_n^d , kterou jsme definovali vztahem (4.2).

Ještě než přikročíme k odhadu rychlosti konvergence náhodné procházky po toru, uvedeme si jako lemma tzv. Waldovu rovnost, kterou využijeme v důkazu našeho odhadu.

Lemma 20. (Tvrzení 5.18 v [3]) *Uvažujme náhodný součet $S = \sum_{n=1}^N X_n$, kde $X_n, n \in \mathbb{N}$ jsou nezávislé reálné náhodné veličiny se stejnou střední hodnotou a náhodná veličina $N : \Omega \rightarrow \mathbb{N}_0$ je s nimi nezávislá. Potom je-li $\mathbf{E}|X_1| < \infty$ a $\mathbf{E}|N| < \infty$, pak je $\mathbf{E}|S| < \infty$ a platí $\mathbf{E}S = \mathbf{E}N\mathbf{E}X_1$.*

Nyní můžeme pomocí vhodného couplingu odhadnout mixing time líné náhodné procházky po \mathbb{Z}_n^d .

Věta 21. *Pro línou verzi náhodné procházky po d -dimenzionálním toru \mathbb{Z}_n^d platí*

$$t_{mix}(\varepsilon) \leq c(d)n^2 \log_2(\varepsilon^{-1}), \quad (4.6)$$

kde $c(d)$ je konstanta závislá na dimenzi d .

Důkaz. Abychom mohli pomocí *Důsledku 18* dokázat tuto větu, zkonstruujeme coupling řetězců pro každou dvojici počátečních stavů (\mathbf{x}, \mathbf{y}) a shora odhadneme střední hodnotu času koalescence $\tau_{couple} = \tau_{\mathbf{x}, \mathbf{y}}$.

Nechť tedy (\mathbf{X}_t) a (\mathbf{Y}_t) jsou líné verze náhodné procházky po \mathbb{Z}_n^d s počátečními stavy \mathbf{x} a \mathbf{y} . Nyní slovně popíšeme náš coupling $(\mathbf{X}_t, \mathbf{Y}_t)$. Nejprve rovnoměrně náhodně vybereme jednu z d souřadnic. Jestliže se pozice obou procházek ve vybrané souřadnici shodují, posuneme stavy obou řetězců v těchto souřadnicích o stejnou hodnotu $+1, -1$ nebo 0 , a to s pravděpodobnostmi $1/4, 1/4$ a $1/2$. Jestliže se naopak hodnoty obou procházek v dané souřadnici liší, náhodně vybereme jeden z řetězců, jehož pozici ve vybrané souřadnici posuneme o $+1$ nebo -1 , kde každá z obou možností nastane s pravděpodobností $1/2$. Pozici druhého řetězce v tomto případě nezměníme.

Všimněme si, že jakmile se jednou pozice obou řetězců v nějaké souřadnici sjednotí, pak tyto už nikdy nebudou různé. Nechť $\mathbf{X}_t = (X_t^1, \dots, X_t^d)$ a $\mathbf{Y}_t = (Y_t^1, \dots, Y_t^d)$ a označme

$$\tau_i := \min\{t \geq 0 : X_t^i = Y_t^i\}$$

čas, ve kterém se sjednotí pozice obou řetězců v souřadnici i .

Uvědomme si dále, že rozdíl mezi hodnotami X_t^i a Y_t^i měřený po směru hodinových ručiček a vnímaný pouze v časech, kdy je vybrána souřadnice i , se chová stejně jako coupling náhodných procházek po n -cyklu, který jsme popsali v předchozí sekci. Střední hodnota počtu posunů v souřadnici i , který je potřeba vykonat, aby se pozice obou řetězců v dané souřadnici shodovaly, je proto menší nebo rovna $n^2/4$.

