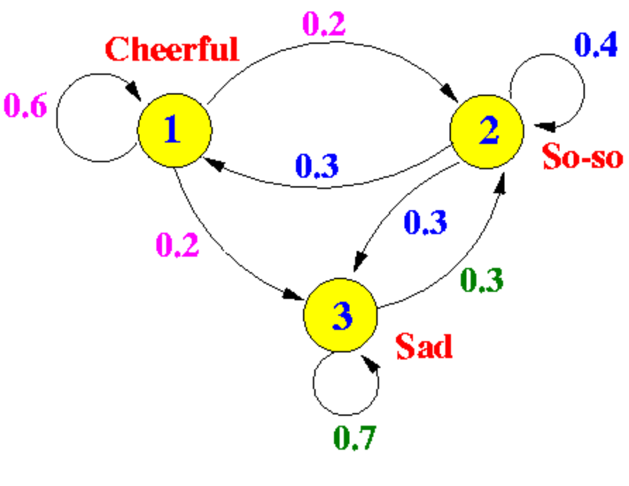
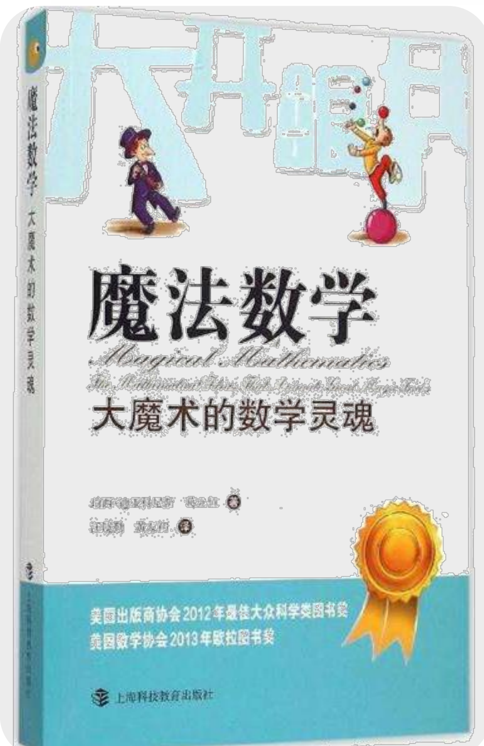
Random Walks and Markov Chains

longhuan@sjtu.edu.cn





- 戴康尼斯（Persi Diaconis，1945年1月31日 - ）：美国数学家、统计学家。斯坦福大学的数学与统计学教授。

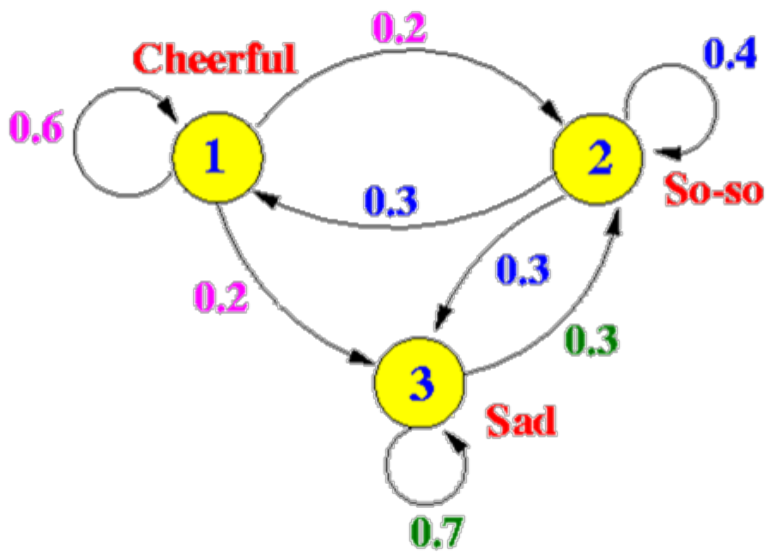
- 他解决了一些随机性的问题，包括掷币和洗牌。1992年，他和David Bayer证明完美的洗牌至少要洗七次。他又和说明从高处跌下的猫为何总能以脚着地的Richard Montgomery合作，证明了掷币哪面向上，物理因素比运气重要得多。

- 自14岁，他便跟随一个叫Dai Vernon的魔术师行走江湖。后来在赌场，他尝试研究防止他和其他魔术师被骗的方法。他18岁时买了一本An Introduction to Probability and Its Applications，但因为不懂微积分而看不明。24岁，他在City College of New York上数学课。其间他在《科学美国人》杂志投稿，介绍了他的两个纸牌戏法。马丁·葛登能认为那两个戏法十分精彩，注意到他的才华，为他写了一封推荐信。当时，哈佛大学的统计学家Fred Mosteller正在研究魔术，因此决定让Diaconis成为他的研究生。

——维基百科

Random walk

- **Random walk.** on a directed graph, a sequence of vertices generated from a start vertex by probabilistically selecting an incident edge, traveling the edge to a new vertex, and repeat the process.



Random walk

Probability distribution. $p = [p_1, p_2, \dots, p_n]$, where $\sum_{i=1}^{n} p_i$=1

Starting. $p = p(0) = [p_1(0), p_2(0), \dots, p_n(0)]$, $\sum_{i=1}^{n} p_i(0)$=1 and p_x is the probability of staring at x.

The probability of being at vertex x at time $t + 1$:

$$p_x(t+1) = \sum_{(y,x)\in E} p_y(t) \cdot \Pr(y \to x)$$

Transition Matrix P: P_{ij} is the probability of the walk at vertex i selecting the edge to vertex j.

$$p(t) \cdot P = p(t+1)$$

Random walk

Fundamental property. in the limit, the long-term average property of being at a particular vertex is *independent of the start vertex*, or an initial probability distribution over vertices (provided the underlying graph is strongly connected) – the *stationary probabilities*.

Markov chain

- A finite set S of states
- Transition probability: For $x, y \in S$, p_{xy} is the probability going from state x to y.
- $\sum_y p_{xy} = 1$

Markov chain➔ Random graph

① A vertex ➔ a state

② p_{xy}➔weighted edge from x to y.

