Def A Markov chain ~~with limiting probs π_i~~ is called ~~time~~ reversible if

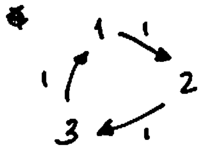
$$Q_{ij} = P_{ij} \quad \forall i, j.$$

~~Equivalently~~ This happens if (if lim prob's are π_i)

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \forall i, j. \qquad (*)$$

Examples:

(a)



$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

NOT reversible;

(b) "Most" of M. chains are not reversible (if you put arbitrary numbers).
– see "Kolmogorov's criterion" ~~in Wikipedia~~, ~~Ross~~ (Thm 4.2)

~~(c) Symmetric Simple random walk on Z is the reversible.~~

(c) $\forall$ M. chain with symmetric P is reversible:

Prop Suppose $P_{ij} = P_{ji}$ $\forall i, j$. on a finite irreducible M. chain.
Then $\pi_i = \frac{1}{n}$ $\forall i$
i.e. the stationary distr. is uniform,
and M. chain is reversible.

Proof To check that $\pi_i = \frac{1}{n}$, all we need to check (by Thm p. 124)
is that $\frac{1}{n} = \sum_{i=1}^{n} \frac{1}{n} P_{ij}$,
i.e. that $\sum_{i=1}^{n} P_{ij} = 1$. But this is true, since
$$\sum_{i=1}^{n} P_{ij} = \sum_{i=1}^{n} P_{ji} \quad \text{(by symmetry)}$$
$$= 1 \quad \text{(by def of } P_{ij}\text{} - \text{total prob).}$$

The M. chain is reversible since ~~$\frac{1}{n} P_{ij} = \frac{1}{n} P_{ji}$ by symmetry, so (*)~~
(*) holds: $\frac{1}{n} P_{ij} = \frac{1}{n} P_{ji}$ by symmetry.

Ex (d) M. chain from p. 125 (3×3) is reversible.

Also,

(e) $P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 2/3 & 0 & 1/3 \\ 1/3 & \frac{1}{3} & 1/3 \end{bmatrix}$ is reversible; limiting prob's are $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$.

"doubly stochastic matrix".

~~Prop (vice-versa) Suppose we can find numbers $\pi_i \geq 0$, $\sum_i \pi_i = 1$, and s.t.~~

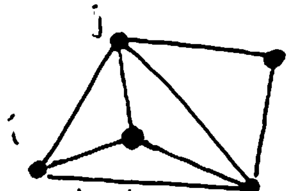
~~$\pi_i P_{ij} = \pi_j P_{ji}$ $\forall i, j$.~~

~~Then M. chain is reversible by~~

Application: random walks on graphs.

- Consider a simple, connected undirected graph

  no loops



- Random walk: move from vertex to vertex along edges, ~~to a randomly chosen neighbor~~).
  every time picking an edge at random (uniformly among available)

~~Irreducible, aperiodic M. chain. (why?)~~
~~Hence ∃ limit probabilities.~~
~~Let's compute them, using~~
~~Tra $P_{ij} =$~~

- Markov chain, with transition prob's

$$P_{ij} = \frac{w_{ij}}{\sum_j w_{ij}} \quad \text{where} \quad w_{ij} = \begin{cases} 1 & \text{if } \exists \text{ edge } i - j \\ 0, & \end{cases}$$

Irreducible, aperiodic (why?)

Thus ∃ limiting prob's π_i (by Thm p.124)

• Let's compute π_i using Thm p. 127.

If we find numbers $\pi_i \geq 0$, $\Sigma \pi_i = 1$ satisfying time reversibility equations

$$\pi_i P_{ij} = \pi_j P_{ji} \quad , \qquad \forall ij$$

then π_i are automatically the limiting prob's.

$$\pi_i \frac{w_{ij}}{\sum_j w_{ij}} = \pi_j \frac{w_{ji}}{\sum_i w_{ji}} \qquad \forall ij$$

Since $w_{ij} = w_{ji}$ by construction,

$$\frac{\pi_i}{\sum_j w_{ij}} = \frac{\pi_j}{\sum_i w_{ij}} \qquad \forall ij$$

Thus all these numbers are the same, $\forall i$!

ie.

$$\frac{\pi_i}{\sum_j w_{ij}} = c \quad (\text{const}) . \quad \forall i$$

• To calculate c;

$$\pi_i = c \sum_j w_{ij}$$

Sum over i,

$$1 = c \sum_{ij} w_{ij}$$

$$\Rightarrow \pi_i = \frac{\sum_j w_{ij}}{\sum_{ij} w_{ij}} .$$

$\sum_j w_{ij} = d_i$, the <u>degree</u> of i (# of neighbors)

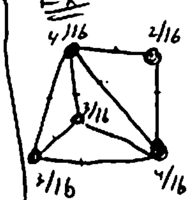
$\sum_{ij} w_{ij} = N$, total # of edges.