Je vidět, že doba čekání než bude v našem couplingu vybrána i -tá souřadnice má geometrické rozdělení s parametrem $1/d$. Její střední hodnota je proto rovna d a užitím *Lemmatu 20* dostáváme

$$\mathbf{E}_{\mathbf{x}, \mathbf{y}}(\tau_i) \leq \frac{dn^2}{4}. \quad (4.7)$$

Jelikož pro náš coupling zřejmě platí $\tau_{couple} = \max_{1 \leq i \leq d} \tau_i$, můžeme psát

$$\mathbf{E}_{\mathbf{x}, \mathbf{y}}(\tau_{couple}) = \mathbf{E}_{\mathbf{x}, \mathbf{y}}\left(\max_{1 \leq i \leq d} \tau_i\right) \leq \mathbf{E}_{\mathbf{x}, \mathbf{y}}\left(\sum_{i=1}^d \tau_i\right) = d\mathbf{E}_{\mathbf{x}, \mathbf{y}}(\tau_i) \leq \frac{d^2 n^2}{4}. \quad (4.8)$$

Odhad (4.8) je nezávislý na dvojici počátečních stavů a užitím Markovovy nerovnosti dostáváme

$$\mathbf{P}_{\mathbf{x}, \mathbf{y}}\{\tau_{couple} > t\} \leq \frac{\mathbf{E}_{\mathbf{x}, \mathbf{y}}(\tau_{couple})}{t} \leq \frac{d^2 n^2}{4t}. \quad (4.9)$$

Pravá strana (4.9) je rovna $1/4$ pro $t_0 = d^2 n^2$. Užitím *Důsledku 18* tedy získáváme $t_{mix} \leq d^2 n^2$. Odtud potom díky (2.44) máme

$$t_{mix}(\varepsilon) \leq d^2 n^2 \lceil \log_2 \varepsilon^{-1} \rceil$$

a důkaz věty je tak dokončen. \square

S využitím nerovnosti (4.7) můžeme odhad mixing time líné verze náhodné procházky po \mathbb{Z}_n^d ještě vylepšit. Zkusme podrobněji zanalyzovat čas koalescence couplingu dvou líných náhodných procházek po d -dimenzionálním toru $(\mathbf{X}_t), (\mathbf{Y}_t)$ s nějakými počátečními stavy $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_n^d$, který jsme vytvořili v důkazu minulé věty.

Pro $i = 1, \dots, d$ opět označme $\tau_i := \min\{t \geq 0 : X_t^i = Y_t^i\}$ čas, ve kterém dojde ke sjednocení obou řetězců v i -té souřadnici. Zkoumejme nyní pro pevné $i \in \{1, \dots, d\}$ veličinu τ_i . Předpokládejme, že platí

$$t \geq kdn^2, \quad (4.10)$$

kde k je nějaké přirozené číslo. Matematickou indukcí ukážeme, že pro všechna $k \in \mathbb{N}$ platí vztah

$$\mathbf{P}_{\mathbf{x}, \mathbf{y}}\{\tau_i > kdn^2\} \leq \left(\frac{1}{4}\right)^k. \quad (4.11)$$

Postupným užitím Markovovy nerovnosti a zmiňovaného odhadu (4.7) z důkazu předchozí věty dostáváme

$$\mathbf{P}_{\mathbf{x}, \mathbf{y}}\{\tau_i > kdn^2\} \leq \frac{\mathbf{E}_{\mathbf{x}, \mathbf{y}}(\tau_i)}{dn^2} \leq \frac{1}{4}.$$

Předpokládejme nyní, že vztah (4.11) platí pro nějaké $k \in \mathbb{N}$. Díky zřejmé inkluzi jevů $\{\tau_i > (k+1)dn^2\} \subset \{\tau_i > kdn^2\}$ můžeme psát

$$\begin{aligned} \mathbf{P}_{\mathbf{x}, \mathbf{y}}\{\tau_i > (k+1)dn^2\} &= \mathbf{P}_{\mathbf{x}, \mathbf{y}}\{\tau_i > (k+1)dn^2 \mid \tau_i > kdn^2\} \mathbf{P}_{\mathbf{x}, \mathbf{y}}\{\tau_i > kdn^2\} \\ &\leq \frac{1}{4} \left(\frac{1}{4}\right)^k. \end{aligned}$$

