微专题：

马尔可夫链基础

Markov Chains

马尔可夫链基础 Markov Chains

Levin & Peres, Markov Chains and Mixing times, 2nd edition.

- definition
- property
- stationary
- convergence.
- MCMC
- mixing time
- coupling

1. 马尔可夫性质. 转移与稳态

def (Random Process) 随机过程 为一族随机变量 $\{X_t : t \in T\}$

X_t 在状态空间 (state space) Ω 中取值 通: $\mathbb{R}$

T 是一组下标. 每个 $t \in T$ 可被解释成"时间"

- discrete-time: $t = 0, 1, 2, \cdots$
- continuous-time: $T = [0, \infty)$

def (Markov Chain) 一个离散时间的随机过程 $X_0, X_1, X_2, \cdots$ 为马尔可夫链

若 $\Pr(X_{t+1} = x_{t+1} \mid X_0 = x_0, X_1 = x_1, \cdots, X_t = x_t) = \boxed{\Pr(X_{t+1} = x_{t+1} \mid X_t = x_t)} = P^{(t)}_{x_t x_{t+1}}$

对 $\forall t \in \mathbb{N}$, $\forall x_0, x_1, \cdots, x_t, x_{t+1} \in \Omega$ 成立

$\boxed{\text{Markov Property}}$ (memoryless)

The next state X_{t+1} depends on the current state X_t but is independent of the history $X_0, X_1, \cdots, X_{t-1}$ of how the process arrived at state X_t.

$|\Omega|$ 有限 $\Rightarrow$ 有穷 Markov Chain.

$X_t \to X_{t+1}$

$X_0 \to X_1 \to \cdots \to X_t \to X_{t+1}$

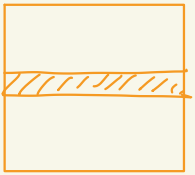
X_{t+1} 和 $X_0, \cdots, X_{t-1}$ given X_t: conditionally indep

时间同质 (time-homogeneous)

转移矩阵 (transition matrix) $P \in \mathbb{R}^{|\Omega| \times |\Omega|}$: $P_{xy} = \Pr(X_{t+1} = y \mid X_t = x)$

↳ row-stochastic (行随机): $\forall x \in \Omega \quad \sum_{y \in \Omega} P_{xy} = 1$

P_{xy} 刻画了从状态 x 到状态 y 的转移概率.



MC 随机过程. $\Rightarrow$ X_t 随机变量 distribution

↑ pmf 概率质量函数

$$X_t \sim \pi^{(t)} \in \mathbb{R}_{\geq 0}^{|\Omega|}$$ 行向量. 代表一个概率分布

$$\pi_x^{(t)} = Pr(X_t = x)$$

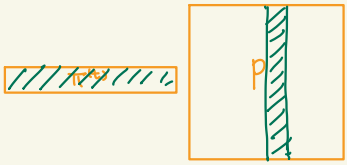
Memorylessness: $Pr(X_{t+1} = y \mid X_t = x) = P_{xy}$ ← 转移概率

$$\pi_y^{(t+1)} = Pr(X_{t+1} = y)$$

$$= \sum_{x \in \Omega} Pr(X_t = x) \cdot Pr(X_{t+1} = y \mid X_t = x)$$

$$= \sum_{x \in \Omega} \pi_x^{(t)} \cdot P_{xy}$$

$$= (\pi^{(t)} \cdot P)_y$$



$$\Rightarrow \boxed{\pi^{(t+1)} = \pi^{(t)} P} \quad \forall t \in \mathbb{N}$$

第t+1时刻的分布　第t时刻的分布

右乘一次P. 进行一次状态转移.

$$\Rightarrow \pi^{(m+n)} = \pi^{(m)} \cdot P^n$$

$$\Rightarrow \boxed{\pi^{(t)} = \pi^{(0)} \cdot P^t}$$

↑ Chapman-Kolmogorov 方程.

每一时刻的状态 X_t 的分布. 可由初始时刻状态 X_0 的分布 $\pi^{(0)}$ 和转移矩阵 P 通过线性运算得出

MC: $M = (\Omega, P, \pi^{(0)}) \xrightarrow{?} M = (\Omega, P)$

状态空间　转移矩阵　初始分布

· 状态空间 Ω 上的马尔可夫链等价于图上的随机游走 (random walk)

每个状态 $x \in \Omega$ 代表图上的一个顶点 $G=(V,E)$

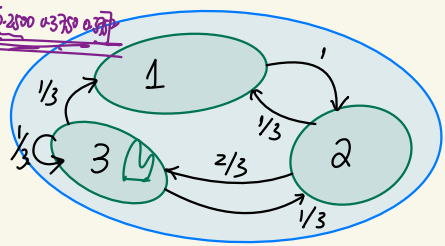
两个状态 x, y 之间的边为带权边，权重为 $P_{xy} = Pr(X_{t+1}=y \mid X_t = x)$

$\Omega = \{1,2,3\}$

$$P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

$$\overset{\pi^{(0)}}{\begin{bmatrix} \pi_1^{(0)}, \pi_2^{(0)}, \pi_3^{(0)} \end{bmatrix}} \cdot \underbrace{P^{10}}_{\approx} = \begin{bmatrix} 0.2500 & 0.3750 & 0.3750 \\ 0.2500 & 0.3750 & 0.3750 \\ 0.2500 & 0.3750 & 0.3750 \end{bmatrix} = \begin{bmatrix} 0.2500 & 0.3750 & 0.3750 \end{bmatrix}$$

$$\pi = \left(\frac{1}{4}, \frac{3}{8}, \frac{3}{8}\right)$$



def (Stationary distribution) 令 π 为状态空间 Ω 上的一个概率分布，即 $\pi \in \Delta^{|\Omega|}$ (i.e. $\pi \in \mathbb{R}^{|\Omega|}_{\geq 0}$ 且 $\|\pi\|_1 = 1$)。若分布 π 满足 $\pi = \pi P$，则称 π 为转移矩阵 P 对应的马尔可夫链的一个稳定分布。

$\pi^{(t+1)} = \pi^{(t)} \cdot P$

$X_0 \to X_1 \to \cdots \to X_t \to X_{t+1} \to X_{t+2} \to X_{t+3} \to \cdots$ "吸收分布"

$\pi \quad \pi P \quad \pi P^2 \quad \pi P^3$

$\quad\;\; = \quad = \quad =$

$\quad\;\; \pi \quad \pi \quad \pi$

找 π：解线性系统

π 是线性变换 P 的一个不动点 (fixed point, equilibrium)

(特征值为 1 的左特征向量)

Thm (Perron-Frobenius) 令 $A \in \mathbb{R}^{n \times n}_{\geq 0}$ 为一非负方阵，谱半径为 $\rho = \rho(A) > 0$。则：

$\max\{|\lambda_1|, |\lambda_2|, \cdots, |\lambda_n|\}$

- ρ 是 A 的一个特征值，且存在一个非负(左或右)特征向量
- 如果 A 不可约 (irreducible)，则特征值 ρ 的重数为 1，且存在一个正的特征向量
- 如果 A 是(行或列)随机的，则其谱半径 $\rho = \rho(A) = 1$

转移矩阵 P 是 row-stochastic

$\Rightarrow$ 任何一个有穷的 MC 都存在稳定分布 π。

irreducible: the graph G is strongly-connected

任何一个有穷的不可约的 MC 都存在唯一的稳定分布 π。

def (time-reversible) 一个马不可夫链 $m=(\Omega, P)$ 是时间可逆的，

如果存在一个 Ω 上的分布 π 满足如下的细致平衡方程 (Detailed Balanced Equation, DBE)

$$\forall x, y \in \Omega: \quad \boxed{\underbrace{\pi_x P_{xy}}_{x\to y} = \underbrace{\pi_y P_{yx}}_{y\to x}}$$

sufficient condition of stationary!

Proposition. 任何满足DBE的分布 π 必然是一个稳态分布

$$(\pi P)_y = \sum_{x\in\Omega} \pi_x \cdot P_{xy} = \sum_{x\in\Omega} \pi_y \cdot P_{yx}$$

$$= \pi_y \cdot \sum_{x\in\Omega} P_{yx} = \pi_y \Rightarrow \pi P = \pi.$$

$\pi P = \pi$: macro equilibrium.

DBE: refined equilibrium.

Proposition. 对于一个可逆的马不可夫链 $(X_t)_{t=0,1,2,\cdots}$，如果初始状态 X_0 的分布满足DBE $\pi^{(0)} \Rightarrow \pi$

则对于任意 $t \geq 0$，随机向量 $(X_0, X_1, \cdots, X_t)$ 和 $(X_t, X_{t-1}, \cdots, X_0)$ 同分布.

