**COMP 480/580 — Probabilistic Algorithms and Data Structure** April 9, 2020

Lecture 20

*Lecturer: Anshumali Shrivastava* *Scribe By: Noushin Quazi*

# 1 Review

## 1.1 Markov Property

The Markov property describes a random process where the probability of arriving to the current state is only dependent on the previous state. This leads to the following update dynamic for a distribution $\pi$ over a Markov chain with transition matrix $P$:

$$\pi_t = \pi_{t-1} P$$

or equivalently

$$\pi_t = \pi_0 P^t$$

## 1.2 Stationary Distribution

A distribution $\pi$ over a Markov chain with transition matrix $P$ is said to be stationary once it no longer changes for further timesteps ie. it has converged. Mathematically, a stationary distribution exhibits

$$\pi = \pi P$$

Note that $\pi$ is an eigenvector of $P$ with an eigenvalue of 1.

# 2 Properties of Markov Chains

Going forward, I have marked the timestamp that each concept is brought up in the original lecture video.

Now, consider a Markov chain with state space $\Omega$ and transition matrix $P$.

## 2.1 Irreducibility

**Timestamp 7:23**
A Markov chain is irreducible if every state is reachable from every other state ie. there exists a path between them. Formally, a Markov chain is irreducible if it satisfies the following criterion:

$$\forall x, y \in \Omega \left( \exists t (P^t(x, y) > 0) \right)$$

Note that $P^t(x, y)$ refers to the element corresponding to states $(x, y)$ of matrix $P^t$. Irreducibility is key for convergence; otherwise, the resulting distribution will depend on the starting state.

## 2.2 Aperiodicity

**Timestamp 11:53**
A Markov chain is aperiodic if no sequence of steps taken between any pair of states has a period. This includes the sequence of steps taken to start from and return to the same state. As an example of a periodic Markov chain, consider a simple undirected graph with two states connected by an edge; the number of steps required to start from one state and return to that state is divisible by 2, so the period is 2. Formally, a Markov chain is aperiodic if it satisfies the following:

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

where $GCD$ refers to the greatest common divisor.

### 2.2.1 Ehrenfest Urn

A well-known example of a periodic Markov chain with a known stationary distribution is the Ehrenfest process. It involves two urns and $n$ balls split randomly among the urns. The random process comprises of picking one of the balls and swapping the urn it belongs to. The random variable $X$ = number of balls observed in urn 1, so the state space is $[1, 2...n]$. This leads to the following transition probabilities:

$$P(X_{t+1} = i + 1|X_t = i) = 1 - i/n$$
$$P(X_{t+1} = i - 1|X_t = i) = i/n$$

The intuition is straightforward: urn 1 can only gain a ball if a ball from urn 2 has been selected, and urn 1 can only lose a ball if a ball from urn 1 has been selected. This is periodic because the period for starting and returning to the same state is 2. The stationary distribution is $\pi(x) = \frac{\binom{n}{x}}{2^n}$

## 2.3 Reversibility

**Timestamp 32:13**
A Markov chain is reversible w.r.t a distribution $\pi$ if all pairs of states have the same joint probability:

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

**Theorem.** The distribution corresponding to a reversible Markov chain is a stationary distribution.

*Proof.* Assume $\pi(x)P(x, y) = \pi(y)P(y, x)$. We show that $\pi(x)$ remains unchanged after applying $P$.

$$\begin{aligned}(\pi P)_x &= \Sigma_z \pi(z)P(z, x) \\ &= \Sigma_z \pi(x)P(x, z) \\ &= \pi(x)\Sigma_z P(x, z) \\ &= \pi(x)\end{aligned}$$

□

This result is important for constructing a Markov chain for a given $\pi$ and is used in the Metropolis process shown later.

# 3 Fundamental Theorem of Markov Chains

**Timestamp 18:45**

**Theorem.** If the Markov chain with transition matrix $P$ is irreducible and aperiodic, then there is a unique stationary distribution that is the first left eigenvector of P with eigenvalue 1.