Protože vzhledem k podmínce (4.10) platí $\{\tau_i > t\} \subset \{\tau_i > kdn^2\}$, dostáváme ze vztahu (4.11) nerovnost

$$\mathbf{P}_{\mathbf{x}, \mathbf{y}}\{\tau_i > t\} \leq \left(\frac{1}{4}\right)^k. \quad (4.12)$$

Dále vzhledem k rovnosti jevů

$$\{\tau_{couple} > t\} = \left\{ \bigcup_{i=1}^d \{\tau_i > t\} \right\},$$

která platí nezávisle na dvojici počátečních stavů, můžeme díky odhadu (4.12) psát

$$\mathbf{P}_{\mathbf{x},\mathbf{y}}\{\tau_{couple} > t\} = \mathbf{P}_{\mathbf{x},\mathbf{y}}\left\{\bigcup_{i=1}^d\{\tau_i > t\}\right\} \leq \sum_{i=1}^d \mathbf{P}_{\mathbf{x},\mathbf{y}}\{\tau_i > t\} \leq \left(\frac{1}{4}\right)^k d. \quad (4.13)$$

Podle *Důsledku 18* je tedy $d(t) \leq (1/4)^k d$. Pravá strana této nerovnosti nabývá hodnoty $(1/4)^k$ pro $k = (1/2) \log_2(4d)$. Vzhledem k definici (2.41) a faktu, že jsme po celou dobu předpokládali (4.10), je potom

$$t_{mix} \leq \frac{1}{2}(\log_2 4d)dn^2.$$

Hlubší analýzou času koalescence jsme tedy pro mixing time líné verze náhodné procházky po \mathbb{Z}_n^d získali odhad řádu $\mathcal{O}((d \log d)n^2)$, což pro vysoké dimenze představuje znatelné zlepšení oproti původnímu $\mathcal{O}(d^2n^2)$ z *Věty 21*.

4.3 Vítězná série

Na závěr kapitoly pro zajímavost ukážeme odhad rychlosti konvergence jednoho klasického Markovova řetězce. Představme si tvora s omezenou pamětí, který opakovaně hází mincí a snaží se zapamatovat si délku poslední (nepřerušené) série po sobě padlých líců. Pokud tedy při posledním hození padl rub, tvor si pamatuje číslo 0. Pokud padlo více jak n líců po sobě, tvor si pamatuje jen posledních n z nich, protože jeho paměť je omezená.

Ekvivalentně, uvažujme okénko délky n , posouvající se směrem doprava po nekonečné posloupnosti náhodných rovnoměrně rozdělených nezávislých bitů a nechť X_t značí délku podposloupnosti samých jedniček, začínající na pravém kraji okénka.

Definice. *Vítěznou sérii* definujeme jako Markovův řetězec s množinou stavů $\{0, \dots, n\}$ a následujícími nenulovými pravděpodobnostmi přechodu

$$\begin{aligned} P(i, 0) &= 1/2 \quad \text{pro } 0 \leq i \leq n, \\ P(i, i+1) &= 1/2 \quad \text{pro } 0 \leq i < n, \\ P(n, n) &= 1/2. \end{aligned} \quad (4.14)$$

Spočtením soustavy rovnic (1.1) snadno ověříme, že stacionární rozdělení tohoto řetězce je

$$\pi(i) = \begin{cases} 1/2^{i+1} & \text{pro } i = 0, 1, \dots, n-1, \\ 1/2^n & \text{pro } i = n. \end{cases} \quad (4.15)$$

Pokusme se nyní pomocí naší metody vyšetřit, jak rychle vítězná série ke svému stacionárnímu rozdělení konverguje.

Zvolme libovolně počáteční hodnoty $a, b \in \{0, \dots, n\}$ a nechť x a y jsou nějaké posloupnosti bitů délky n , které končí úsekem jedniček délky a , resp. b . Uvažujme dvě vítězné série v jejichž okénkách jsou na počátku posloupnosti x a y . Coupling obou řetězců vytvoříme tak, že za úseky x i y připojíme *stejnou* posloupnost náhodných rovnoměrně rozdělených nezávislých bitů. Oba řetězce se pak budou vyvíjet dle této posloupnosti, po které se bude posouvat okénko délky n a stav

každého řetězce v daném čase bude určen délkou posloupnosti po sobě jdoucích jedniček začínající na pravém kraji jeho okénka.