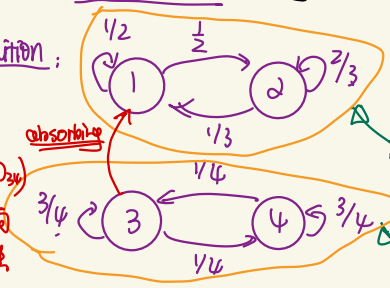
任一时刻的分布都是稳态分布

2. 马不可夫链基本定理

$M=(\Omega, P, \pi^{(0)}) \to M=(\Omega, P)$

Motivation: 什么条件下，马不可夫链可以从任意初值出发向稳态分布 π 收敛？

(A) Intuition:



$\pi = (\pi_{12}, 0_{34})$

吸收不到的稳态分布

从它出发

从它出发

非连通图

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 1/4 & 3/4 \end{pmatrix}$$

disconnected!

$$\Rightarrow P^{20} \approx \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

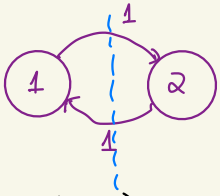
$\pi = (\pi_{12}, \pi_{34})$

$\sum_{k\in\Omega} P_{ik} P_{kj}$

def (irreducibility)

- 状态 x 到状态 y 是可达的 (accessible). $x \rightsquigarrow y$，若图中存在一条从 x 到 y 的有向路径 (# of hops) $\exists n \geq 1$ s.t. $P^n_{xy} > 0$
- 状态 x 和状态 y 是互通的 (communicable). $x \leftrightsquigarrow y$
- 对一个马不可夫链 (Ω, P) 而言，若所有状态都彼此互通 (即对应的图为强连通图)，则称这个链是不可约的 (irreducible) $\forall x, y \in \Omega$. $\exists n \geq 1$. s.t. $P^n_{xy} > 0$

(B)



$P=\begin{bmatrix}0&1\\1&0\end{bmatrix} \Rightarrow P^{2n}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$, $P^{2n+1}=\begin{bmatrix}0&1\\1&0\end{bmatrix}$.

周期性 $(d=2)$

def (aperiodicity)

$d = \gcd\{n\geq 1 \mid P^n_{xx}>0\}$

· 一个状态 $x\in\Omega$ 具有周期 d (period), 如果图中所有包含 x 的回路(未必是单纯回路)的长度都能被 d 整除, 且 d 是满足该性质的最大正整数.

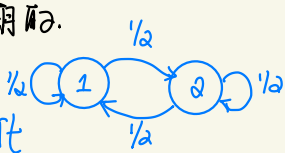
· 对一个马不可夫链 (Ω, P) 而言, 若所有状态的周期均为 1, $d_x=1 \ \forall x\in\Omega$

则称这个链是非周期的 (aperiodic)

Proposition. 若所有状态 $x\in\Omega$ 满足 $P_{xx}>0$, 则该链是非周期的.

$n=1$. $d=1$.

break symmetry by being lazy 懒惰性

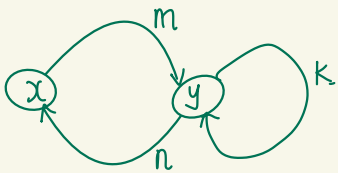


"laziness"

Proposition. 如果状态 x 和状态 y 是互通的, 则其周期 $d_x=d_y$.

$\exists m,n$. $P^m_{xy}>0$. $P^n_{yx}>0$.

周期是SCC中所有点的共同属性



$d(x) \mid m+n$,

$d(x) \mid m+n+k$ $\Rightarrow d(x) \mid k$.

$d(y)=k$ $\Rightarrow d(x) \leq d(y)$

similarly, $d(x)\geq d(y)$ $\Rightarrow d(x)=d(y)$.

Thm (Fundamental Theorem of Markov Chain)

任何一个不可约、非周期的有限马不可夫链都仅会收敛到唯一的稳态分布. 也即:

Ω 上存在唯一的稳态分布 $\pi=\pi P$, 使得对于 Ω 上的任意初始分布 $\pi^{(0)}$, 都满足

$$\pi = \lim_{t\to\infty} \pi^{(0)} P^t$$

存在性和唯一性 by Perron-Frobenius

convergence thm. of MC

limit thm of MC

* 无穷马尔可夫链的收敛性：

Ω countably infinite.

def (ergodicity)

- 状态 $x \in \Omega$ 是瞬时的 (transient)，若 $\Pr(X_n = x \text{ for inf. many } n \mid X_0 = x) = 0$
- 状态 $x \in \Omega$ 是常返的 (recurrent)，若 $\Pr(X_n = x \text{ for inf. many } n \mid X_0 = x) = 1$

  (i.e. 从 x 出发会以概率 1 再次访问 x) a.s.

Zero-One Law: $f_{xx} = \Pr(\exists\, n \geq 1,\ X_n = x \mid X_0 = x)$

$x \in \Omega$ is recurrent iff $f_{xx} = 1$, transient iff $f_{xx} < 1$

- 状态 $x \in \Omega$ 是正常返的 (positive recurrent)，若 x 常返且平均返回时间有穷.

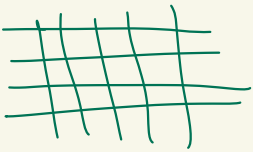
$$\mathbb{E}\left[\min\{t \geq 1 : X_t = x\} \mid X_0 = x\right] < \infty$$

- 对一个马尔可夫链 (Ω, P) 而言，若所有状态都是正常返的，且同期为 1.

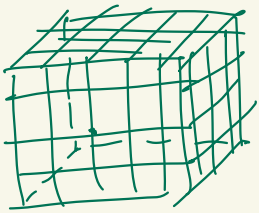
  则称这个链是遍历的 (ergodic)

ergodicity $\begin{cases} \text{positive recurrent} \\ \text{aperiodicity} \end{cases}$

Thm 任何一个不可约、遍历的马尔可夫链都收敛到唯一的稳态分布.



$\mathbb{Z}^2$: recurrent



$\mathbb{Z}^3$: transient

Shizuo Kakutani (角谷静夫):

A drunk man will find his way home, but a drunk bird may get lost forever.

(醉汉易归家，醉鸟难返巢)

Recall: 对于有穷的马尔可夫链，若其不可约 (irreducible) 且无周期 (aperiodic)

则从任意初始状态 π^0 出发，都会收敛到唯一的稳态分布 π.

存在性 唯一性 ✓

只需证：从任意 $\pi^{(0)}$ 出发，都会收敛到稳定分布 π，

$$\lim_{t\to\infty} \pi^{(0)} P^t = \pi, \quad \forall \pi^{(0)} \in \mathcal{L}^{(\Omega)}$$

证明：(X_t)：初始分布为任意 $\pi^{(0)}$

(Y_t)：初始分布为稳定 π.

$\xrightarrow{\text{coupling}}$ (X_t, Y_t) 耦合链

s.t. X_t 和 Y_t 的边缘分布都分别与链 (X_t), (Y_t) 一致.

$$\begin{array}{ccccccccccccc}
\pi^{(0)} & \sim & X_0 & \to & X_1 & \to & \cdots & \to & X_n & & & & \\
 & & \# & & \# & & & & \| & & & & \\
\pi & \sim & Y_0 & \to & Y_1 & \to & \cdots & \to & \underbrace{Y_n}_{\pi} & \to & Y_{n+1} & \to \cdots \to & Y_t \sim \pi
\end{array}$$

若 $X_t \neq Y_t$：根据转移矩阵 P 独立地生成 X_{t+1} 和 Y_{t+1}.

即：$\Pr(X_{t+1}=x' \wedge Y_{t+1}=y' \mid X_t=x \wedge Y_t=y) = P_{xx'} \cdot P_{yy'}$

若 $X_t = Y_t$：则 $X_{t+1} = Y_{t+1}$，且符合转移 P.

即：$\Pr(X_{t+1}=Y_{t+1}=y \mid X_t=Y_t=x) = P_{xy}$

$\Rightarrow$ 如果 $X_n = Y_n$，则对 $\forall t \geq n$：$X_t = Y_t$.

而 $Y_t \sim \pi$ 为稳定分布 $\Rightarrow$ 只需证：$\boxed{\Pr(X_t = Y_t) = 1}$ $t \to \infty$

Coupling Lemma，耦合引理

$\pi^{(0)} P^t = \pi$

$$\|\pi^{(0)} P^t - \pi\|_{TV} \leq \Pr(X_t \neq Y_t) = 0$$

irreducibility：$\exists$ finite $s>0$，$\forall x,y \in \Omega$，$\exists t \leq s$ s.t. $P^t_{xy} > 0$

$s \leq s$

aperiodicity：$\exists$ finite $r>0$，$\forall x \in \Omega$ $\forall t \geq r$，$P^t_{xx} > 0$

$r \geq r$，$r+s \geq r$

(对于 gcd 为 1 的正整数集合 $S=\{a_1, a_2, \cdots\}$，必存在一个有界的 r，使得对于任意 $t \geq r$ 都可以仅通过 S 中的数相加构成)

Frobenius #

Frobenius coin problem

$\forall x, y \in \Omega$: $P_{xy}^{r+s} = Pr(X_{t+r+s} = y \mid X_t = x)$

$\geq Pr(X_{t+r+s} = y \mid X_{t+s} = y) \cdot Pr(X_{t+s} = y \mid X_t = x)$

选取 $x \xrightarrow{s} y \xrightarrow{r} y$

$= P_{yy}^{r} \cdot P_{xy}^{s} > 0, \quad \forall t$

finite

↓

每 $r+s$ 步都有非零的概率独立地使得 $X_t = Y_t$.

