6.896: Probability and Computation

Spring 2011

lecture 2

Constantinos (Costis) Daskalakis
costis@mit.edu

Recall: the MCMC Paradigm

Input: a. very large, but finite, set Ω ;
b. a positive weight function $w : \Omega \rightarrow R^+$.

Goal: Sample $x \in \Omega$, with probability $\pi(x) \propto w(x)$.

in other words: $\pi(x) = \frac{w(x)}{Z}$

the "partition function"
$Z = \sum_{x \in \Omega} w(x)$

MCMC approach:

construct a Markov Chain (think sequence of r.v.'s) $(X_t)_t$ converging to π, i.e.

$\Pr[X_t = y \mid X_0 = x] \rightarrow \pi(y)$ as $t \rightarrow +\infty$ (independent of x)

Markov Chains

Def: A ***Markov Chain*** on Ω is a stochastic process $(X_0, X_1, \ldots, X_t, \ldots)$ such that

a. $X_t \in \Omega, \ \forall t$

b. $\Pr[X_{t+1} = y \mid X_t = x, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0] \equiv \Pr[X_{t+1} = y \mid X_t = x]$

the ***transition probability*** from state x to state y $\quad P(x, y)$

Properties of the matrix P:

Non-negativity: $\forall x, y \in \Omega, P(x, y) \geq 0;$

Stochasticity: $\displaystyle\sum_{y \in \Omega} P(x, y) = 1, \forall x \in \Omega\,.$

such a matrix is called ***stochastic***

Card Shuffling

Sample a random permutation of a deck of cards

Ω = {all possible permutations}

$w(x) = 1$, for all permutations x

Markov Chain:

and repeat forever

X_t: state of the deck after the t-th riffle; X_0 is initial configuration of the deck;

X_{t+1} is independent of $X_{t-1},\ldots,X_0$ conditioning on X_t.

Evolution of the Chain

$p_x^{(t)} \in \mathbb{R}_+^{1\times|\Omega|}$: distribution of X_t conditioning on $X_0 = x$.

then

$$\boxed{\begin{aligned} p_x^{(t+1)} &= p_x^{(t)} P \\ p_x^{(t)} &= p_x^{(0)} P^t \end{aligned}}$$

Graphical Representation

Represent Markov chain by a graph $G(P)$:

- nodes are identified with elements of the state-space Ω

-there is a directed edge between states x and y if $P(x, y) > 0$, with edge-weight $P(x,y)$;