Všimněme si, že jakmile obě okénka dojdou k první nule v napojené posloupnosti, oba řetězce přejdou do stavu 0 a od tohoto momentu se budou vyvíjet identicky. Vzhledem k rovnoměrnému rozdělení a nezávislosti bitů v posloupnosti proto platí

$$\mathbf{P}\{\tau_{couple} > t\} \leq 2^{-t},$$

a to pro všechny dvojice počátečních stavů $a, b \in \{0, \dots, n\}$. Užitím *Důsledku 18* tedy dostáváme

$$d(t) \leq 2^{-t}. \tag{4.16}$$

Pravá strana (4.16) je rovna ε pro $t = \log_2(1/\varepsilon)$ a vzhledem k definici (2.40) tedy máme

$$t_{mix}(\varepsilon) \leq \left\lceil \log_2 \left(\frac{1}{\varepsilon} \right) \right\rceil,$$

což je odhad nezávislý na n . Speciálně pro všechna $n \in \mathbb{N}$ dostáváme $t_{mix} \leq 2$.

5. Grand coupling a rychlost konvergence Metropolisova řetězce

Závěrečnou kapitolu věnujeme konstrukci odhadu mixing time Metropolisova řetězce na množině přípustných obarvení grafu. Metropolisův řetězec je užitečný příklad tzv. Markov Chain Monte Carlo algoritmů, které jsme představili v úvodu práce. Velmi podobná konstrukce odhadu pro jiný MCMC algoritmus je (společně s touto) k nalezení v páté kapitole knihy [1].

5.1 Grand coupling

Uvažujme situaci, kdy chceme zkoumat rychlost konvergence nějakého Markovova řetězce se stavovým prostorem Ω . Někdy může být pro tento účel užitečné zkonstruovat právě $|\Omega|$ kopií tohoto řetězce, z nichž každá má jiný počáteční stav $x \in \Omega$. V případě, kdy $|\Omega| > 2$ si již nevystačíme s klasickým couplingem Markovových řetězců definovaným v sekci 3.1. V tuto chvíli využijeme nástroj zvaný grand coupling.

Definice. Nechť (X_t) je Markovův řetězec s maticí P a stavovým prostorem Ω . Kolekci náhodných veličin $\{X_t^x : x \in \Omega, t = 0, 1, 2, \dots\}$ splňující, že pro každé $x \in \Omega$ je náhodný proces $(X_t^x)_{t=0}^\infty$ Markovův řetězec s maticí P a počátečním stavem $X_0^x = x$, budeme nazývat *grand coupling*.

K pohodlné konstrukci grand couplingu lze využít reprezentaci řetězce (X_t) pomocí náhodného přiřazení, která byla definována v (1.4). Mějme tedy dvojici (f, Z) , kde $f : \Omega \times \Lambda \rightarrow \Omega$ je přechodová funkce a Z je náhodná veličina s hodnotami v Λ , která splňuje pro všechny dvojice stavů $x, y \in \Omega$ rovnost $P(x, y) = \mathbf{P}\{f(x, Z) = y\}$. Existenci takové dvojice (f, Z) zaručuje *Věta 4*. Dále nechť Z_1, Z_2, \dots je posloupnost nezávislých stejně rozdělených náhodných veličin, jejichž rozdělení je stejné jako rozdělení veličiny Z a položíme

$$X_0^x = x, \quad X_t^x = f(X_{t-1}^x, Z_t) \text{ pro } t \geq 1. \quad (5.1)$$

Jelikož zřejmě každý z procesů $(X_t^x)_{t=0}^\infty$ je Markovův řetězec s maticí P a počátečním stavem x , právě zkonstruovaná kolekce náhodných veličin je skutečně grand coupling.