$Pr(X_{t+r+s} = Y_{t+r+s} = y \mid X_t = x, Y_t = y)$

$= Pr(X_{t+r+s} = y \mid X_t = x) \cdot Pr(Y_{t+r+s} = y \mid Y_t = y)$

$= P_{xy}^{r+s} \cdot P_{yy}^{r+s} > 0$

$t \to \infty$: $Pr(X_t \neq Y_t) = 0 \Rightarrow Pr(X_t = Y_t) = 1$.

$\lim_{t \to \infty} \pi^{(0)} P^t = \pi$

□

3. 马尔可夫链蒙特卡罗方法 (Markov Chain Monte Carlo, MCMC)

MCMC 是一类用于近似复杂概率分布的采样方法，其主要思想是：通过构建一个马尔可夫链来生成一系列样本点（在对应的图上进行随机游走），这些样本点最后将逐渐逼近目标概率分布.

Bayes Estimation

X: data

θ: parameter

→ posterior $\boxed{\lambda(\theta \mid X)} = \dfrac{\lambda(\theta) \cdot P_\theta(X)}{\int_\Omega \lambda(\eta) \cdot P_\eta(X)\, d\eta}$

flat

Jeffrey prior

likelihood

evidence

High-Dimensional Ω

Computational Strategy: set up a MC with stationary distribution $\propto \pi(\theta)\pi(x)$

Probabilistic Graphic Model

often intractable

$\alpha, \beta \sim \lambda_0$

$\theta \sim \mathrm{Beta}(\alpha, \beta)$

$x \sim B(n, \theta)$

Hierarchical Bayes

· 图上随机游走：给定一个无向图 $G=(V,E)$，我们考虑该无向图上的简单随机游走

· 每一步从当前点 $u \in V$ 的邻居中按均匀分布随机选一个点 v，走到 v

转移矩阵 P：
$$P_{uv} = \begin{cases} \frac{1}{\deg(u)}, & \text{if } (u,v) \in E \\ 0 & \text{otherwise} \end{cases}$$

稳定分布 π：
$$\pi_v = \frac{\deg(v)}{\sum_{u \in V} \deg(u)} = \frac{\deg(v)}{2|E|}$$

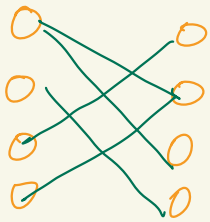
可逆! by DBE

if $(u,v) \in E$: $\pi_u P_{uv} = \pi_v P_{vu} = \frac{1}{2|E|}$.

if $(u,v) \notin E$: $\pi_u P_{uv} = \pi_v P_{vu} = 0$

☆ 随机游走
- 不可约 $\Leftrightarrow$ 图 G 连通
- 非周期 $\Leftrightarrow$ 图 G 不是二分图

↓ 避免 periodicity



· 懒情版本的简单随机游走：

· 每一步以 $\frac{1}{2}$ 概率什么都不做，

$\hookrightarrow 1/\Delta$

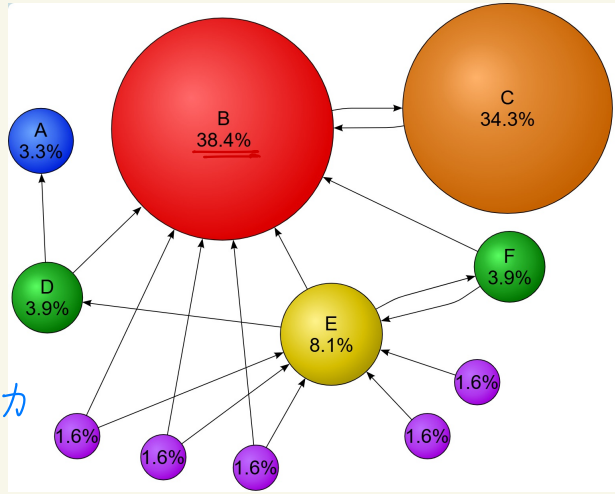
· 否则从当前点 $u \in V$ 的邻居中按均匀分布随机选择一个点 v，走到 v

$$P^{LZ} = \frac{1}{2} I + \frac{1}{2} P$$

Proposition. 该懒情版本的简单随机游走不可约当且仅当图 G 连通，且总是非周期

"懒情凡"

一个应用：PageRank 算法



- 为每个网页 $x \in \Omega$ 分配一个排名 $r(x)$
- 一个网页拥有高排名，如果它被很多高排名的网页指向
- 高排名的网页有更高的影响力
- 指向更少其他网页的网页有更高的影响力

$d \to d^+$：out-degree

$$r(x) = \sum_{y \to x} \frac{r(y)}{d^+(y)} \quad \to \quad P_{xy} = \begin{cases} \frac{1}{d^+(x)}, & \text{if } x \to y \\ 0 & \text{o.w.} \end{cases}$$

Back to MCMC:

- 给定一个无向图 $G=(V,E)$，我们希望得到一个稳态分布为 V 上均匀分布的随机游走，如何设计转移矩阵？ posterior

$$\forall u,v \in V. \quad \boxed{P_{uv} = \begin{cases} 1 - \frac{\deg(u)}{\Delta}, & \text{if } u=v \\ \frac{1}{\Delta}, & \text{if } (u,v) \in E \\ 0, & \text{ow.} \end{cases}} \quad (*)$$

$\Delta \Downarrow$ complete graph.

$\Delta = \max_{v \in V} \deg(v)$ 是图 G 的最大度数

① P：row-stochastic $\quad P\vec{1} = \vec{1}$

② 可逆，且稳态分布为均匀分布（P：对称矩阵 + 均匀分布 $\Rightarrow$ DBE）

③ 不可约性 $\Leftrightarrow$ 图 G 连通 ∴非周期性 ✓

- 进一步地，如果希望指定的一个 $\boxed{V\text{上的分布}\pi\text{作为稳态分布}}$ 如何设计对应的转移矩阵 P' posterior

by DBE，$\dfrac{P'_{xy}}{P'_{yx}} = \dfrac{\pi_y}{\pi_x}$.

Metropolis-Hastings 算法：

通过"提议-过滤"(propose-filter)的范式来将稳态分布为均匀分布的转移矩阵 P 转变为转移矩阵 P'. 在每一步：

1. 随机选择一个当前点 $u \in V$ 的邻居 v，每个邻居以概率 $1/\Delta$ 被选中，

proposal 提议走到 v；剩余的 $1-\frac{\deg(u)}{\Delta}$ 概率停留在 u.

→ 根据(u)进行状态转移.

2. 以概率 $\min\left\{1, \frac{\pi_v}{\pi_u}\right\}$ 接受这个提议，走到 v，否则停留在 u.

accept　filter

消除了对 evidence 的计算

$$\forall u,v \in V: \quad P'_{uv} = \begin{cases} P_{uv} \cdot \min\left\{1, \frac{\pi_v}{\pi_u}\right\} & u \neq v \\ 1 - \sum_{w \neq u} P'_{uw} & u = v \end{cases}$$

propose　accept

Proposition. MC $m = (\Omega, P')$ 非周期、可逆、且稳态分布为 π.

(can be proved by DBE: $\pi_u P'_{uv} = \pi_v \cdot P'_{vu}$ $\Leftrightarrow$ $\pi_u \cdot P_{uv} \min\left\{1, \frac{\pi_v}{\pi_u}\right\} = \pi_v P_{vu} \min\left\{1, \frac{\pi_u}{\pi_v}\right\}$ $u \neq v$)

P 对称 $\begin{cases} \pi_v > \pi_u: \ \cancel{\pi_u} P_{uv} = \cancel{\pi_u} \cdot P_{vu} \cdot \frac{\cancel{\pi_u}}{\cancel{\pi_v}} \\ \pi_v < \pi_u: \ \cancel{\pi_u} \cdot P_{uv} \cdot \frac{\cancel{\pi_v}}{\cancel{\pi_u}} = \cancel{\pi_v} P_{vu} \end{cases}$

Example. 图染色均匀采样

给定一个无向图 $G=(V,E)$. 图 G 的一个 q-染色 (q-coloring) 是一个映射 $\sigma: V \to [q]$.