This is a useful theorem as it means that we can start with any initial distribution $\pi_0$ and guarantee the same solution every time. Also, we can ascertain the dynamics of our solution under this theorem.

## 3.1 Analysis

**Timestamp 22:23**
This is where we bring in our linear algebra. As stated above, if a Markov chain is irreducible and aperiodic, then the stationary distribution $\pi$ is the left eigenvector with eigenvalue ($\lambda_1$) of 1. Assume that all other eigenvalues are less than 1, which is necessary to ensure convergence. Then, consider the initial distribution $\pi_0$ rewritten as a linear combination of $P$'s eigenvectors $\{e_i\}$.

$$\begin{aligned}
\pi_0 &= \Sigma_{i=0}^{n} a_i e_i \\
\pi_t &= \pi_0 P^t \\
&= \Sigma_{i=0}^{n} a_i e_i P^t \\
&= \Sigma_{i=0}^{n} a_i e_i \lambda_i^t \\
&= a_1 e_1 + \Sigma_{i=0}^{n} a_i e_i \lambda_i^t \\
&= a_1 e_1 + O(\lambda_2^t \Sigma_{i=0}^{n} a_i e_i) \\
&= a_1 e_1 && t \to \infty
\end{aligned}$$

The Markov chain converges to the stationary distribution at a rate of $\lambda_2^t$.

## 3.2 Fixing Markov Chains

If a Markov chain is neither aperiodic or irreducible, it is easy to fix the chain such that the fundamental theorem may still apply. If it is not irreducible, that means there are a set of states that cannot reach other; simply adding transition probabilities with very small probabilities ensures irreducibility without drastically altering the chain. If the chain is periodic, adding self-loops breaks periodicity because every pair of states can add +1 to the number of steps between them.

# 4 Metropolis Process

**Timestamp 29:15**

Recall our goal to use Markov Chains to efficiently sample from some distribution $\pi$. Until now, we have looked at examples where the Markov chain was given, but how do we find Markov chains for $\pi$ in the first place? The Metropolis process uses the reversibility property discussed in previous sections to construct a Markov chain. In essence, it finds a Markov chain that is reversible w.r.t $\pi$.

First, we start with a random irreducible Markov chain with arbitrary transition matrix $P$. Then, we update $P$ according to another probability matrix $A$ in the following way:

$$P' = PA$$

$A$ provides accept/reject conditions for updating P, which is illustrated in the following simple algorithm:

**Algorithm 1:** Metropolis Process (Timestamp 59:55)

Start at state $x$;
Sample next state $y$ from P(x,·);
Move to $y$ with probability A(x,y);
Else remain at $x$ with probability 1 - A(x,y);

Visually, the algorithm is accomplishing the following:

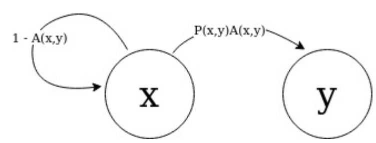


Figure 1: Metropolis process updating $P(x, y)$

We want to find a probability matrix $A$ such that $\pi(x)P'(x, y) = \pi(y)P'(y, x)$. Then substituting $A$ gives

$$\pi(x)P(x, y)A(x, y) = \pi(y)P(y, x)A(y, x)$$
$$\frac{A(x, y)}{A(y, x)} = \frac{\pi(y)P(y, x)}{\pi(x)P(x, y)}$$

Since $A$ is a probability matrix, we must constrain the solution space to

$$A(x, y) = \min\left(1, \frac{\pi(y)P(y, x)}{\pi(x)P(x, y))}\right)$$
$$A(y, x) = \min\left(1, \frac{\pi(x)P(x, y)}{\pi(y)P(y, x))}\right)$$

To convince ourselves that this simple fix works, consider when A(x,y) = 1. Then, $\frac{\pi(x)P(x,y)}{\pi(y)P(y,x))} \leq 1$, $A(y, x) = \frac{\pi(x)P(x,y)}{\pi(y)P(y,x))}$, and $\frac{1}{A(y,x)} = \frac{\pi(y)P(y,x)}{\pi(x)P(x,y))}$.