5.2 Metropolisův řetězec

Uvažujme situaci, kdy máme nějaký Markovův řetězec s maticí Ψ , stavovým prostorem Ω a *nějakým* stacionárním rozdělením. Zvolme libovolně nějaké pravděpodobnostní rozdělení π na Ω a položme si otázku, zda existuje způsob, jak modifikovat matici pravděpodobností přechodu Ψ tak, aby nový řetězec měl stacionární rozdělení π . Obecný návod na řešení tohoto problému dává Metropolisův-Hastingsův algoritmus, popsáný např. v sekci 3.2.2 knihy [1].

Pro naši aplikaci v následující části však postačí zjednodušená verze tohoto algoritmu nazvaná Metropolisův řetězec, která se omezuje na případy, kdy je matice Ψ symetrická, tj. pro všechna $x, y \in \Omega$ platí $\Psi(x, y) = \Psi(y, x)$. Nyní ukážeme,

jak lze upravit takovou symetrickou matici Ψ , abychom získali řetězec s libovolně zvoleným stacionárním rozdělením π , čímž Metropolisův řetězec odvodíme.

Budeme chtít, aby se nový řetězec choval následovně: pokud se řetězec nachází ve stavu x , bude podle rozdělení $\Psi(x, \cdot)$ vybrán kandidát na následující stav. Takový kandidát y bude *přiját* s pravděpodobností $a(x, y)$ nebo bude naopak *zamítnut*, a to s doplňkovou pravděpodobností $1 - a(x, y)$. V případě přijetí kandidáta y se řetězec posune do stavu y , je-li naopak kandidát zamítnut, řetězec zůstane ve stavu x . Klíčovým krokem pro odvození fungujícího algoritmu je samozřejmě správná volba pravděpodobností přijetí $a(x, y)$ pro každou dvojici stavů $x, y \in \Omega$. Všimněme si, že matice pravděpodobností přechodu P nového řetězce je

$$P(x, y) = \begin{cases} \Psi(x, y)a(x, y) & \text{pro } y \neq x, \\ 1 - \sum_{z: z \neq x} \Psi(x, z)a(x, z) & \text{pro } y = x. \end{cases} \quad (5.2)$$

Z *Věty 3* plyne, že π je stacionárním rozdělením řetězce s maticí P , jestliže pro všechny dvojice různých stavů $x \neq y$ platí

$$\pi(x)\Psi(x, y)a(x, y) = \pi(y)\Psi(y, x)a(y, x). \quad (5.3)$$

Vzhledem k tomu, že předpokládáme symetrickou matici Ψ , rovnice (5.3) platí právě tehdy, když pro všechny dvojice různých stavů $x \neq y$ platí

$$b(x, y) = b(y, x), \quad (5.4)$$

kde $b(x, y) = \pi(x)a(x, y)$.

Dále vzhledem k faktu, že pravděpodobnost přijetí musí splňovat $a(x, y) \leq 1$, je funkce $b(x, y)$ za platnosti (5.4) shora ohraničena následovně:

$$b(x, y) \leq \pi(x), \quad (5.5)$$

$$b(x, y) = b(y, x) \leq \pi(y). \quad (5.6)$$

Protože zamítání kandidátů při výběru dalšího stavu zpomaluje celý řetězec, čímž se snižuje efektivita algoritmu, chceme vybrat největší možné $b(x, y)$ splňující podmínky (5.5) a (5.6). Největším takovým řešením je zřejmě $b(x, y) = \min\{\pi(x), \pi(y)\}$. Po dosazení do definice $b(x, y)$ pak získáváme

$$a(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\}. \quad (5.7)$$

Definice. *Metropolisův řetězec* odpovídající symetrické matici pravděpodobností přechodu Ψ a stacionárnímu rozdělení π definujeme jako Markovův řetězec s maticí P , kde

$$P(x, y) = \begin{cases} \Psi(x, y) \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\} & \text{pro } y \neq x, \\ 1 - \sum_{z: z \neq x} \Psi(x, z) \min \left\{ 1, \frac{\pi(z)}{\pi(x)} \right\} & \text{pro } y = x. \end{cases} \quad (5.8)$$

Poznámka. Výše uvedenou diskuzí jsme ukázali, že π je skutečně stacionárním rozdělením Metropolisova řetězce. To však samo o sobě nezajišťuje, že marginální rozdělení Metropolisova řetězce k rozdělení π konverguje. To, jestli je π pro Metropolisův řetězec i rozdělením limitním, závisí na vlastnostech dané matice Ψ . Tuto skutečnost je proto nutné v konkrétních případech zvláště ověřit, a to například tím, že ukážeme, že jsou splněny předpoklady *Věty 15*.