该染色是一个合法染色 (proper coloring). 如果任意相邻的点的颜色不同. 即

$\forall (u,v) \in E: \sigma(u) \neq \sigma(v)$

→ Δ: maximum degree

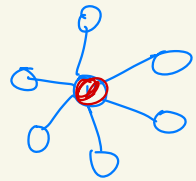
记 Ω 为图 G 的所有合法 q-染色的集合. 通过 Metropolis-Hastings 算法构造图染色链 (称为图染色的单点 Metropolis 链, single-site Metropolis chain)，使该 MC 的稳态分布为均匀分布

posterior ⇒ 永远接受提议

每一步：
- 以 $\frac{1}{2}$ 概率什么也不做. lazy
- 独立且均匀分布地选取顶点 $v\in V$ 和颜色 $c\in[q]$
- 如果 v 染颜色 c 和邻居当前颜色不冲突，则把 v 的颜色染成 c.

symmetric
⇓
stationary distribution
π is uniform over Ω

$\boxed{X_t} \longrightarrow \boxed{X_{t+1}}$ Goal：采样尽可能多的合法染色



$q \le \Delta+1$：reducible,

$q \ge \Delta+2$：irreducible & aperoidic.（每次改变一个点的颜色 $x\to y$）

Gibbs Sampling

Bayes estimation 中的参数 (θ)

- 考虑如图染色这样的联合分布的随机向量 $X=(X_1,X_2,\cdots,X_n)\in[q]^n$，它的分布是空间 $\Omega=[q]^n$ 上的一个分布 μ.

Gibbs采样是一种从 μ 中采样的马尔可夫链，称作 Glauber Dynamics.

heat-bath random walk

假设当前状态 $X\in\Omega$：

- 以 $\frac{1}{2}$ 的概率什么也不做 → lazy.
- 均匀分布地选取一个 $i\in\{1,2,\cdots,n\}$ single component (coordinate)
- 从边缘分布 $\mu_i(\cdot\mid X_{-i})$ 中采样一个随机值 $c\in[q]$，令 $X_i=c$.

令 $\sigma\in[q]^n$ 为分布为 μ 的随机向量

则 $\mu_i(\cdot\mid X_{-i})$ 为在条件 $\forall j\ne i,\ \sigma_j=X_j$ 下 σ_i 的分布.

X_1
X_2
X_3
X_4

$\mu_1(\cdot\mid X_2,X_3,X_4)$ conditional distribution

↳ 采样得到 X_1'

$\mu_3(\cdot\mid X_1,X_2,X_4)$

↳ 采样得到 X_3'

① Gibbs sampler 是 MH 的一个特例

② MC 非周期、可逆，且稳定分布是 μ.

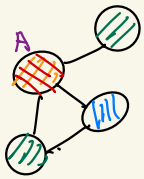
(by DBE)

Example. 图染色均匀采样 —— 从本身是均匀分布

每一步：
- 以 $\frac{1}{n}$ 概率什么也不做
- 均匀分布地选取顶点 $v \in V$
- 从颜色集合 $[q]$ 中去掉 v 邻居当前使用的颜色. 剩下的颜色成为 v 的可用颜色. 均匀分布地选取一个可用颜色 c. 将 v 染成颜色 c.

condition distribution

Proposition. 该MC非周期. 可逆. 且稳态分布是图 G 所有合法 q-colorings 的均匀分布



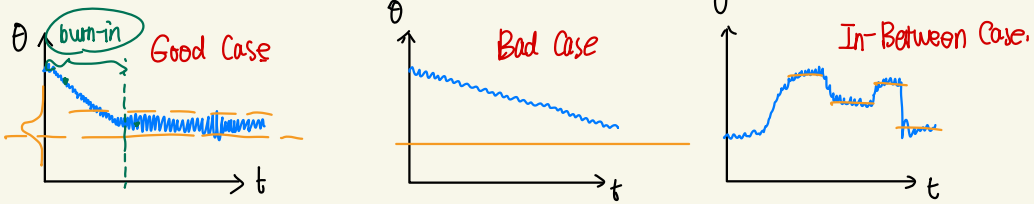
4. 耦合. 混合时间(理论)

Recall: (MC基本定理) 不可约. 非周期 的有穷 MC 会在极限时间收敛到唯一的稳态分布

收敛的速率

MCMC in practice:

trace plot



def (total variation distance) 令 p 和 q 为同一状态空间 Ω 上的概率分布. $p, q \in \Delta^{|\Omega|}$

分布 p 和 q 之间的总变化距离. 记为 $\|p-q\|_{TV}$ 或 $d_{TV}(p, q)$ 定义为:

$$\|p-q\|_{TV} = \boxed{\frac{1}{2}\|p-q\|_1 = \max_{A \subseteq \Omega} |p(A) - q(A)|}$$

任一事件 $A \subseteq \Omega$ 在 p, q 下的概率差的最大值

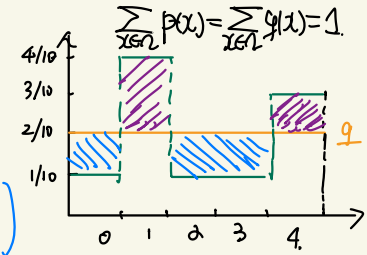
Proof: $B = \{x : p(x) \geq q(x)\}$ $A \subseteq \Omega$.

$$p(A) - q(A) = p(A \cap B) - q(A \cap B) + p(A \cap B^c) - q(A \cap B^c) \leq p(A \cap B) - q(A \cap B)$$

$$p(B) - q(B) = p(A \cap B) - q(A \cap B) + p(A^c \cap B) - q(A^c \cap B) \geq p(A \cap B) - q(A \cap B)$$

$$\Rightarrow p(A) - q(A) \leq p(B) - q(B)$$

Similiarly, $q(A)-p(A) \le q(B^c)-p(B^c)$



$$\|p-q\|_{TV} = \frac{1}{2}\|p-q\|_1 = \frac{1}{2}\sum_{x\in\Omega}|p(x)-q(x)|$$

$$= \frac{1}{2}\left(p(B)-q(B) + q(B^c)-p(B^c)\right)$$

$$= \max_{A\subseteq\Omega}|p(A)-q(A)| \qquad (A=B \text{ 或 } A=B^c) \quad \square$$

Remark. $\|p-q\|_{TV} = \sum_{\substack{x\in\Omega \\ p(x)\ge q(x)}}(p(x)-q(x)) = \sum_{\substack{x\in\Omega \\ p(x)\le q(x)}}(q(x)-p(x))$

def (Coupling) 两个概率分布 p 和 q 的耦合是一对联合分布的随机变量 (X,Y). 其中 X 的边缘分布是 p, Y 的边缘分布是 q.

令 μ 为 (X,Y) 的分布 则 $p=\sum_y \mu(\cdot,y)$, $q=\sum_x \mu(x,\cdot)$.

假设我们希望构造一个 p 和 q 的耦合 (X,Y), 使 X 和 Y 尽可能一样.

我们最好能做到多好? 总变化距离给出了答案.

Convergence Thm:

$$\max_{x\in\Omega} d_{TV}(P^t(x,\cdot),\pi) \le C\cdot\alpha^t.$$

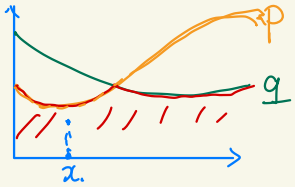
$\alpha\in(0,1)$

耦合引理

Thm (Coupling Lemma) 令 p 和 q 为样本空间 Ω 上的概率分布 则:

1. 对于 p 和 q 的任意耦合 (X,Y), 都有 $\Pr(X\ne Y) \ge \|p-q\|_{TV}$.
2. 存在 p 和 q 的耦合 (X,Y), 使得 $\Pr(X\ne Y) = \|p-q\|_{TV}$.

TV距离衡量了分布 p 和 q 在最优耦合下的概率差.



$$\Pr(X\ne Y) = 1-\Pr(X=Y) = 1-\sum_{x\in\Omega}\Pr(X=Y=x)$$

$$\ge 1-\sum_{x\in\Omega}\min\{p(x),q(x)\} = \sum_{x\in\Omega}\left(p(x)-\min\{p(x),q(x)\}\right)$$

$$= \max_{A\subseteq\Omega}(p(A)-q(A)) = d_{TV}(p,q)$$

def (mixing time) 令 $M=(\Omega,P)$ 为一收敛到稳定分布为 π 的马尔可夫链.

- $\Delta(t)=\max_{x\in\Omega}\|P^t_{x\cdot}-\pi\|_{TV}$ ← 从最坏的初始情况出发, t时刻的状态分布和稳定分布之间的距离
- ε-混合时间: $T(\varepsilon)=\min\{t \mid \Delta(t)\le\varepsilon\}$
- 混合时间: $\boxed{T_{mix}=T(\frac{1}{2e})}$ ← mixing time
- 若 $T_{mix}=(\log|\Omega|)^{O(1)}$, 则称链 M 快速混合 (rapidly mixing)

bound
① 耦合法
② 线性代数
③ 高维扩张图

Proposition 1. $\forall t\ge 0$. $\Delta(t+1)\le\Delta(t)$ 单调性!