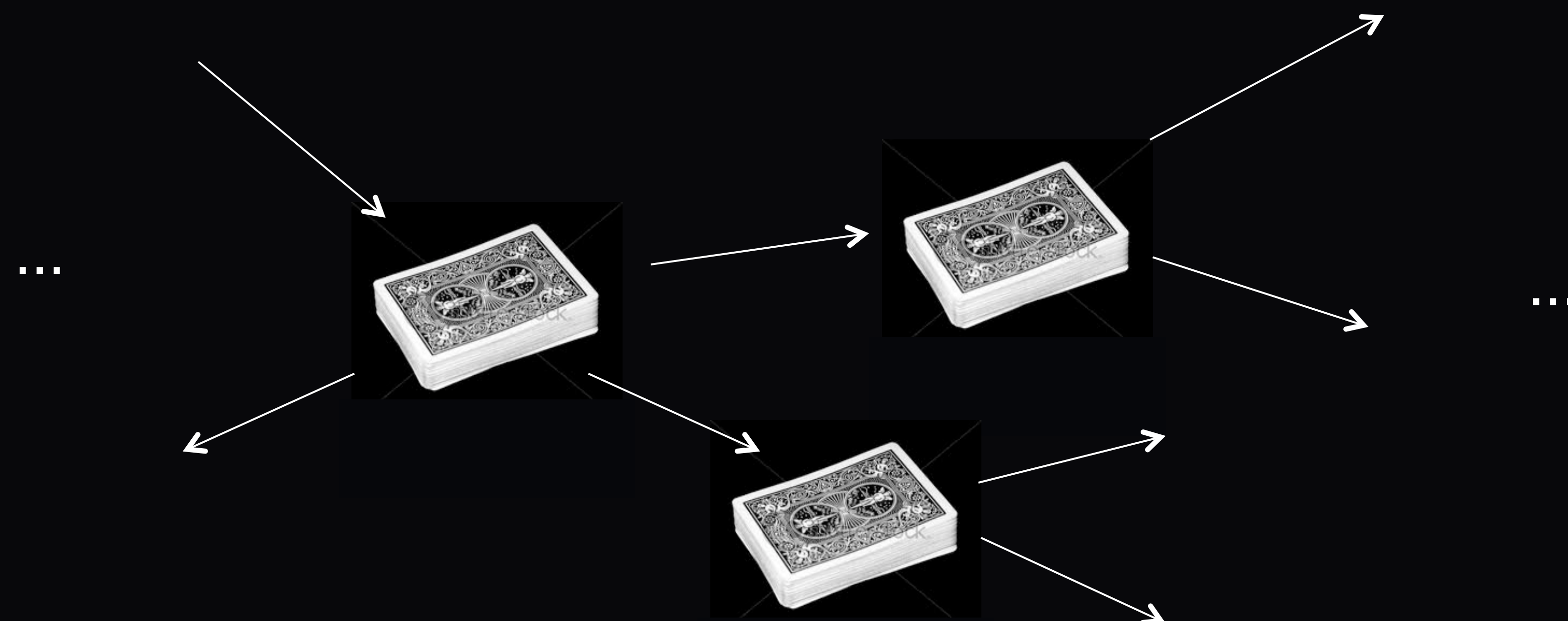
- no edge if P(x,y)=0;

- self loops are allowed (when $P(x,x) > 0$)

Much of the theory of Markov chains only depends on the topology of $G(P)$, rather than its edge-weights.

Many natural Markov Chains have the property that $P(x, y)>0$ iff $P(y, x)>0$. In this case, we'll call $G(P)$ ***undirected*** (ignoring the potential difference in the weights on an edge).

e.g. card Shuffling



“ $\longrightarrow$ ” : reachable via a cut and riffle

e.g. of non-edge: no way to go from permutation 1234 to 4132

e.g. of directed edge: Can go from 123456 to 142536, but not vice versa

Ir-reducibility and A-periodicity

Def: A Markov chain P is ***irreducible*** if for all x, y, there exists some t such that $P^t(x, y) > 0$.

[Equivalently, $G(P)$ is strongly connected. In case the graphical representation is an undirected graph, then <==> $G(P)$ being connected.]

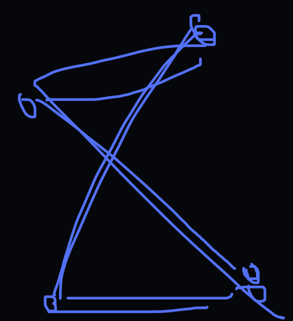
Def: A Markov chain P is ***aperiodic*** if for all x, y we have

$$\gcd\{t : P^t(x, y) > 0\} = 1.$$

True or False

For an irreducible Markov chain P, if $G(P)$ is undirected then aperiodicity is equivalent to $G(P)$ being non-bipartite.

A: true, look at lecture notes



True or False (ii)

Define the period of x as $\gcd\{t : P^t(x, x) > 0\}$. For an irreducible Markov chain, the period of every $x \in \Omega$ is the same.

A: true, 1 point exercise

[Hence, if $G(P)$ is undirected, the period is either 1 or 2.]

True or False (iii)

Suppose P is irreducible. Then P is aperiodic iff there exists t such that $P_t(x,y) > 0$ for all $x, y \in \Omega$.

A: true, 1 point exercise to fill in the details of the sketch we discussed in class. For the forward direction, you may want to use the concept of the ***Frobenius number*** (aka the ***Coin Problem***).

True or False (iv)

Suppose P is irreducible and contains at least one self-loop (i.e., $P(x, x) > 0$ for some x). Then P is aperiodic.