Thus: $\boxed{\pi_i \sim \text{degree of } i}$

THM Consider a ~~simple~~ random walk on ~~an undirected~~ a simple, undirected, connected graph. Then this M. chain is reversible; the limiting prob's are proportional to the degrees of vertices.

Ex

Ex:



Random walk on directed graphs:

Page Rank

= algorithm used by Google to rank the importance of websites, in their search engine results.

- Web graph: vertices = web pages,
  edges = links (directed!)

- Simplified Page Rank algorithm:
  - perform a random walk on the web graph (just like before, but in the direction of links)
  - ~~estimate limiting prob's by fraction~~
  - output limiting prob's = "rank" of pages.
    (compute them e.g. as the fraction of time spent).

- Problem: graph is directed ⇒ NOT irreducible!
  Walk will get stuck e.g. in a webpage with no outgoing links "sink"

- Solution: damping:
  - at any step, flip a coin that comes heads with prob. $p \simeq 0.85$.
  - If H ⇒ ~~move to~~ follow a random link (as before)
  - If T ⇒ jump to a randomly chosen webpage (from the entire internet).

  Makes M. chain irreducible, aperiodic ⇒ ∃ lim, prob π_i = "rank".

(Motivation for damping: "surfer gets bored".)

- Remark In contrast to r. walks on undirected graphs, Pagerank's (limiting prob's) are NOT proportional to simply # of links to/from a page:
  [links from an important page (with higher rank) increase the rank of your page fast]

Application: Markov Chain Monte Carlo (MCMC)

Recall Monte-Carlo method for approximating

$$\int_\Omega f(x)\,dx \quad \text{where } \Omega \subset \mathbb{R}^d.$$

~~Sample $X_1, X_2, \dots, X_N \sim \mathrm{Unif}(\Omega)$, then~~

Represent $\frac{1}{|\Omega|}\int_\Omega f(x)\,dx = \mathbb{E}[f(X)]$ where $X \sim \mathrm{Unif}(\Omega)$.

Sample $X_1, \dots, X_N \sim \mathrm{Unif}(\Omega)$ iid;

LLN $\Rightarrow$ $\frac{1}{N}\sum_1^N f(X_i) \longrightarrow \mathbb{E}[X]$ as $N \to \infty$.

compute.

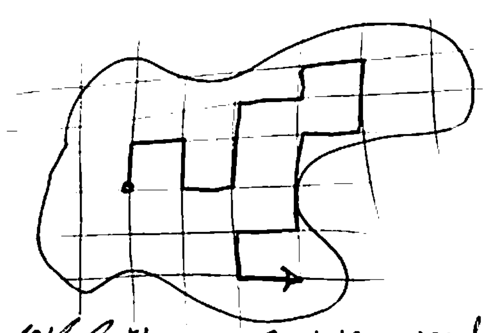
Difficulty. How to sample from Unif(Ω)?

(e.g. when Ω is given by "oracle": $x \in \Omega$ / $x \notin \Omega$).

could be hard!

Markov Chain approach: perform a random walk on Ω (computable!)

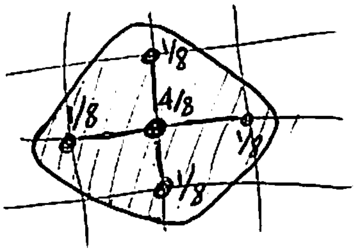
Stop after long time $\to X_1$, do again $\to X_2, \dots$



$\boxed{\text{MCMC.}}$

Question: We would sample from the stationary distr. π. Is it the same as Unif(Ω)?

Not always! e.g. for



Random walk on this graph has ← ~~pos~~ limiting prob's $\neq$ const! (by Thm p.130)

Boundary effect! It is a problem especially in high dim's, (will spend most time near boundary)

So, a general Question: How to sample from a given distribution?

More precisely,
~~General~~ MCMC Question. Consider a set S and some distribution π on S (think about S = finite set, $\pi = \text{Unif}(S)$).

Construct a Markov chain $X_0, X_1, X_2, \dots$ on state space S with limiting distribution π, e.g.

$$\lim_{n\to\infty} \mathbb{P}\{X_n = i\} = \pi_i .$$

Solution: <u>Hastings – Metropolis Algorithm</u>

~~Solution. Metropolis algorithm (1953)~~

~~Hastings (1970)~~

- Idea: choose a right acceptance/rejection policy.