Proof: $\Leftrightarrow$ $\forall x\in\Omega$: $\|P^{t+1}_{x\cdot}-\pi\|_{TV}\le\|P^t_{x\cdot}-\pi\|_{TV}$.

MC: $(X_t)_{t\ge0}$ 转移矩阵 P. 初始状态 $X_0=x$

$(Y_t)_{t\ge0}$ ~ P ~ $Y_0=\pi$

By Coupling Lemma, 存在一个耦合 (X_t,X_t) s.t. $\Pr(X_t\ne Y_t)=\|P^t_{x\cdot}-\pi\|_{TV}$.

X_t的分布 Y_t的分布

在此基础上构造耦合 (X_{t+1},Y_{t+1}):

$X_t=Y_t\Rightarrow X_{t+1}=Y_{t+1}$. i.e. $X_{t+1}\ne Y_{t+1}\Rightarrow X_t\ne Y_t$

By Coupling Lemma, $\|P^{t+1}_{x\cdot}-\pi\|_{TV}=\Pr(X_{t+1}\ne Y_{t+1})$

$\le\Pr(X_t\ne Y_t)=\|P^t_{x\cdot}-\pi\|_{TV}$ □

Proposition 2. $\forall 0<\varepsilon<\frac{1}{2}$. $k\ge1$: $\Delta(k\cdot T(\varepsilon))\le(2\varepsilon)^k$ n指数衰减

我们可以仅用 $T(\varepsilon)$ 在一个具体的常数 $\varepsilon=\frac{1}{2e}$ 上的取值来代表 Mixing Time!

Proof: MC: $(X_t)_{t\ge0}$: 从 $x\in\Omega$ 出发 } 转移矩阵为 P $X_t\sim P^t_{x\cdot}$

$(Y_t)_{t\ge0}$: 从 $y\in\Omega$ 出发 $Y_t\sim P^t_{y\cdot}$

记 $\tau=\tau(\varepsilon)$，要证 $\Delta(k\cdot\tau)\le(2\varepsilon)^k$

$$\Leftarrow \quad \forall x,y\in\Omega:\ \|P_{x\cdot}^{k\tau}-P_{y\cdot}^{k\tau}\|_{TV}\le(2\varepsilon)^k. \qquad (*)$$

Claim: $\Delta(t)=\max_{x\in\Omega}\|P_{x\cdot}^{t}-\pi\|_{TV}\le\max_{x,y\in\Omega}\|P_{x\cdot}^{t}-P_{y\cdot}^{t}\|_{TV}$

（π 可以表示成为 $P_{y\cdot}^{t}$，$y\in\Omega$ 的凸组合）

Induction: • $k=1$: $\forall x,y\in\Omega$

$$\|P_{x\cdot}^{\tau}-P_{y\cdot}^{\tau}\|_{TV}\le\|P_{x\cdot}^{\tau}-\pi\|_{TV}+\|P_{y\cdot}^{\tau}-\pi\|_{TV}\le 2\varepsilon$$

三角不等式

By Coupling Lemma, 存在耦合 (X_τ,Y_τ) s.t.

$$\Pr(X_\tau\ne Y_\tau)=\|P_{x\cdot}^{\tau}-P_{y\cdot}^{\tau}\|_{TV}\le 2\varepsilon.$$

• $k>1$: 利用耦合 (X_τ,Y_τ) 构造耦合 $(X_{k\tau},Y_{k\tau})$

• 若 $X_\tau=Y_\tau$: 对所有 $\tau'\ge\tau$: $X_{\tau'}=Y_{\tau'}$ $(X_{k\tau}=Y_{k\tau})$

• 否则，$\forall x'\ne y'$. 在 $(X_\tau,Y_\tau)=(x',y')$ 的条件下，

令 $(X_{k\tau},Y_{k\tau})$ 为该情况下两个随机变量的最优耦合：

$$\Pr\left(X_{k\tau}\ne Y_{k\tau}\,\middle|\,(X_\tau,Y_\tau)=(x',y')\right)=\|P_{x'\cdot}^{(k-1)\tau}-P_{y'\cdot}^{(k-1)\tau}\|_{TV}$$

"初始化"

By Coupling Lemma, $\|P_{x\cdot}^{k\tau}-P_{y\cdot}^{k\tau}\|\le\Pr(X_{k\tau}\ne Y_{k\tau})$

$$=\Pr(X_\tau\ne Y_\tau)\,\Pr(X_{k\tau}\ne Y_{k\tau}\mid X_\tau\ne Y_\tau)$$

$$\le 2\varepsilon\cdot\sum_{x'\ne y'}\Pr\left((X_\tau,Y_\tau)=(x',y')\right)\cdot\Pr\left(X_{k\tau}\ne Y_{k\tau}\,\middle|\,(X_\tau,Y_\tau)=(x',y\right.$$

5. 耦合. 混合时间(实例)

通过耦合法(coupling method)分析马尔可夫链的混合时间.

Coupling Lemma, P. Q $\quad \forall$ coupling (X, Y) of $p, q \Rightarrow \Pr(X \neq Y) \geq \|p-q\|_{TV}$

$\exists$ coupling (X_0, Y_0) of $p, q \Rightarrow \Pr(X_0 \neq Y_0) = \|p-q\|_{TV}$

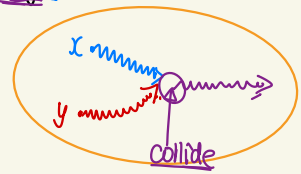
Mixing time: $\tau(\varepsilon) = \max_{x \in \Omega} \min\{t \geq 1 \mid \|P_{x\cdot}^t - \pi\|_{TV} \leq \varepsilon\}$

non increasing: $\Delta(t) \geq \Delta(t+1)$

geometric convergence: $\tau(\varepsilon) = \lceil \log \frac{1}{\varepsilon} \rceil \cdot \tau_{mix}$

def (Coupling of Markov Chains) 马尔可夫链 $\mathcal{M} = (\Omega, P)$ 的一个耦合, 为一个状态空间 $\Omega \times \Omega$ 的马尔可夫链 $\{(X_t, Y_t)\}_{t \geq 0}$, 满足: Markovian Coupling

- $\forall t \geq 0, \forall x, y \in \Omega$, $\Pr(X_{t+1} = y \mid X_t = x) = \Pr(Y_{t+1} = y \mid Y_t = x) = P_{xy}$. "独立性关系"
- 若 $X_t = Y_t$, 则 $X_{t+1} = Y_{t+1}$. "粘合"



$$\begin{array}{l} x \sim X_0 \to X_1 \to \cdots \to X_n \\ \qquad\qquad\qquad\quad \| \\ y \sim Y_0 \to Y_1 \to \cdots \to Y_n \end{array} \Rightarrow \underset{(Y_{n+1})}{X_{n+1}} \to \cdots \to \underset{(Y_t)}{X_t} \to \cdots$$

Thm 令 $\{(X_t, Y_t)\}_{t \geq 0}$ 为马尔可夫链 $\mathcal{M} = (\Omega, P)$ 的一个耦合. 如果对于 $t = t(\varepsilon) > 0$.

满足: $\forall x, y \in \Omega$, $\Pr(X_t \neq Y_t \mid X_0 = x \wedge Y_0 = y) \leq \varepsilon$

则 $\mathcal{M}$ 的混合时间有上界: $\tau(\varepsilon) \leq t(\varepsilon)$. 耦合法分析混合时间上界

Proof: $\tau(\varepsilon) = \min\{t \mid \Delta(t) \leq \varepsilon\}$ 只需证 $\Delta(t) \leq \varepsilon$.

$$\begin{aligned} \Delta(t) &= \max_{x \in \Omega} \|P_{x\cdot}^t - \pi\|_{TV} \\ &\leq \max_{x \in \Omega} \|P_{x\cdot}^t - P_{y\cdot}^t\|_{TV} \\ &\leq \max_{x \in \Omega} \Pr(X_t \neq Y_t \mid X_0 = x \wedge Y_0 = y) \quad \text{(Coupling Lemma)} \\ &\leq \varepsilon \qquad \square \end{aligned}$$

Example. 超立方体随机游走

$n=5$: $x = 10011$, $y = 00110$ $\quad d_{Hamming}(x,y)=3$

- 考虑 n 维超立方体 (hypercube) $\Omega=\{0,1\}^n$. 任意 x,y 相邻当且仅当 x 和 y 之间的 Hamming 距离为 1, 即 x 与 y 刚好相差一个比特.
- 超立方体上的简单均匀随机游走 (simple uniform random walk). 每一步对当前状态 $x\in\Omega$ 做如下操作:
  - 以 $\frac{1}{2}$ 的概率, 什么也不做
  - 否则, 均匀分布地随机选取一个 $i\in[n]$, 将 x 的第 i 位 x_i 改为 $1-x_i$.