Markov chain

Markov chain➔ Random graph

① A vertex ➔ a state

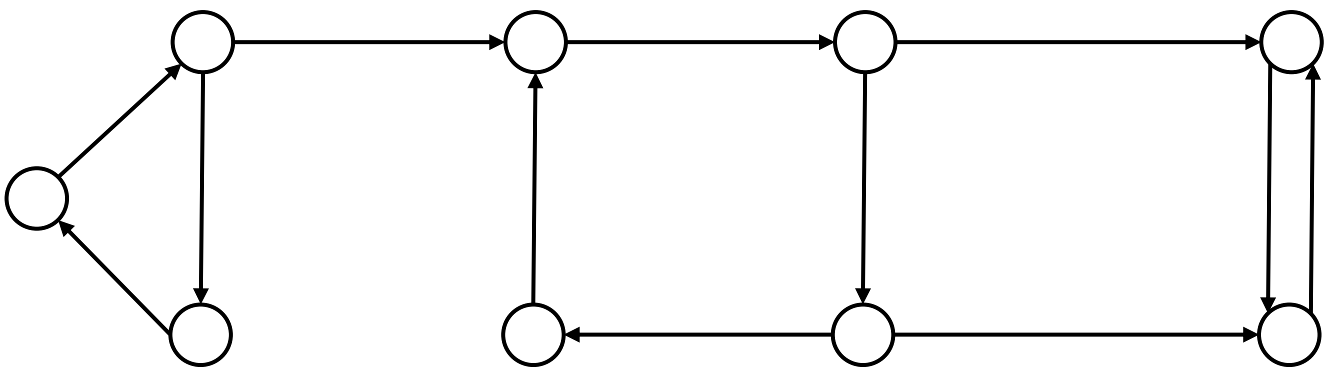
② p_{xy}➔weighted edge from x to y.

Connected Markov chain: if the underlying directed graph is strongly connected.

Transition probability matrix P: P_{xy} is the probability of changing from state x to y.

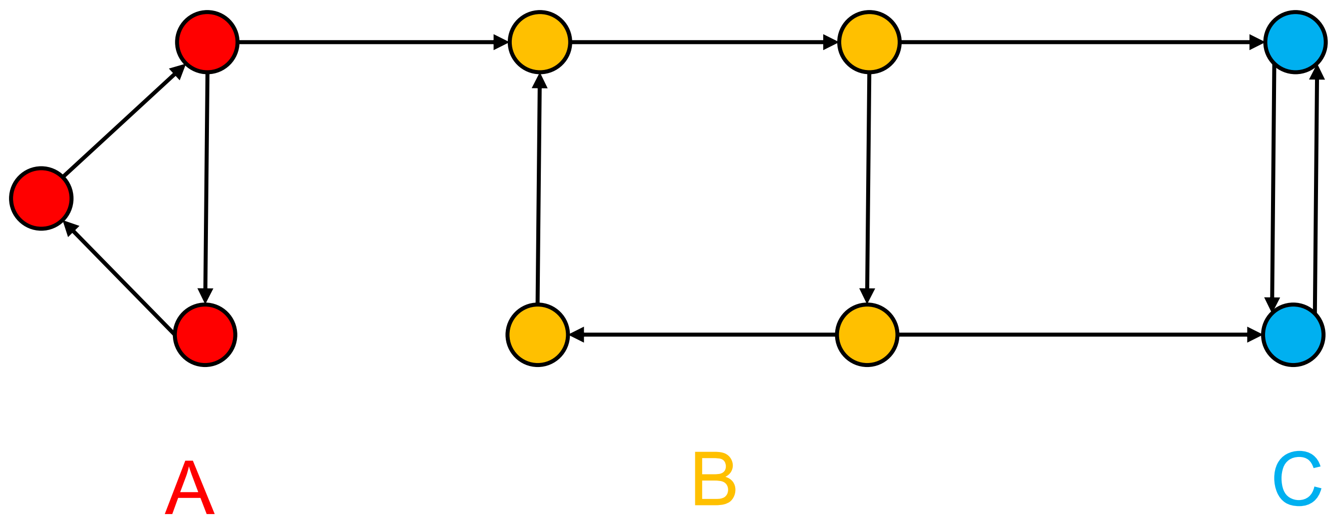
Markov chain

Persistent state. If the state ever be reached, the random process will return to it with probability 1.



Markov chain

Persistent state. If the state ever be reached, the random process will return to it with probability 1.



Markov chain

Aperiodic. If the greatest common divisor of the lengths of directed cycles in one.

Random walk	**Markov Chain**
Graph	Stochastic process
Vertex	State
Strongly connected	Persistent/Connected
Aperiodic	Aperiodic
Strongly connected +Aperiodic	Ergodic
Undirected graph	Time reversible

We will assume strongly connectness by default.

Stationary distribution

$p(t)$: the probability distribution after t steps of a random walk.

Long-term average probability distribution:

$$\boldsymbol{a(t)} = \frac{1}{t}(p(0) + p(1) + \cdots + p(t-1))$$

Fundatmental theorem of Markov chains:
For a connected MC, $\boldsymbol{a(t)}$ converges to a limit probability x which satisfies $x \cdot P = x$.

Fundamental Theorem

Lemma: Let P be the transition probability matrix for a connected Markov chain. The $n \times (n+1)$ matrix $A = [P - \boldsymbol{I}, \boldsymbol{1}]$ obtained by augmenting the matrix $P - \boldsymbol{I}$ with an additional column of ones has rank n.

Fundamental Theorem of Markov Chains: For a connected Markov chain there is a unique vector $\boldsymbol{\pi}$ satisfying $\boldsymbol{\pi} \cdot P = \boldsymbol{\pi}$. Moreover, for any starting distribution, $\lim_{t \to \infty} \boldsymbol{a}(\boldsymbol{t})$ exists and equals $\boldsymbol{\pi}$.