- Start from some ~~known~~ Markov chain on S that is easy to implement (e.g. random walk as above).
  Suppose its trans. prob. matrix is Q.

- Modify Q s.t. the limiting distr. is π.
  For each i,j, build an acceptance probability a_{ij} (tbd).

  $$i \xrightarrow{a_{ij}} j$$

  Modify the transitions as follows:

  If the original ~~initial~~ M.C. proposes a transition from i to j,
  accept it with prob a_{ij} (and stay at i w/prob $1-a_{ij}$).

- The new M.C. obviously has trans. probabilities

$$P_{ij} = Q_{ij}\, a_{ij}, \qquad i \neq j$$

- In order that π is the ~~transit~~ limiting probability for this M.C, it suffices that the <u>time reversibility</u> condition holds.

$$\pi_i P_{ij} = \pi_j P_{ji} \qquad \forall\, i \neq j$$

$$(*) \qquad \iff \pi_i Q_{ij}\, a_{ij} = \pi_j Q_{ji}\, a_{ji} \qquad \forall\, i \neq j$$

- It's time to choose a_{ij} so that this ineq. is true. Metropolis choise is

$$\boxed{a_{ij} = \min\left(1, \frac{\pi_j Q_{ji}}{\pi_i Q_{ij}}\right)}$$

[Check (*)]!

- <u>Conclusion</u>. lim. prob $= \pi$. ~~Important remark: Works for not prob. measures, whose sums we don't know — it relies on ratio π_i/π_j only~~

We have proved:

THM (Hastings-Metropolis Algorithm)

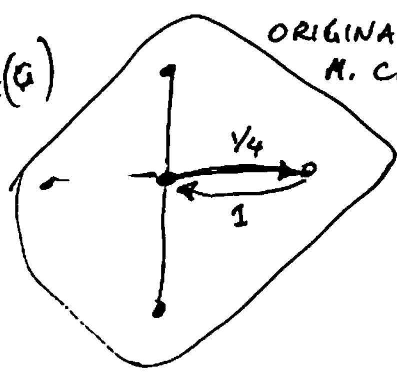
> Consider an irreducible, aperiodic Markov chain on a finite state space S, with transition prob's Q_{ij}.
> Let $\pi_i \geq 0$, $\sum_{i \in S} \pi_i = 1$.
> Construct a Markov chain that accepts transitions $i \to j$ with prob.
> $$a_{ij} := \min\left(1, \frac{\pi_j Q_{ji}}{\pi_i Q_{ij}}\right).$$
> (and stays at i with prob $1 - a_{ij}$).
> Then the new M. chain is reversible, and has lim. prob's π_i

Remark: H-M Algorithm works even if ~~π_i's are not~~ π IS NOT a prob. measure, i.e. $\sum_{i \in S} \pi_i \neq 1$ — since it depends on ratios π_j / π_i rather than actual values.

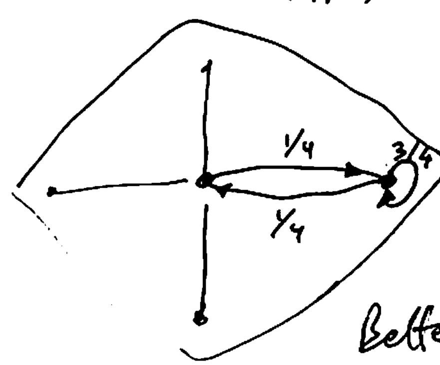
Example: To sample from $\mathrm{Unif}(\Omega)$, $\pi_i = 1$ IS OK (rather than $\frac{1}{|\Omega|}$, which could be unknown).

Example (G)

ORIGINAL M. CHAIN



NEW M. CHAIN:

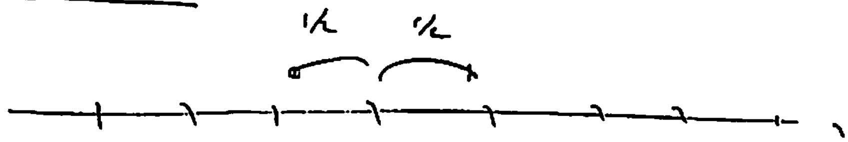


Better!

Examples (6)

Simulate a Poisson(λ) ~~r.v. with~~ r.v.

$S = \mathbb{Z}$



Original M.C: Simple random walk, $Q_{ij} = \frac{1}{2}$, $|i-j| = 1$.