随机游走

Proposition. 该随机游走不可约, 非周期, 可逆, 且稳态分布为 Ω 上的均匀分布

等价于每一步对当前状态 $x\in\Omega$ 做如下操作:

- $\boxed{\text{独立且均匀分布地选取 } i\in\{1,2,\cdots,n\} \text{ 和 } b\in\{0,1\}, \text{ 将 } x_i \text{ 赋值为 } b}$

w.p. $\frac{1}{2}$ $\frac{1}{2}$

$\{(X_t,Y_t)\}_{t\geq 0}$:

初态不同, 随机性相同 (X_t 和 Y_t 共享相同的 i 和 b)

$(X_t,Y_t)\to(X_{t+1},Y_{t+1})$ $\quad$ (一旦 $X_t=Y_t$, 就有 $X_{t+1}=Y_{t+1}$. $\Rightarrow$ 称该 coupling)

$\tau(\varepsilon)\leq t$: $\Pr(X_t\neq Y_t \mid X_0=x \wedge Y_0=y)\leq\varepsilon$

Observation: 一旦全部 n 个 bit 都被重新赋值过, 就一定有 $X_t=Y_t$. (t)

Balls-into-Bins $\to$ n bins

Coupon-Collector

X: total # of balls thrown to make all n bins nonempty.

X_i: # of balls thrown while there are exactly $(i-1)$ non-empty bins

$\hookrightarrow$ 几何分布, $p_i=1-\frac{i-1}{n}=\frac{n-i+1}{n}$

Hoeffding inequality

$\Rightarrow \Pr(t>n\log n+cn)\leq e^{-c}=\varepsilon$

$X=\sum_{i=1}^n X_i \Rightarrow \mathbb{E}[X]=\sum_{i=1}^n \mathbb{E}[X_i]=\sum_{i=1}^n\frac{n}{n-i+1}=n\cdot\sum_{i=1}^n\frac{1}{i}\doteq n\log n$

$t=n\log n+\lceil n\cdot\log\frac{1}{\varepsilon}\rceil \Rightarrow \Pr(X_t\neq Y_t\mid X_0,Y_0)\leq\varepsilon \Rightarrow \tau(\varepsilon)\leq n\log n+\lceil n\log 1/\varepsilon\rceil$

Example. 图染色均匀采样

Recall：图染色的单点 Metropolis 链，有一步如下：

- 独立且均匀分布地选取顶点 $v \in V$ 和颜色 $c \in [q]$ "proposal"
- 如果 v 染颜色 c 与邻居当前颜色不冲突，就把 v 的颜色染成 c "accept/filter"

若 $q \geq \Delta+2$，则该MC是不可约的

maximum degree

$x \rightsquigarrow y$

Intuition：q 相对 Δ 越大，状态空间 Ω 的连通性越好，更容易快速混合

$q \geq \alpha \cdot \Delta + \beta$

(SOTA：$q \geq \frac{11}{6}\Delta$. 1999 ；$q \geq \left(\frac{11}{6} - \varepsilon_0\right)\Delta$. 2018

$\varepsilon_0 \geq 9.4 \times 10^{-5}$

Thm 当 $q \geq 4\Delta+1$ 时，图染色链满足 $T(\varepsilon) \leq \left\lceil \frac{q}{q-4\Delta} \cdot n \log \frac{n}{\varepsilon} \right\rceil$

rapidly mixing

$n = |V|$ 是顶点数

Proof：从一对合法的 q-coloring 出发，

构造耦合 $(X, Y) \to (X', Y')$：两条链选择相同顶点 v 和颜色 c. (一步)

(sharing randomness) (o)

$d(X, Y) = \left|\{v \in V \mid X_v \neq Y_v\}\right|$ 为染色 X 和 Y 之间的 Hamming 距离

Goal：$0 < \delta < 1$. $\forall X, Y \in \Omega$

$\frac{q-4\Delta}{nq}$

☆ "期望意义下" $\boxed{\mathbb{E}[d(X', Y') \mid X, Y] \leq (1-\delta) \cdot d(X, Y)}$ (*)

假设(*)式成立，构造耦合链 $\{(X_t, Y_t)\}_{t \geq 0}$

每一步转移 $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$ 都由 (o) 给出

$$\mathbb{E}[d(X_t, Y_t) \mid X_0, Y_0] = \mathbb{E}[\mathbb{E}[d(X_t, Y_t) \mid X_0, Y_0, X_{t-1}, Y_{t-1}] \mid X_0, Y_0]$$

$$= \mathbb{E}[\mathbb{E}[d(X_t, Y_t) \mid X_{t-1}, Y_{t-1}] \mid X_0, Y_0]$$ (Markov Property)

$$\le \left(1-\frac{q-4\Delta}{nq}\right)\cdot \mathbb{E}\left[d(X_{t-1},Y_{t-1})\mid X_0,Y_0\right] \quad (*)$$

$\mathbb{E}[X]=\mathbb{E}[\mathbb{E}[X\mid Y]]$

$\mathbb{E}[X\mid Z]=\mathbb{E}_X[\mathbb{E}_Y[X\mid Y,Z]\mid Z]$

$$\Rightarrow \mathbb{E}\left[d(X_t,Y_t)\mid X_0,Y_0\right] \le \left(1-\frac{q-4\Delta}{nq}\right)^t\cdot d(X_0,Y_0)$$

$$\le \left(1-\frac{q-4\Delta}{nq}\right)^t\cdot n$$

$$\Pr(X_t\ne Y_t\mid X_0,Y_0) = \Pr(d(X_t,Y_t)\ge 1\mid X_0,Y_0)$$

Markov inequality ☆

$$\le \mathbb{E}\left[d(X_t,Y_t)\mid X_0,Y_0\right]$$

$$\le \left(1-\frac{q-4\Delta}{nq}\right)^t\cdot n \le \varepsilon$$

$\downarrow \left(1+\frac{1}{x}\right)^x \to e,\ x\to\infty$

取 $t=\left\lceil \frac{q}{q-4\Delta}\cdot n\log\frac{n}{\varepsilon}\right\rceil$ 满足上述不等式.

by MC Coupling Lemma, 混合时间上界 $\tau(\varepsilon)\le \left\lceil \frac{q}{q-4\Delta}\cdot n\log\frac{n}{\varepsilon}\right\rceil$.

下面证明: $\mathbb{E}\left[d(X',Y')\mid X,Y\right]\le(1-\delta)\cdot d(X,Y)$

$X\to X'$	$Y\to Y'$	$d(X,Y)$
变	变	?
不变	变	?
变	不变	?
不变	不变	0

→ 取决于具体的随机选择 $(v,c)\in V\times[q]$

$d(X,Y)$ 每轮: (共享随机性)

-1: 减小1 $\boxed{好(v,c)}$ 当 $X_v\ne Y_v$ 且 c 和 [v 的邻居在 X,Y 的颜色] 不冲突

$d(X',Y')=d(X,Y)-1$ 至少 $d(X,Y)\cdot(q-2\Delta)$ 种

$+1$: 增加1 $\boxed{坏(v,c)}$ 当 $X_v=Y_v$ 且 c 在 X 或 Y 删的一个邻居中与 v 的某个邻居颜色冲突 (只在一个转移中更新)

$\Rightarrow d(X',Y')=d(X,Y)+1$ 至多 $2\cdot d(X,Y)\cdot\Delta$ 种

$\Rightarrow$ v 必须是某个在 (X,Y) 中颜色不一致的点 u 的邻居, 且 $c\in\{X_u,Y_u\}$

⓪. 不变 → 其他情况

每种随机选择等概率 $\frac{1}{nq}$ 被采样. 因此:

$$\mathbb{E}[d(X',Y')-d(X,Y)\mid X,Y] \le \left(-d(X,Y)\cdot(q-2\Delta)+2d(X,Y)\cdot\Delta\right)\cdot\frac{1}{nq}$$

$$= -\frac{d(X,Y)(q-4\Delta)}{nq}$$

$$\Rightarrow \mathbb{E}[d(X',Y')\mid X,Y] \le \left(1-\frac{(q-4\Delta)}{nq}\right)\cdot d(X,Y) \qquad \square$$

未必一定要使用相同的随机性!

· 考虑如下更加复杂的耦合 $(X,Y)\to(X',Y')$:

☆ < 合法耦合 距离衰减

· 使用相同的均匀分布的随机顶点 $v\in V$.

· 按如下方式刷新两条链中的随机颜色 $(C_X,C_Y)\in[q]\times[q]$:

· 若 $X_v\neq Y_v$, 则 $C_X=C_Y$ 为 $[q]$ 中均匀分布的随机颜色

· 若 $X_v=Y_v$, 令 $S_v^X=\{X_u \mid (u,v)\in E\}$ 和 $S_v^Y=\{Y_u\mid (u,v)\in E\}$

分别表示 v 的邻居在 X 和 Y 中使用的颜色的集合.