Lemma: For a random walk on a strongly connected graph with probabilities on the edge, if the vector $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi}_x p_{xy} = \pi_y p_{yx}$ for all x and y and $\sum_x \pi_x = 1$, then $\boldsymbol{\pi}$ is the stationary distribution of the walk.

Fundamental Theorem

Lemma: Let P be the transition probability matrix for a connected Markov chain. The $n \times (n+1)$ matrix $A = [P - \boldsymbol{I}, \mathbf{1}]$ obtained by augmenting the matrix $P - \boldsymbol{I}$ with an additional column of ones has rank n.

Fundamental Theorem

Fundamental Theorem of Markov Chains: For a connected Markov chain there is a unique vector $\boldsymbol{\pi}$ satisfying $\boldsymbol{\pi} \cdot P = \boldsymbol{\pi}$. Moreover, for any starting distribution, $\lim_{t \to \infty} \boldsymbol{a}(\boldsymbol{t})$ exists and equals $\boldsymbol{\pi}$.

Fundamental Theorem

Lemma: For a random walk on a strongly connected graph with probabilities on the edge, if the vector $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi}_x p_{xy} = \pi_y p_{yx}$ for all x and y and $\sum_x \pi_x = 1$, then $\boldsymbol{\pi}$ is the stationary distribution of the walk.

Markov Chain Monte Carlo

MCMC. A technique for sampling a multivariate probability distribution $p(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$.

Application. to estimate the expected value of a function $f(x)$

$$E(f) = \sum_{\boldsymbol{x}} f(\boldsymbol{x}) \cdot p(\boldsymbol{x})$$

Markov Chain Monte Carlo

Application. to estimate the expected value of a function $f(x)$

$$E(f) = \sum_{\boldsymbol{x}} f(\boldsymbol{x}) \cdot p(\boldsymbol{x})$$

Realization:

① Draw a set of samples. Each sample $\boldsymbol{x}$ is selected with probability $p(\boldsymbol{x})$.

② Averaging f over these samples.

Markov Chain Monte Carlo

Sample according to $p(\boldsymbol{x})$. Design a MC whose states correspond to the value space of $\boldsymbol{x}$ and whose stationary probability distribution is $p(\boldsymbol{x})$.

Recall:

- ✓ $p(\boldsymbol{t})$ is the row vector of probabilities of the random walk being at each state at time t.
- ✓ $\boldsymbol{a}(\boldsymbol{t}) = \frac{1}{t}(p(0) + p(1) + \cdots + p(t-1))$

$$E(r) = \sum_i f_i(\frac{1}{t}\sum_{j=1}^{t} \Pr(walk\ is\ in\ state\ i\ at\ time\ j)) = \sum_i f_i a_i(t)$$

Markov Chain Monte Carlo

Sample according to $p(\boldsymbol{x})$. Design a MC whose states correspond to the value space of $\boldsymbol{x}$ and whose *stationary probability distribution* is $p(\boldsymbol{x})$.

$$E(r) = \sum_i f_i\left(\frac{1}{t}\sum_{j=1}^{t} \Pr(walk\ is\ in\ state\ i\ at\ time\ j)\right) = \sum_i f_i a_i(t)$$

$$\left|\sum_i f_i p_i - E(r)\right| \le f_{max} \cdot \sum_i |p_i - a_i(t)|$$

$$= f_{max} \cdot \boxed{\|p - a(t)\|_1}$$

Markov Chain Monte Carlo

Sample according to $p(\boldsymbol{x})$. Design a MC whose states correspond to the value space of $\boldsymbol{x}$ and whose stationary probability distribution is $p(\boldsymbol{x})$.

Two general approach:

- The Metropolis-Hastings algorithm
- The Gibbs sampling

Metropolis-Hastings Algorithm

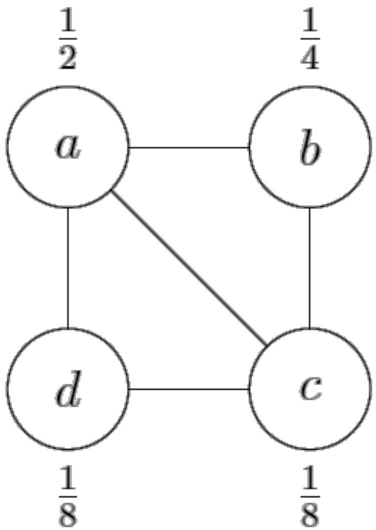
MHA. A general method to design a Markov chain whose stationary distribution is a given target distribution $\boldsymbol{p}$.

Given random graph G, with $\Delta(G) = r$. The transitions of the MC are defined as

$$\begin{cases} p_{ij} = \dfrac{1}{r}\min\left(1, \dfrac{p_j}{p_i}\right) \\ p_{ii} = 1 - \displaystyle\sum_{i \neq j} p_{ij} \end{cases}$$

Given random graph G, with $\Delta(G) = r$. The transitions of the MC are defined as $\begin{cases} p_{ij} = \frac{1}{r}\min\left(1, \frac{p_j}{p_i}\right) \\ p_{ii} = 1 - \sum_{i \neq j} p_{ij} \end{cases}$.

$p(a) = \frac{1}{2}$
$p(b) = \frac{1}{4}$
$p(c) = \frac{1}{8}$
$p(d) = \frac{1}{8}$



$a \to b \quad \frac{1}{3}\,\frac{1}{4}\,\frac{2}{1} = \frac{1}{6}$
$a \to c \quad \frac{1}{3}\,\frac{1}{8}\,\frac{2}{1} = \frac{1}{12}$
$a \to d \quad \frac{1}{3}\,\frac{1}{8}\,\frac{2}{1} = \frac{1}{12}$
$a \to a \quad 1 - \frac{1}{6} - \frac{1}{12} - \frac{1}{12} = \frac{2}{3}$

$b \to a \quad \frac{1}{3}$
$b \to c \quad \frac{1}{3}\,\frac{1}{8}\,\frac{4}{1} = \frac{1}{6}$
$b \to b \quad 1 - \frac{1}{3} - \frac{1}{6} = \frac{1}{2}$