5.3 Problém přípustného obarvení grafu

Definice. Necht' $\{1, 2, \dots, q\}$ je množina barev a necht' $G = (V, E)$ je nějaký graf s množinou vrcholů V a množinou hran E . *Přípustným q -obarvením* grafu G rozumíme takové přiřazení barev vrcholům V , že žádné sousední vrcholy nemají stejnou barvu.

Uvažujme následující Markovův řetězec, jehož stavovým prostorem je množina *všech* (ne nutně přípustných) q -obarvení nějakého grafu $G = (V, E)$. Nejprve je náhodně zvolen nějaký vrchol $v \in V$, každý se stejnou pravděpodobností $1/|V|$. Dále je náhodně zvolena *jakákoliv* barva z množiny $\{1, 2, \dots, q\}$, každá s pravděpodobností $1/q$. Vrchol v je poté obarven zvolenou barvou.

Na tento řetězec nyní aplikujeme Metropolisův algoritmus, kde za π vezmeme rovnoměrné rozdělení na množině všech *přípustných* q -obarvení grafu G . Je-li aktuálním stavem řetězce nějaké přípustné q -obarvení a barva k je navrhována k obarvení vybraného vrcholu, Metropolisův algoritmus přijme toto navrhované přebarvení s pravděpodobností 1 právě tehdy, je-li výsledné obarvení *přípustné*, jinak jej zamítne a novým stavem řetězce zůstane původní přípustné obarvení.

5.4 Odhad rychlosti konvergence

Nyní máme k dispozici všechny potřebné prostředky, abychom mohli odhadnout rychlost konvergence marginálního rozdělení Metropolisova řetězce pro přípustná q -obarvení grafu k jeho stacionárnímu rozdělení, kterým je rovnoměrné rozdělení na množině přípustných q -obarvení. K důkazu našeho odhadu uijeme grand coupling, který jsme definovali v prvním odstavci této kapitoly.

Věta 22. *Necht' $G = (V, E)$ je graf s n vrcholy a maximálním stupněm vrcholu Δ . Necht' $q > 3\Delta$. Označíme-li $c_{met}(\Delta, q) := 1 - (3\Delta/q)$, pak pro Metropolisův řetězec na množině přípustných q -obarvení grafu G , který byl popsán v sekci 5.3, platí*

$$t_{mix}(\varepsilon) \leq c_{met}(\Delta, q)^{-1} n [\log n + \log(1/\varepsilon)]. \quad (5.9)$$

Důkaz. Nejprve označme Ω množinu *přípustných* q -obarvení grafu G a $\tilde{\Omega}$ množinu *všech* q -obarvení grafu G .

Definujme následující grand coupling Metropolisových řetězců (X_t^x) se stavovým prostorem $\tilde{\Omega}$ a stacionárním rozdělením π rovnoměrným na Ω . V každém kroku bude rovnoměrně náhodně a nezávisle na minulosti vygenerována dvojice vrcholu a barvy (v, k) z kartézského součinu $V \times \{1, \dots, q\}$. Pro každé $x \in \tilde{\Omega}$ dojde při kroku z X_t^x na X_{t+1}^x k pokusu o přebarvení vrcholu v na barvu k . Tento krok bude přijat tehdy a jen tehdy, když barva k není mezi barvami vrcholů sousedících s vrcholem v . Pokud není zvolený krok přijat, řetězec (X_t^x) zůstane ve stavu X_t^x . Důležité je, že v našem grand couplingu bude pro každý řetězec (X_t^x) v daném čase navrhována stejná dvojice (v, k) .

Pro každá q -obarvení $x, y \in \tilde{\Omega}$ definujme

$$\rho(x, y) := \sum_{v \in V} \mathbf{1}_{\{x(v) \neq y(v)\}}$$

jako počet vrcholů, které mají v x a y rozdílnou barvu. Všimněme si, že ρ je metrika na prostoru $\tilde{\Omega}$.