记 $S_v^X\setminus S_v^Y=\{a_1,\cdots,a_k\}$, $S_v^Y\setminus S_v^X=\{b_1,\cdots,b_\ell\}$

X邻居用的颜色 但Y的邻居没用

构造 $[q]$ 上的双射 ϕ: $\phi(c)=\begin{cases} b_i, & \text{if } c=a_i \wedge i\le\min\{k,\ell\} \\ c, & \text{otherwise}\end{cases}$ $\forall c\in[q]$

令 $\boxed{C_X}$ 为 $[q]$ 中均匀分布的随机颜色, $C_Y=\phi(C_X)$.

· 如果染色 (C_X,C_Y) 为分别在X Y链中合法的(即和邻居不冲突), 则进行MC的转移

对于图 Y 顶点 v, 记 $A_v=\{u\mid (u,v)\in E,\ X_u=Y_u\}$ 颜色相同的邻居

$D_v=\{u\mid (u,v)\in E,\ X_u\neq Y_u\}$ 颜色不同的邻居

· 好 (v,C_X): $X_v\neq Y_v \Rightarrow X'_v=Y'_v$ 至少 $\sum_{v:X_v\neq Y_v}\left(q-(2\Delta-|A_v|)\right)$ 种

• 坏 (v, C_X): $X_v = Y_v \Rightarrow X_v' \neq Y_v'$ 至多 $\sum_{v: X_v = Y_v} |D_v| = \sum_{v: X_v \neq Y_v} |A_v|$ 种

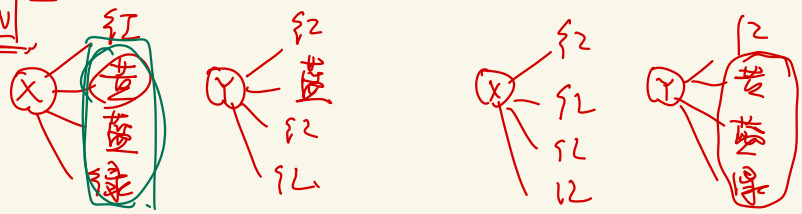
C_X:
- v 在 X 的邻居有
  - ① X 有 — Y 有 ⇒ 冲突 ⇒ 不更新 ⇒ d 不变
  - ② X 无 — Y 无
  - X 无 — Y 有
  - Y 无 ⇒ 不冲突 ⇒ 更新 ⇒ d 不变

① X 冲突: $C_X = a_i$ $|D_v|$
- $i \le l$: $C_Y = b_i$ ⇒ Y 冲突 ⇒ 都不更新 ⇒ d 不变
- $i > l$: $C_Y = C_X$ ⇒ Y 不冲突 ⇒ 更新 Y ⇒ d += 1

OR

② X 不冲突: $C_Y = C_X$ Y 有可能冲突 ⇒ 更新 X 不更新 Y ⇒ d += 1.

$|D_v|$



$\sum_{v: X_v = Y_v} |D_v| = \sum_{v: X_v \neq Y_v} |A_v|$

(u,v)
- $X_u = Y_u$ 且 $X_v = Y_v$ ⇒ $u \in A_v$. $v \in A_u$
- $X_u \neq Y_u$ 且 $X_v \neq Y_v$ ⇒ $u \in D_v$. $v \in D_u$
- "混合边": 假设 $X_u = Y_u$ 且 $X_v \neq Y_v$ ⇒ $v \in D_u$, $u \in A_v$

贡献相同

$$\mathbb{E}[d'(X,Y) - d(X,Y) \mid X, Y] \le -\frac{1}{nq} \sum_{v: X_v \neq Y_v} \left(q - (2\Delta - |A_v|)\right) + \frac{1}{nq} \sum_{v: X_v \neq Y_v} |A_v|$$

$$\le -\frac{q - 2\Delta}{nq} \cdot d(X,Y)$$

$$\Rightarrow \mathbb{E}[d(X', Y') \mid X, Y] \le \left(1 - \frac{q - 2\Delta}{nq}\right) \cdot d(X,Y)$$

⇒ 当 $q \ge 2\Delta + 1$ 时, 图染色 Metropolis 链有 $\tau(\varepsilon) \le \left\lceil \frac{q}{q - 2\Delta} n \log \frac{n}{\varepsilon} \right\rceil$

6. 路径耦合

Ω 度量 (metric) $d: \Omega^2 \to \mathbb{R}_{\geq 0}$ $\begin{cases} \text{正定} \\ \text{对称} \\ \text{三角不等式} \end{cases}$ (eg Hamming distance)

$\mathbb{E}[d(X', Y') \mid X, Y] \leq (1-\delta)\, d(X, Y) \to$ 期望意义衰减.

Thm 令 d 为 Ω 上的一个度量 (metric). 且定义 $d_{max} \triangleq \max_{x,y \in \Omega} d(x,y)$. $d_{min} \triangleq \min_{\substack{x,y \in \Omega \\ x \neq y}} d(x,y)$

如果存在一个耦合 $(X, Y) \to (X', Y')$ 使得对于某个 $0 < \delta < 1$.

对任意 $X, Y \in \Omega$ 都有: $\boxed{\mathbb{E}[d(X', Y') \mid X, Y] \leq (1-\delta)\, d(X, Y)}$

则混合时间有上界 $\tau(\varepsilon) \leq \left\lceil \frac{1}{\delta} \log \frac{D}{\varepsilon} \right\rceil$. 其中 $D = d_{max} / d_{min}$

Proof: $\Pr(X_t \neq Y_t \mid X_0, Y_0) = \Pr(d(X_t, Y_t) \geq d_{min} \mid X_0, Y_0)$

$$\leq \mathbb{E}[d(X_t, Y_t) \mid X_0, Y_0] / d_{min}$$ Markov inequality

$$= \frac{1}{d_{min}} \cdot \mathbb{E}[\mathbb{E}[d(X_t, Y_t) \mid X_{t-1}, Y_{t-1}, X_0, Y_0] \mid X_0, Y_0]$$

$$= \frac{1}{d_{min}} \mathbb{E}[\mathbb{E}[d(X_t, Y_t) \mid X_{t-1}, Y_{t-1}] \mid X_0, Y_0]$$ Markov property

$$\leq \frac{1}{d_{min}} \cdot (1-\delta) \cdot \mathbb{E}[d(X_{t-1}, Y_{t-1}) \mid X_0, Y_0]$$

$$\leq \frac{1}{d_{min}} \cdot (1-\delta)^t \cdot d(X_0, Y_0) \leq D(1-\delta)^t \leq \varepsilon$$

$(1+\frac{1}{x})^x < e$. $x > 0$

令 $t = \left\lceil \frac{1}{\delta} \log \frac{D}{\varepsilon} \right\rceil$ 满足上式 则由 coupling lemma 即得 □

def (pre-metric) 状态空间 Ω 上的预度量 是以 Ω 为顶点集合的边带正权的连通无向图. 且任何一条边 (x,y) 都是 x 和 y 之间的最短路径

i.e. x 和 y 之间任意路径的长度 (边的权重之和) 都至少为边 (x,y) 的权重.

x 和 y 相邻 iff $(x,y) \in E$.

度量 ⇒ 只在相邻而定义了距离 （一步耦合）

度量（Ω上）：$\forall x,y\in\Omega$，$d(x,y)$ 为 x 和 y 之间最短路径长度（耦合链）

路径耦合（Path Coupling）

Thm（Bubley & Dyer）给定 Ω 上的一个度量，如果存在一个只对相邻 $X,Y\in\Omega$ 定义的耦合 $(X,Y)\to(X',Y')$，使得对于某个 $0<\delta<1$，对任意相邻 $X,Y\in\Omega$：

一步耦合

$$\mathbb{E}[d(X',Y')\mid X,Y]\le(1-\delta)\,d(X,Y)$$

则该耦合 $(X,Y)\to(X',Y')$ 可拓展到任意 $X,Y\in\Omega$ 且仍有上述不等式成立。

（不需限制 X,Y 相邻）

Proof：X,Y 为任意一对初始状态

$X\leftrightsquigarrow Y$ 最短路径：$X=Z_0\to Z_1\to\cdots\to Z_k=Y$

$$d(X,Y)=\sum_{i=0}^{k-1}d(Z_i,Z_{i+1})$$

$$d(X',Y')\le\sum_{i=0}^{k-1}d(Z_i',Z_{i+1}')$$

有一对相邻的状态 (Z_i,Z_{i+1})，构造耦合 $(Z_i,Z_{i+1})\to(Z_i',Z_{i+1}')$

- 利用耦合 $(Z_0,Z_1)\to(Z_0',Z_1')$ 根据当前 (Z_0,Z_1) 产生 (Z_0',Z_1')，
- $i=1,\cdots,k-1$，依次在当前已生成 Z_i' 的条件下，根据耦合 $(Z_i,Z_{i+1})\to(Z_i',Z_{i+1}')$，根据当前 (Z_i,Z_{i+1}) 和 Z_i' 产生 Z_{i+1}'

⇒ 合作的耦合 $(X,Y)\to(X',Y')$

且 $$\mathbb{E}[d(X',Y')\mid X,Y]\le\mathbb{E}\Big[\sum_{i=0}^{k-1}d(Z_i',Z_{i+1}')\mid X,Y\Big]$$

$$=\sum_{i=0}^{k-1}\mathbb{E}[d(Z_i',Z_{i+1}')\mid Z_i,Z_{i+1}]\qquad\text{（三角不等式）}$$

$$\le(1-\delta)\cdot\sum_{i=0}^{k-1}d(Z_i,Z_{i+1})=(1-\delta)\,d(X,Y)\ \text{（最短路）}\ \square$$

Cor 若MC链满足Bubley-Dyer定理的条件，则混合时间 $\tau(\varepsilon)\le\left\lceil\frac{1}{\delta}\log\frac{D}{\varepsilon}\right\rceil$.