$c \to a \quad \frac{1}{3}$
$c \to b \quad \frac{1}{3}$
$c \to d \quad \frac{1}{3}$
$c \to c \quad 1 - \frac{1}{3} - \frac{1}{3} - \frac{1}{3} = 0$

$d \to a \quad \frac{1}{3}$
$d \to c \quad \frac{1}{3}$
$d \to d \quad 1 - \frac{1}{3} - \frac{1}{3} = \frac{1}{3}$

$$\begin{aligned} p(a) &= p(a)p(a \to a) + p(b)p(b \to a) + p(c)p(c \to a) + p(d)p(d \to a) \\ &= \frac{1}{2}\,\frac{2}{3} + \frac{1}{4}\,\frac{1}{3} + \frac{1}{8}\,\frac{1}{3} + \frac{1}{8}\,\frac{1}{3} = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} p(b) &= p(a)p(a \to b) + p(b)p(b \to b) + p(c)p(c \to b) \\ &= \frac{1}{2}\,\frac{1}{6} + \frac{1}{4}\,\frac{1}{2} + \frac{1}{8}\,\frac{1}{3} = \frac{1}{4} \end{aligned}$$

$$\begin{aligned} p(c) &= p(a)p(a \to c) + p(b)p(b \to c) + p(c)p(c \to c) + p(d)p(d \to c) \\ &= \frac{1}{2}\,\frac{1}{12} + \frac{1}{4}\,\frac{1}{6} + \frac{1}{8}\,0 + \frac{1}{8}\,\frac{1}{3} = \frac{1}{8} \end{aligned}$$

$$\begin{aligned} p(d) &= p(a)p(a \to d) + p(c)p(c \to d) + p(d)p(d \to d) \\ &= \frac{1}{2}\,\frac{1}{12} + \frac{1}{8}\,\frac{1}{3} + \frac{1}{8}\,\frac{1}{3} = \frac{1}{8} \end{aligned}$$

Metropolis-Hastings Algorithm

Given random graph G, with $\Delta(G) = r$. The transitions of the MC are defined as $\begin{cases} p_{ij} = \frac{1}{r}\min\left(1, \frac{p_j}{p_i}\right) \\ p_{ii} = 1 - \sum_{i \neq j} p_{ij} \end{cases}$

Correctness.

To prove the stationary distribution is indeed the target distribution $\boldsymbol{p}$.

$$p_i p_{ij} = \frac{p_i}{r}\min\left(1, \frac{p_j}{p_i}\right) = \frac{1}{r}\min(p_i, p_j) = \frac{p_j}{r}\min\left(1, \frac{p_i}{p_j}\right) = p_j p_{ji}$$

Gibbs Sampling

Let $p(\boldsymbol{x})$ be the target distribution where $\boldsymbol{x} = (x_1, \dots, x_d)$. Now the undirected random graph is a hyper cube: there is an edge between $\boldsymbol{x}$ and $\boldsymbol{y}$ if $\boldsymbol{x}$ and $\boldsymbol{y}$ differ in only 1 coordinate.

Sampling process: for $\boldsymbol{x} = (x_1, \dots, x_d)$

① Choose one of the x_i to update;

② $x_i{}'$ is chosen based on the marginal probability of x_i

$$p_{xy} = \frac{1}{d} \cdot p(y_i \,|\, x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)$$

where $x_i \neq y_i$ and $x_j = y_j$ for all $i \neq j$,

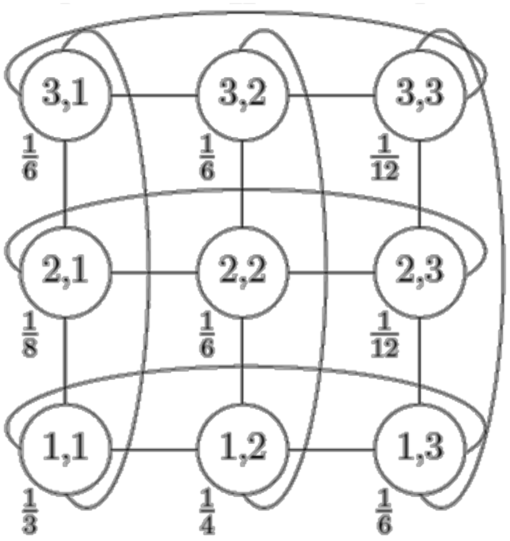
(i.e., $x_{i \neq j}$ does not change).

Sampling process: for $\boldsymbol{x} = (x_1, \dots, x_d)$

① Choose one of the x_i to update;

② $x_i{}'$ is chosen based on the marginal probability of x_i (i.e., $x_{i \neq j}$ will not change).

$p_{xy} = \frac{1}{d} \cdot p(y_i | \, x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)$, where $x_i \neq y_i$ and $x_j = y_j$ for all $i \neq j$.

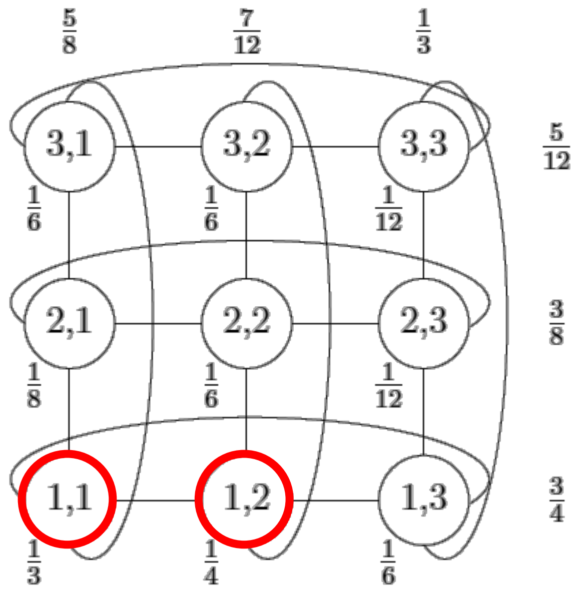


Sampling process: for $\boldsymbol{x} = (x_1, \dots, x_d)$

① Choose one of the x_i to update;

② $x_i{}'$ is chosen based on the marginal probability of x_i (i.e., $x_{i \neq j}$ will not change).