$Q_{ij} = Q_{ji} \Rightarrow$

$$a_{ij} = \min\left(1, \frac{\pi_j}{\pi_i}\right).$$

Algorithm then:
- Proposes a transition $i \to j$,
- Is $\pi_j > \pi_i$?
  - yes $\Rightarrow$ move
  - no $\Rightarrow$ move w/prob $\frac{\pi_j}{\pi_i}$, otherwise stay.

In our case,

$$\frac{\pi_{i+1}}{\pi_i} = \frac{\lambda^{i+1}/(i+1)!}{\lambda^i/i!} = \frac{\lambda}{i+1},$$

$$\frac{\pi_{i-1}}{\pi_i} = \frac{\lambda^{i-1}/(i-1)!}{\lambda^i/i!} = \frac{i}{\lambda}.$$

$$\begin{cases} a_{i,i+1} = \min\left(1, \frac{\lambda}{i+1}\right) \\ a_{i,i-1} = \min\left(1, \frac{i}{\lambda}\right). \end{cases}$$

Interpretation: $i < \lambda \Rightarrow \frac{\pi_{i+1}}{\pi_i} \geq 1 \Rightarrow$ always accept moves to the right

$i > \lambda \Rightarrow \frac{\pi_{i-1}}{\pi_i} \geq 1 \Rightarrow$ always accept moves to the left

$i = \lambda \Rightarrow$ wander around λ.

Another solution to Q. p.133.

4.9. Gibbs sampler

We want to sample from a given ~~distribution $\pi(x_1,\dots,x_d) \in \mathbb{R}^d$~~ distribution with PMF or PDF $p(x_1,\dots,x_d)$ on $\mathbb{R}^d$.

Gibbs Sampler: "sample one coordinate at a time":

- Start with some vector $(x_1,\dots,x_d)$.
- Choose a coordinate i at random (uniformly among $1,\dots,n$)
- Fix all other coordinates x_j $(j \neq i)$ and sample the i'th coordinate conditionally on x_j $(j \neq i)$.
- Repeat

Example: To sample from $\mathrm{Unif}(\Omega)$,

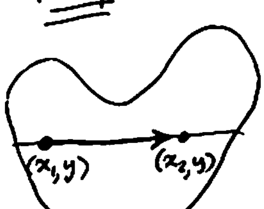
etc.

2nd coord. chosen

Random walk. Markov chain. Does ~~it~~ lim. prob. converge to $\mathrm{Unif}(\Omega)$?

THM (Gibbs sampler) Consider a ~~function~~ distribution $p(x_1,\dots x_d)$ supported on a finite subset of $\mathbb{Z}^d$ (= $\forall$ distr. in $\mathbb{R}^d$ is OK actually). Then ~~the M. chain~~ the limiting ~~prob's~~ distr in Gibbs sampler equals p. ~~converge to π.~~

Proof (for $\mathbb{R}^2$, for simplicity). Let's check the reversibility condition (Thm p.127):



$$\pi_{(x_1,y)}\, P_{(x_1,y)(x_2,y)} \overset{?}{=} \pi_{(x_2,y)}\, P_{(x_2,y)(x_1,y)}$$

$$p(x_1,y)\cdot \frac{1}{2}\, \frac{p(x_2,y)}{\sum_x p(x,y)} \overset{?}{=} p(x_2,y)\cdot \frac{1}{2}\, \frac{p(x_1,y)}{\sum_x p(x,y)}$$

choose first coordinate among two

Yes! $\square$

Ex Show that Gibbs sampler is a particular case of Hastings-Metropolis algorithm, where all transitions are accepted.

Example of a Gibbs sampler NOT on $\mathbb{R}^n$:

Gibbs sampler for random permutations:

~~Choose any~~

- S_d = {all $n!$ permutations on d symbols}

  "Symmetric group"

  e.g. $S_3 = \{abc, acb, bac, bca, cab, cba\}$

- How to sample a random permutation $\sim Unif(S_d)$?

Gibbs sampler:

- choose a pair symbols at random
- swap them ("transposition")
- repeat

$\boxed{\text{THM The result of Gibbs sampler converges to } Unif(S_d).}$

Let's check the reversibility condition of Thm p. 127.

$\forall$ permutations $\rho, \eta \in S_d$ that differ by one transposition,

~~let's check the~~

$$\underbrace{\pi_\rho}_{=\frac{1}{n!}} \cdot \underbrace{P_{\rho,\eta}}_{=\frac{1}{\binom{n}{2}}} \overset{?}{=} \underbrace{\pi_\eta}_{=\frac{1}{n!}} \cdot \underbrace{P_{\eta,\rho}}_{=\frac{1}{\binom{n}{2}}}. \quad \text{Yes!} \quad \square$$

(choose a pair of symbols to swap)