A: true, easy to see.

Stationary Distribution

Def: A probability distribution π over Ω is a ***stationary distribution*** for P if $\pi = \pi P$.

Theorem (Fundamental Theorem of Markov Chains) :

If a Markov chain P is ***irreducible*** and ***aperiodic*** then it has a unique stationary distribution π.

In particular, π is the unique (normalized such that the entries sum to 1) left eigenvector of P corresponding to eigenvalue 1.

Finally, $P^t(x, y) \rightarrow \pi(y)$ as $t \rightarrow \infty$ for all $x, y \in \Omega$.

In light of this theorem, we shall sometimes refer to an irreducible, aperiodic Markov chain as **ergodic**.

Reversible Markov Chains

Def: Let $\pi > 0$ be a probability distribution over Ω. A Markov chain P is said to be ***reversible wrt π*** if

$$\forall\ x, y \in \Omega: \ \pi(x)\, P(x, y) = \pi(y)\, P(y,x).$$

Note that any symmetric matrix P is trivially reversible (w.r.t. the uniform distribution π).

Lemma: If a Markov chain P is reversible w.r.t. π, then π is a stationary distribution for P.

Reversible Markov Chains

Representation by ***ergodic flows*:**

detailed balanced condition

$$Q(x,y) := \boxed{\pi(x) \cdot P(x,y) \equiv \pi(y)P(y,x)}$$

the amount of probability mass flowing from *x* to *y* under π

From flows to transition probabilities:

$$P(x,y) = \frac{Q(x,y)}{\sum_x Q(x,y)} \qquad \text{(verify)}$$

From flows to stationary distribution:

$$\frac{\pi(x)}{\pi(y)} = \frac{P(y,x)}{P(x,y)} \qquad \text{(verify)}$$

Mixing of Reversible Markov Chains

Theorem (Fundamental Theorem of Markov Chains) :

If a Markov chain P is ***irreducible*** and ***aperiodic*** then it has a unique stationary distribution π.

In particular, π is the unique (normalized such that the entries sum to 1) left eigenvector of P corresponding to eigenvalue 1.

Finally, $P^t(x, y) \to \pi(y)$ as $t \to \infty$ for all $x, y \in \Omega$.

Proof of FTMC: For reversible Markov Chains (today on the board-see lecture notes); full proof next time (probabilistic proof).

Mixing in non-ergodic chains

When P is irreducible (but not necessarily aperiodic), then π still exists and is unique, but the Markov chain does not necessarily converge to π from every starting state.

For example, consider the two-state Markov chain with P = [0 1 ; 1 0].

This has the unique stationary distribution $\pi = (1/2,1/2)$, but does not converge from either of the two initial states.

Notice that in this example $\lambda_0 = 1$ and $\lambda_1 = -1$, so there is another eigenvalue of magnitude 1.

Lazy Markov Chains

Observation: Let P be an irreducible (but not necessarily aperiodic) stochastic matrix. For any $0 < \alpha < 1$, the matrix $P' = \alpha P + (1 - \alpha) I$ is stochastic, irreducible and aperiodic, and has the same stationary distribution as P.

This operation going from P to P' corresponds to introducing a self-loop at all vertices of $G(P)$ with probability $1 - \alpha$.

Such a chain P' is usually called a ***lazy version of P***.

e.g. Card Shuffling

Argue that the following shuffling methods converge to the uniform distribution:

- Random Transpositions

  Pick two cards i and j uniformly at random with replacement, and switch cards i and j; repeat.

- Top-in-at-Random:

  Take the top card and insert it at one of the n positions in the deck chosen uniformly at random; repeat.

- Riffle Shuffle:

  a. Split the deck into two parts according to the binomial distribution Bin(n, 1/2).

  b. Drop cards in sequence, where the next card comes from the left hand L (resp. right hand R) with probability $\frac{|L|}{|L|+|R|}$ (resp. $\frac{|R|}{|L|+|R|}$).

  c. Repeat.