$p_{xy} = \frac{1}{d} \cdot p(y_i | \, x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)$, where $x_i \neq y_i$ and $x_j = y_j$ for all $i \neq j$.

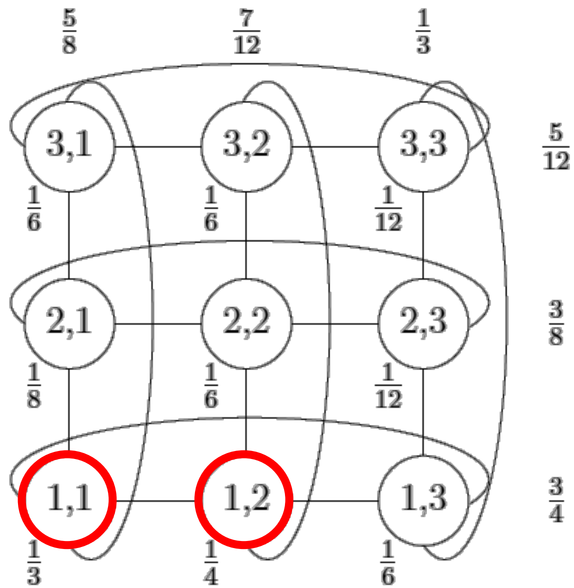


Sampling process: for $\boldsymbol{x} = (x_1, \dots, x_d)$

① Choose one of the x_i to update;

② $x_i{}'$ is chosen based on the marginal probability of x_i (i.e., $x_{i \neq j}$ will not change).

$$p_{xy} = \frac{1}{d} \cdot p(y_i \,|\, x_1 \cdots x_{i-1} x_{i+1} \cdots x_d), \text{ where } x_i \neq y_i \text{ and } x_j = y_j \text{ for all } i \neq j.$$



$$p_{(11)(12)} = \frac{1}{d} p_{12} / (p_{11} + p_{12} + p_{13}) = \frac{1}{2}\ (\frac{1}{4}) / (\frac{1}{3}\ \frac{1}{4}\ \frac{1}{6}) = \frac{1}{8} / \frac{9}{12} = \frac{1}{8}\ \frac{4}{3} = \frac{1}{6}$$

$$p_{(11)(12)} = \frac{1}{2}\ \frac{1}{4}\ \frac{4}{3} = \frac{1}{6} \quad p_{(12)(11)} = \frac{1}{2}\ \frac{1}{3}\ \frac{4}{3} = \frac{2}{9} \quad p_{(13)(11)} = \frac{1}{2}\ \frac{1}{3}\ \frac{4}{3} = \frac{2}{9} \quad p_{(21)(22)} = \frac{1}{2} \frac{1}{6} \frac{8}{3} = \frac{2}{9}$$

$$p_{(11)(13)} = \frac{1}{2}\ \frac{1}{6}\ \frac{4}{3} = \frac{1}{9} \quad p_{(12)(13)} = \frac{1}{2}\ \frac{1}{6}\ \frac{4}{3} = \frac{1}{9} \quad p_{(13)(12)} = \frac{1}{2}\ \frac{1}{4}\ \frac{4}{3} = \frac{1}{6} \quad p_{(21)(23)} = \frac{1}{2} \frac{1}{12} \frac{8}{3} = \frac{1}{9}$$

$$p_{(11)(21)} = \frac{1}{2}\ \frac{1}{8}\ \frac{8}{5} = \frac{1}{10} \quad p_{(12)(22)} = \frac{1}{2}\ \frac{1}{6}\ \frac{12}{7} = \frac{1}{7} \quad p_{(13)(23)} = \frac{1}{2}\ \frac{1}{12}\ \frac{3}{1} = \frac{1}{8} \quad p_{(21)(11)} = \frac{1}{2} \frac{1}{3} \frac{8}{5} = \frac{4}{15}$$

$$p_{(11)(31)} = \frac{1}{2}\ \frac{1}{6}\ \frac{8}{5} = \frac{2}{15} \quad p_{(12)(32)} = \frac{1}{2}\ \frac{1}{6}\ \frac{12}{7} = \frac{1}{7} \quad p_{(13)(33)} = \frac{1}{2}\ \frac{1}{12}\ \frac{3}{1} = \frac{1}{8} \quad p_{(21)(31)} = \frac{1}{2} \frac{1}{6} \frac{8}{5} = \frac{2}{15}$$

Gibbs Sampling

Sampling process: for $\boldsymbol{x} = (x_1, \dots, x_d)$

① Choose one of the x_i to update;

② $x_i{}'$ is chosen based on the marginal probability of x_i (i.e., $x_{i \neq j}$ will not change). $p_{xy} = \frac{1}{d} \cdot p(y_i | x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)$, where $x_i \neq y_i$ and $x_j = y_j$ for all $i \neq j$.

Correctness. To prove the stationary distribution is indeed the target distribution $\boldsymbol{p}$.

$$p_{\boldsymbol{xy}} = \frac{1}{d} \frac{p(y_i | x_1 \cdots x_{i-1} x_{i+1} \cdots x_d) p(x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)}{p(x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)}$$

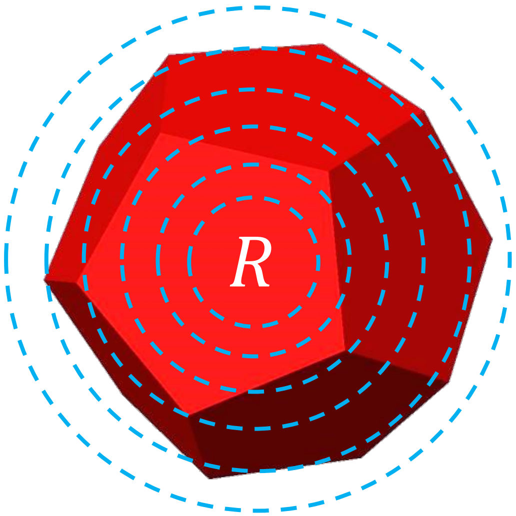
$$= \frac{1}{d} \frac{p(x_1 \cdots x_{i-1} y_i x_{i+1} \cdots x_d)}{p(x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)} = \frac{1}{d} \frac{p(\boldsymbol{y})}{p(x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)}$$

Similarly $p_{\boldsymbol{yx}} = \frac{1}{d} \frac{p(\boldsymbol{x})}{p(x_1 \cdots x_{i-1} x_{i+1} \cdots x_d)}$

It follows that $p(\boldsymbol{x}) p_{\boldsymbol{xy}} = p(\boldsymbol{y}) p_{\boldsymbol{yx}}$.