Example. 图染色采样的路径耦合

对路径耦合，只需为相邻的染色 X, Y 构造耦合 $(X,Y)\to(X',Y')$ ☆

Proposition. 当 $q\ge 2\Delta+1$ 时，图染色链有 $\tau(\varepsilon)\le\left\lceil\frac{q}{q-2\Delta}\cdot n\cdot\log\frac{n}{\varepsilon}\right\rceil$

Proof：对于任意一对相邻的 q-染色 X 和 Y，假设 X 和 Y 只在 v_0 上取不同颜色，我们构造耦合 $(X,Y)\to(X',Y')$ 如下：

- 使用相同的均匀分布的随机顶点 $v\in V$.
- 按如下方式分别生成两条链中的随机颜色 $(C_X,C_Y)\in[q]\times[q]$：
  - 如果 v 和 v_0 相邻，构造 $[q]$ 上的双射 ϕ：

$$\phi(c)=\begin{cases} Y_{v_0}, & \text{if } c=X_{v_0} \\ X_{v_0}, & \text{if } c=Y_{v_0} \\ c, & \text{otherwise}\end{cases}\qquad \forall c\in[q]$$

→ X 冲突，则 Y 冲突

→ dangerous!

令 C_X 为 $[q]$ 中均匀分布的随机颜色，$C_Y=\phi(C_X)$.

  - 其余情况，$C_X=C_Y$ 为 $[q]$ 中均匀分布的随机颜色. ← v 和 v_0 不相邻
- 如果染色 (C_X,C_Y) 为合法的（即和邻居不冲突），则进行MC的转移

好 (v,C_X)：颜色不同 → 颜色相同　　$\underbrace{v=v_0}_{1}$ 且 $\underbrace{C_X=C_Y}_{q-\Delta}$ 且 降成功

至少 $1\cdot(q-\Delta)$ 种

坏 (v,C_X)：颜色相同 → 颜色不同　　$\underbrace{v\text{是}v_0\text{的邻居}}_{\Delta}$ 且 $\underbrace{C_X=Y_{v_0}}_{1}\Rightarrow C_Y=X_{v_0}$

至多 $\Delta\cdot 1$ 种

每个选择仍以 $\frac{1}{nq}$ 的概率采样

$$\Rightarrow \mathbb{E}[d(X',Y')-d(X,Y)\mid X,Y]\le-\frac{q-\Delta}{nq}+\frac{\Delta}{nq}=-\frac{q-2\Delta}{nq}\cdot\underbrace{d(X,Y)}_{\ge 1}$$

$$\Rightarrow \mathbb{E}[d(X',Y')\mid X,Y] \le \left(1-\frac{q-2\Delta}{nq}\right) d(X,Y)$$

$$\Rightarrow \text{混合时间的上界 } \tau(\varepsilon) \le \left\lceil \frac{q}{q-2\Delta}\cdot n\log\frac{n}{\varepsilon}\right\rceil \qquad \square$$

Remark: Ω: (合法或非法) q-colorings $[q]^V$

在 $q \ge \Delta+2$ 条件下, 吸收(absorbing)

稳态分布, 所有合法的 q-colorings 的均匀分布

irreducibility $\begin{cases} \text{disconnected} \times \\ \text{absorbing} \checkmark \end{cases}$

两个合法的 q-colorings X, Y 之间的最短路径未必是它们之间的 Hamming 距离

Dobrushin 条件与快速混合

def (single-site dynamics) 令 V 为 n 个变量的集合. $[q]$ 为单个变量的定义域.

状态空间 $\Omega=[q]^V$ 上的单点动态系统为一抽象马尔可夫链为:

- 每一步, 对当前的状态 $X\in[q]^V$ 操作如下:
  - 均匀分布地选择一个随机的 $v\in V$.
  - 依分布 p_v^X 采样一个 $c\in[q]$, 并将 X_v 赋为 c.

eg.
- Metropolis 链
- Gibbs Sampler

注: 这里的分布 p_v^X 为 $[q]$ 上的一个概率分布. 其定义依赖 v 和 X.

posterior

影响矩阵 $\rho=(\rho_{uv})_{u,v\in V}$. $\rho_{uv}=\boxed{\max_{(\sigma,\tau)\in B_u}\|p_v^\sigma-p_v^\tau\|_{TV}}>0$

$$B_u=\{(\sigma,\tau)\in[q]^V\times[q]^V \mid \sigma_u\ne\tau_u \wedge \forall w\in V\setminus\{u\}, \sigma_w=\tau_w\}$$

只在 u 这一点不同的状态集合.

Thm (Dobrushin-Shlosman) 对于一个单点动态系统，如果存在 $0<\delta<1$，使得 $\|\rho\|_\infty = \max_{u\in V}\sum_{v\in V}\rho_{uv} \le 1-\delta$. 则有混合时间 $T(\varepsilon)\le\left\lceil \frac{n}{\delta}\log\frac{n}{\varepsilon}\right\rceil$

Proof: $X, Y\in[q]^V$ 只在某个点 v_0 上不同.

构造耦合 $(X,Y)\to(X',Y')$:

- 选择相同而均匀分布而随机顶点 $v\in V$.
- 按 P_v^X 和 P_v^Y 的最优耦合来采样随机颜色 $(C_X, C_Y)\in[q]\times[q]$.

  i.e. $\Pr(C_X\ne C_Y) = \|P_v^X - P_v^Y\|_{TV}$

- $X_v\to C_X$: $X\Rightarrow X'$

  $Y_v\to C_Y$: $Y\Rightarrow Y'$

Hamming 距离: $d(x,y)=\left|\{v\in V: x_v\ne y_v\}\right|$

$d(X,Y)=1$

$\downarrow$ 耦合

$$d(X',Y')=\begin{cases}0 & (\text{好事件})\\ 2 & (\text{坏事件})\\ 1.\end{cases}$$

好事件: $v=v_0$, $C_X=C_Y$

w.p. $\frac{1}{n}\left(1-\|P_v^X-P_v^Y\|_{TV}\right)$

坏事件: $v\ne v_0$, $C_X\ne C_Y$

w.p. $\frac{1}{n}\sum_{v\ne v_0}\|P_v^X-P_v^Y\|_{TV}$

$$\Rightarrow \mathbb{E}[d(X',Y')-d(X,Y)\mid X,Y]\le -\frac{1}{n}\left(1-\|P_v^X-P_v^Y\|_{TV}\right)+\frac{1}{n}\sum_{v\ne v_0}\|P_v^X-P_v^Y\|_{TV}$$

$$=-\frac{1}{n}\left(1-\sum_{v\in V}\|P_v^X-P_v^Y\|_{TV}\right)\cdot d(X,Y)$$

$$\Rightarrow \mathbb{E}[d(X',Y')\mid X,Y]\le\left(1-\frac{1}{n}\left(1-\sum_{v\in V}\|P_v^X-P_v^Y\|_{TV}\right)\right)d(X,Y)$$

$$\|\rho\|_\infty = \max_{u\in V} \sum_{v\in V} \rho_{uv} = \max_{u\in V} \sum_{v\in V} \max_{\sigma,\tau\in B_u} \|P_u^\sigma - P_v^\tau\|_{TV}$$

$$\geq \underline{\underline{\sum_{v\in V} \|P_v^X - P_v^Y\|_{TV}}} \geq 1-\delta \quad (X\text{和}Y\text{只在}v_0\text{一点不同})$$

$$\Rightarrow \mathbb{E}[d(X',Y') \mid X,Y] \leq \left(1-\frac{\delta}{n}\right)\cdot d(X,Y)$$

$$\Rightarrow \tau(\varepsilon) \leq \left\lceil \frac{n}{\delta}\log\frac{n}{\varepsilon} \right\rceil \quad \text{by Path coupling} \quad \square$$