Předpokládejme nejprve, že platí $\rho(x, y) = 1$, tzn. že obarvení x a y se liší pouze v nějakém vrcholu v_0 . Označme \mathcal{N} množinu barev, kterými jsou obarveni sousedé vrcholu v_0 v obarvení x . Uvědomme si, že tato množina je stejná jako množina barev, kterými jsou obarveni sousedé vrcholu v_0 v obarvení y . Podívejme se nyní, co se může stát se vzdáleností mezi obarveními x a y po posunu našeho grand couplingu o krok dopředu, tj. zkoumejme $\rho(X_1^x, X_1^y)$. Vzhledem k definici našeho grandcouplingu a předpokladu $\rho(x, y) = 1$ je zřejmé, že $\rho(X_1^x, X_1^y)$ může nabývat pouze hodnot 0, 1 a 2.

Vzdálenost $\rho(X_1^x, X_1^y)$ bude nulová tehdy a jen tehdy, pokud byl k přebarvení vygenerován vrchol v_0 a zároveň nějaká barva, která neleží v množině \mathcal{N} . Vzhledem k tomu, že vybraný vrchol a vybraná barva jsou vzájemně nezávislé veličiny odpovídající rovnoměrnému rozdělení na množinách V a $\{1, 2, \dots, q\}$, platí odhad

$$\mathbf{P}\{\rho(X_1^x, X_1^y) = 0\} = \binom{1}{n} \binom{q - |\mathcal{N}|}{q} \geq \frac{q - \Delta}{nq}. \quad (5.10)$$

Dále si všimněme, že pro každý vrchol w sousedící s vrcholem v_0 je množina barev vrcholů sousedících s w a různých od v_0 stejná v obarvení x a y . Předpokládejme, že tato množina neobsahuje barvy $x(v_0)$ a $y(v_0)$. Je-li k přechodu řetězce vybrána dvojice $(w, x(v_0))$, pak v tomto případě v obarvení y dojde k přebarvení vrcholu w na barvu $x(v_0)$, zatímco v obarvení x bude takový krok zamítnut a toto tak zůstane v původním stavu. Vzdálenost mezi obarveními x a y se tak v důsledku zvětší na dvě. Obdobně bude $\rho(X_1^x, X_1^y) = 2$, pokud bude k přebarvení vybrána dvojice $(w, y(v_0))$. Toto jsou jediné případy, kdy může za předpokladu $\rho(x, y) = 1$ pro náš grandcoupling nastat situace $\rho(X_1^x, X_1^y) = 2$. Pravděpodobnost tohoto jevu můžeme odhadnout shora

$$\mathbf{P}\{\rho(X_1^x, X_1^y) = 2\} \leq \binom{\Delta}{n} \binom{2}{q}. \quad (5.11)$$

S využitím odhadů (5.10) a (5.11) můžeme psát

$$\mathbf{E}(\rho(X_1^x, X_1^y) - 1) \leq \frac{2\Delta}{nq} - \frac{(q - \Delta)}{nq} = \frac{3\Delta - q}{nq}.$$

Vzhledem k linearitě střední hodnoty je potom

$$\mathbf{E}(\rho(X_1^x, X_1^y)) \leq 1 - \frac{3\Delta - q}{nq}.$$

Protože platí $q > 3\Delta$, je $c_{met}(\Delta, q) = 1 - (3\Delta/q) > 0$ a z předchozí nerovnosti máme

$$\mathbf{E}(\rho(X_1^x, X_1^y)) \leq 1 - \frac{c_{met}(\Delta, q)}{n} < 1. \quad (5.12)$$