Areas and Volumes

For general convex sets in d space, there are no close form formulae for volume.

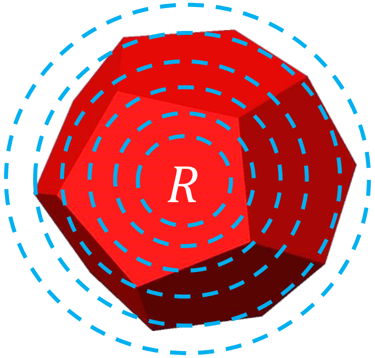


Sequence of concentric spheres:

$$R \supset S_1 \subseteq S_2 \subseteq \cdots \subseteq S_k \supset R$$

$$Vol(R) = Vol(S_k \cap R)$$
$$= \frac{Vol(S_k \cap R)}{Vol(S_{k-1} \cap R)} \cdot \frac{Vol(S_{k-1} \cap R)}{Vol(S_{k-2} \cap R)} \cdots \frac{Vol(S_2 \cap R)}{Vol(S_1 \cap R)} \cdot Vol(S_1)$$

Areas and Volumes



$$Vol(R) = \frac{Vol(S_k \cap R)}{Vol(S_{k-1} \cap R)} \cdot \frac{Vol(S_{k-1} \cap R)}{Vol(S_{k-2} \cap R)} \cdots \frac{Vol(S_2 \cap R)}{Vol(S_1 \cap R)} \cdot Vol(S_1)$$

$$Radius(S_i) = \left(1 + \frac{1}{d}\right) \cdot Radius(S_{i-1})$$

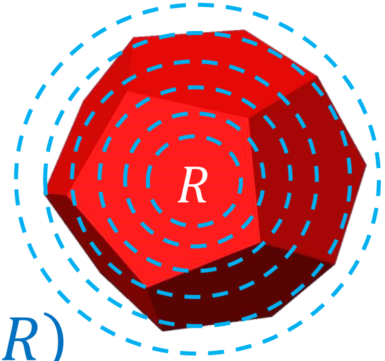
Thus $1 \le \frac{Vol(S_i \cap R)}{Vol(S_{i-1} \cap R)} = \left(1 + \frac{1}{d}\right)^d \le e$

Let $r = \left(1 + \frac{1}{d}\right)^k$ then the number of spheres k is at most

$$O\left(\log_{1+\frac{1}{d}} r\right) = O(d\ln(r))$$

To estimate the overall volume to error $1 \pm \epsilon$:
Estimate each volume ratio to a factor of $1 \pm \frac{\epsilon}{ed\ln(r)}$.

Areas and Volumes



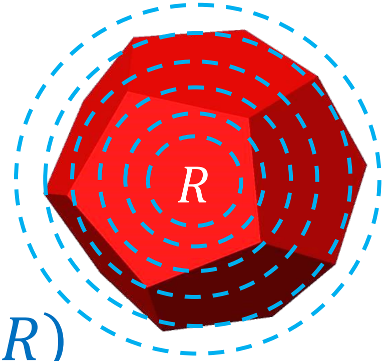
$$Radius(S_i) = \left(1 + \frac{1}{d}\right) \cdot Radius(S_{i+1}),\ 1 \le \frac{Vol(S_i \cap R)}{Vol(S_{i-1} \cap R)} \le e$$

Estimate the ratio $\frac{Vol(S_i \cap R)}{Vol(S_{i-1} \cap R)}$:

① Selecting points in $S_i \cap R$ uniformly at random;

② Computing the fraction in $S_{i-1} \cap R$.

Areas and Volumes

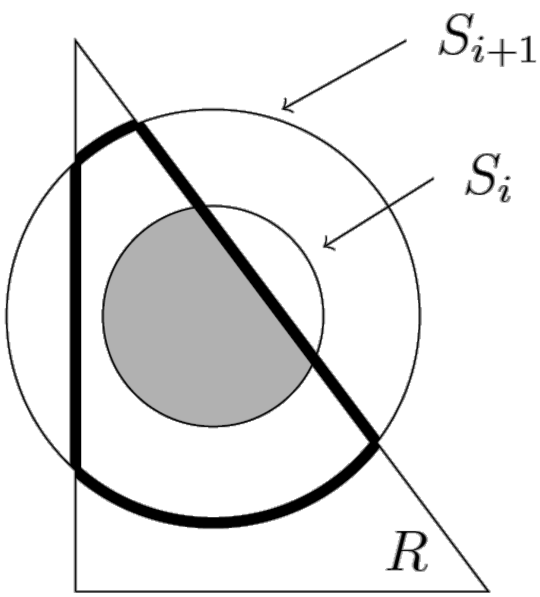


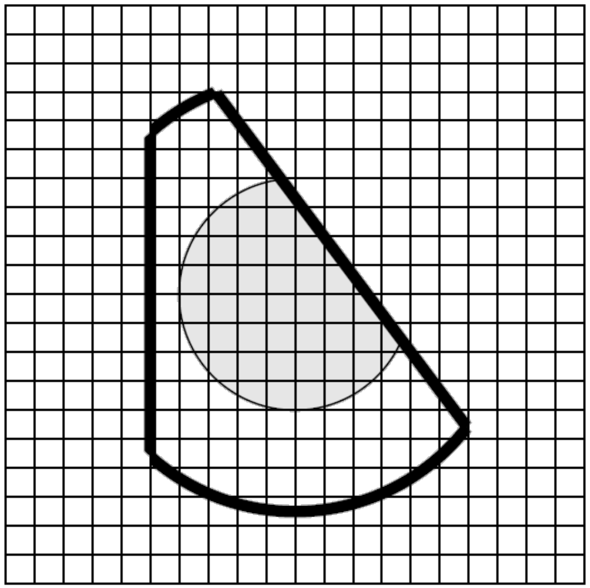
$$Radius(S_i) = \left(1 + \frac{1}{d}\right) \cdot Radius(S_{i+1}),\ \ 1 \leq \frac{Vol(S_i \cap R)}{Vol(S_{i-1} \cap R)} \leq e$$

Estimate the ratio $\frac{Vol(S_i \cap R)}{Vol(S_{i-1} \cap R)}$:

① Selecting points in $S_{i+1} \cap R$ uniformly at random;

② Computing the fraction in $S_i \cap R$.



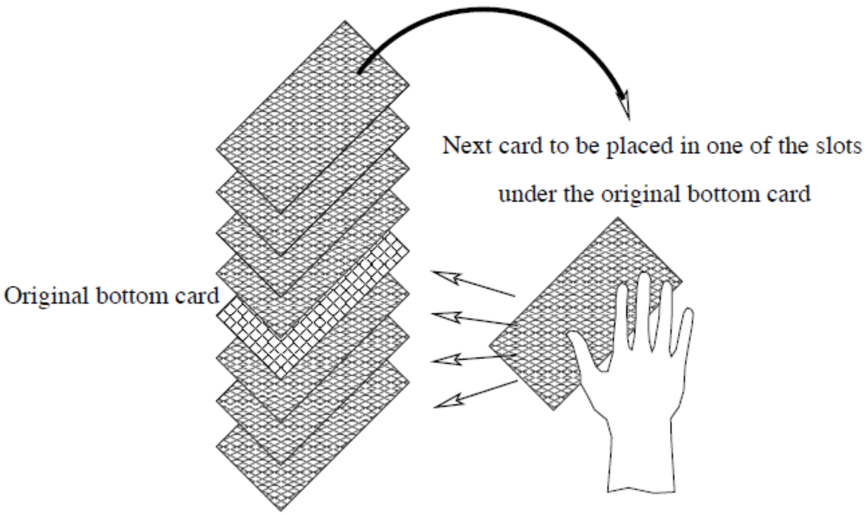


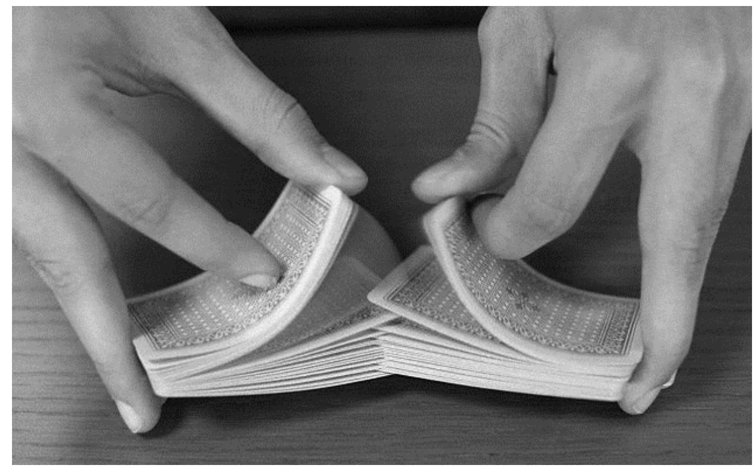
Some conceptions

Mixing time. Fix $\epsilon > 0$. The ϵ −mixing time of a MC is the minimum integer t such that for any starting distribution p_0, the 1-norm distance between the t −step running average probability distribution and the stationary distribution is at most ϵ.

Hitting time h_{xy}. The expected time of a random walk starting at vertex x (or a starting probability distribution) to reach vertex y.

Cover time. The expected time of a random walk starting at vertex x in the graph G to reach each vertex at least once.