Nyní předpokládejme, že x a y jsou obarvení, pro která platí $\rho(x, y) = r$. Zřejmě existuje konečná posloupnost obarvení $x = x_0, x_1, \dots, x_r = y$ takových, že pro $k = 1, 2, \dots, r$ platí $\rho(x_k, x_{k-1}) = 1$. Odhad (5.12) tedy můžeme použít pro každé $\mathbf{E}(\rho(X_1^{x_k}, X_1^{x_{k-1}}))$, což společně s trojúhelníkovou nerovností pro metriku ρ dává

$$\begin{aligned} \mathbf{E}(\rho(X_1^x, X_1^y)) &\leq \sum_{k=1}^r \mathbf{E}(\rho(X_1^{x_k}, X_1^{x_{k-1}})) \\ &\leq r \left(1 - \frac{c_{met}(\Delta, q)}{n} \right) = \rho(x, y) \left(1 - \frac{c_{met}(\Delta, q)}{n} \right). \end{aligned} \quad (5.13)$$

Dále vzhledem k tomu, že rozdělení náhodného vektoru (X_t^x, X_t^y) podmíněné jevem $\{X_{t-1}^x = x_{t-1}, X_{t-1}^y = y_{t-1}\}$ je stejné jako rozdělení vektoru $(X_1^{x_{t-1}}, X_1^{y_{t-1}})$, můžeme podle předchozí nerovnosti psát

$$\begin{aligned} \mathbf{E}(\rho(X_t^x, X_t^y) \mid X_{t-1}^x = x_{t-1}, X_{t-1}^y = y_{t-1}) &= \mathbf{E}(\rho(X_1^{x_{t-1}}, X_1^{y_{t-1}})) \\ &\leq \rho(x_{t-1}, y_{t-1}) \left(1 - \frac{c_{met}(\Delta, q)}{n}\right). \end{aligned}$$

Aplikací střední hodnoty přes (X_{t-1}^x, X_{t-1}^y) pak dostáváme

$$\mathbf{E}(\rho(X_t^x, X_t^y)) \leq \mathbf{E}(\rho(X_{t-1}^x, X_{t-1}^y)) \left(1 - \frac{c_{met}(\Delta, q)}{n}\right).$$

Opakovaným užitím této nerovnosti a aplikací odhadu (5.13) v závěru pak dostáváme

$$\mathbf{E}(\rho(X_t^x, X_t^y)) \leq \rho(x, y) \left(1 - \frac{c_{met}(\Delta, q)}{n}\right)^t. \quad (5.14)$$

Uvědomíme-li si, že platí $\rho(x, y) \geq 1$, právě když je $x \neq y$, můžeme díky Markovově nerovnosti (3.4) a odhadu (5.14) psát

$$\begin{aligned} \mathbf{P}\{X_t^x \neq X_t^y\} &= \mathbf{P}\{\rho(X_t^x, X_t^y) \geq 1\} \leq \mathbf{E}(\rho(X_t^x, X_t^y)) \\ &\leq \rho(x, y) \left(1 - \frac{c_{met}(\Delta, q)}{n}\right)^t \leq ne^{-t(c_{met}(\Delta, q)/n)}, \end{aligned} \quad (5.15)$$

kde v poslední nerovnosti jsme využili faktu, že náš graf má n vrcholů a dále standardního odhadu pro exponenciálu, podle kterého pro $q \in \mathbb{R}$ platí vztah $1 - q \leq e^{-q}$.

Výše uvedené odhady platí pro všechna obarvení $x, y \in \tilde{\Omega}$, proto speciálně platí i pro všechna *přípustná* obarvení $x, y \in \Omega$. Z *Důsledku 18* a nerovnosti (5.15) potom plyne $d(t) \leq ne^{-t(c_{met}(\Delta, q)/n)}$. Odtud dostáváme, že $d(t) \leq \varepsilon$ pro

$$t > c_{met}(\Delta, q)^{-1}n [\log n + \log(1/\varepsilon)],$$

čímž je důkaz vztahu (5.9) dokončen. □

Seznam použité literatury

- [1] LEVIN David A., PERES Yuval, WILMER Elizabeth L.: *Markov Chains and Mixing Times*. American Mathematical Society, 2008.
- [2] PRÁŠKOVÁ Zuzana, LACHOUT Petr: *Základy náhodných procesů*. Karolinum, Praha, 1998.
- [3] LACHOUT Petr: *Teorie pravděpodobnosti*. Karolinum, Praha, 2004.