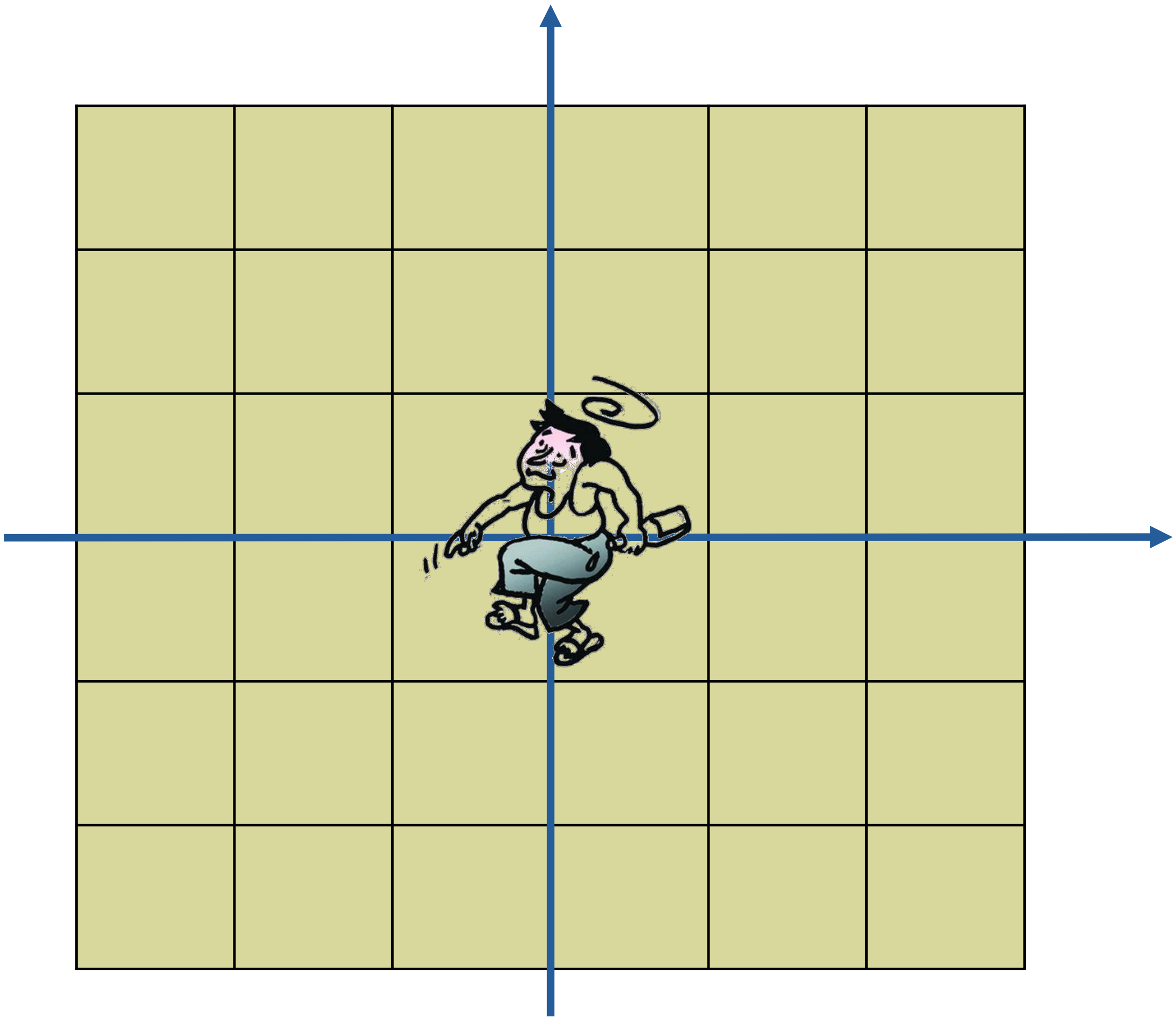

Next card to be placed in one of the slots
under the original bottom card
Original bottom card



$O(n \ln n)$

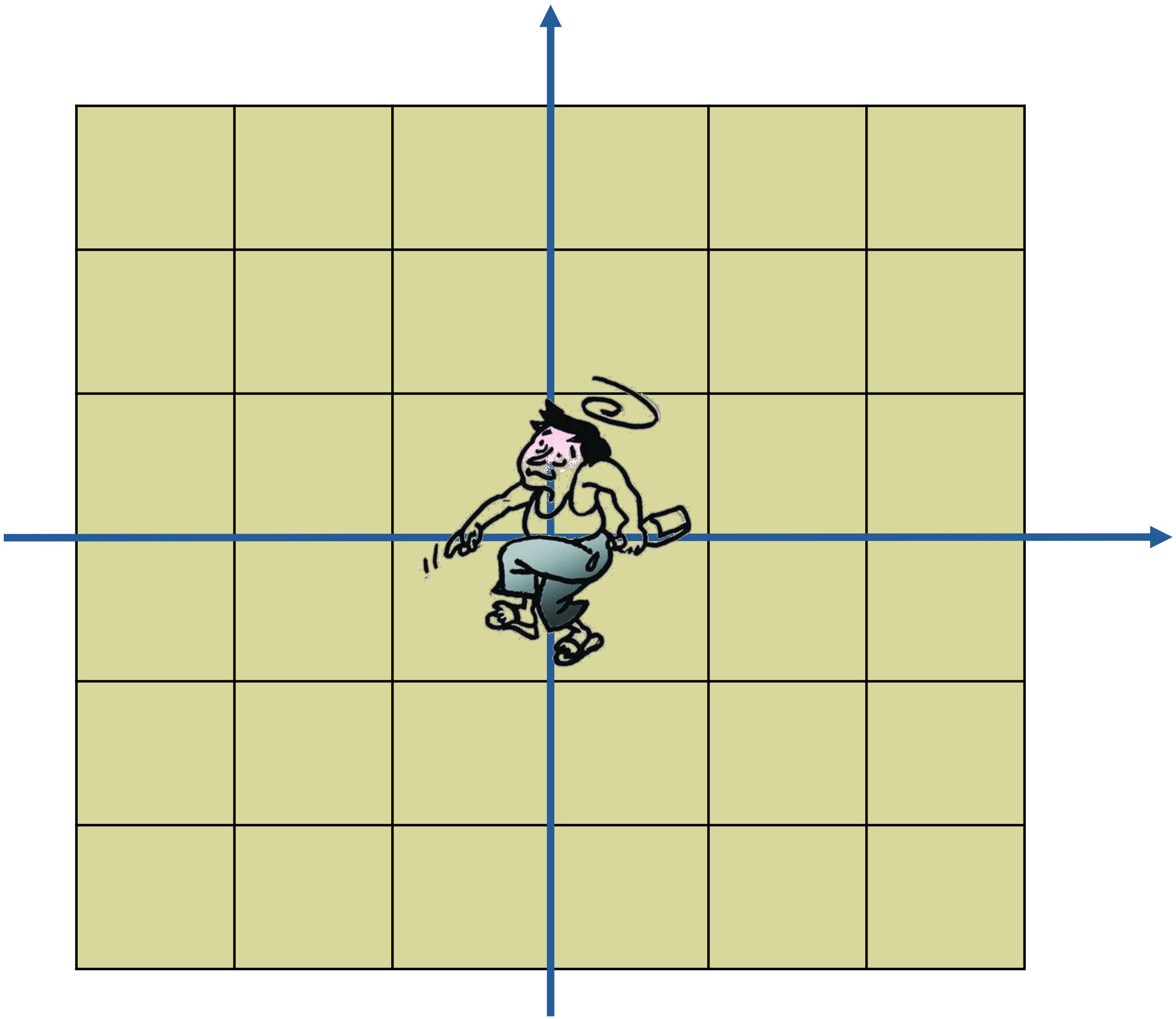
$O(\ln n)$

Random walks in Euclidean Space



George Pólya,1921

Random walks in Euclidean Space



Random walks in Euclidean Space

Random walks in Euclidean Space

- “A drunk person will always find their way home, while a drunk bird may get lost forever.”