Markov Chains II
Dmitri Loguinov
Texas A&M University

February 28, 2017
Agenda

- Markov chains
  - Definitions
  - Transitional probabilities
- Packet loss example
- Stationary distribution
  - Under convergence assumption
Markov Chains

\[ p_{ij} = P(X_{n+1} = j | X_n = i) \]

- Matrix \( P = (p_{ij}) \) is called the one-step transition probability matrix.

- **Example**: determine \( P \) for \( X_n \) being the number of heads in \( n \) coin-flips.

\[
p_{ii} = P(X_{n+1} = i | X_n = i) = 1 - p
\]

\[
p_{i,i+1} = P(X_{n+1} = i + 1 | X_n = i) = p
\]

\[
P = \begin{pmatrix}
1 - p & p & 0 \\
p & 1 - p & p \\
0 & \cdots & \cdots
\end{pmatrix}
\]
Markov Chains 7

• Notice that the sum of each row in $P$ equals 1
  - Any matrix with this property is called *stochastic*
  - Reason: the summation of transition probabilities out of any state $i$ must be 1 (i.e., $\Sigma_j p_{i,j} = 1$)

• In addition, if the sum of each column is 1, the matrix is called *doubly stochastic*
  - $P^T$ is also a valid transition probability matrix
  - Can jump both $i \rightarrow j$ and $j \rightarrow i$
  - Not a requirement for Markov chains
Markov Chains 8

- Suppose the initial probability to find $X$ is any of its states is given by vector $a = (a_1, a_2, \ldots)$:

  $$P(X_0 = i) = a_i$$

- Further denote by $a^{(n)} = (a_1^{(n)}, a_2^{(n)}, \ldots)$ the vector of probabilities to find $X_n$ in each state $i$ at time $n$:

  $$P(X_n = i) = a_i^{(n)}$$

- Next, define $n$-step transition probabilities:

  $$p_{ij}^{(n)} = P(X_n = j | X_0 = i)$$
Markov Chains

- Matrix $P^{(n)}$ consists of probabilities $(p_{ij}^{(n)})$
- Example: for the coin-toss $X_n$, compute $p_{i,i+2}^{(3)}$
- Solution:
  - There must be two heads in 3 tosses and one tail:
    
    $$p_{i,i+2}^{(3)} = (1-p)p^2 + p(1-p)p + p^2(1-p) = 3p^2(1-p)$$
  - You could have also used the binomial distribution:
    
    $$p_{i,i+2}^{(3)} = \binom{3}{2}(1-p)p^2 = 3p^2(1-p)$$
Now, we obtain $P^{(n)}$ in closed form

Write:

\[ P(B) = \sum_i P(BA_i), \{A_i\} \text{ is a partition} \]

\[ p_{ij}^{(n+s)} = P(X_{n+s} = j|X_0 = i) \]
\[ = \sum_k P(X_{n+s} = j, X_n = k|X_0 = i) \]
\[ = \sum_k P(X_{n+s} = j|X_n = k, X_0 = i)P(X_n = k|X_0 = i) \]

Using the Markov property:

\[ p_{ij}^{(n+s)} = \sum_k P(X_{n+s} = j|X_n = k)P(X_n = k|X_0 = i) \]
\[ = \sum_k P(X_s = j|X_0 = k)P(X_n = k|X_0 = i) \]
Thus we get the Chapman-Kolmogorov equation:

\[ p_{ij}^{(n+s)} = \sum_k p_{ik}^{(n)} p_{kj}^{(s)} \]

- Notice that this is a product of the \( i \)-th row of \( P^{(n)} \) and the \( j \)-th column of \( P^{(s)} \)

This leads to:

\[ P^{(n+s)} = P^{(n)} P^{(s)} \]

Applying this once for \( n = s = 1 \), we have \( P^{(2)} = P^2 \) and recursively expanding:

\[ P^{(n)} = P^n \]
Markov Chains 12

• Finding the probability that a chain has moved from state $i$ to any state $j$ involves matrix multiplication.

• It is also simple to express the probability that at time $n$, the chain is in any of its states:

$$a^{(n)} = a^{(n-1)} P$$

Sketch of proof:

$$a^{(n)}_i = P(X_n = i) = \sum_k P(X_{n-1} = k)p_{ki} = \sum_k a^{(n-1)}_k p_{ki}$$

• We can also write:

$$a^{(n)} = aP^n$$
Markov Chains 13

- Markov chains are often used to model simple processes with a small number of states

- One such example is packet loss
  - Suppose we write 0 when there is no loss (or error) in the channel and 1 when there is
  - We run some traffic over a lossy channel and obtain a packet loss pattern: 0001100000101011110000
  - Design a Markov chain for this loss process
Markov Chains 14

• There are 22 bits and 21 transitions
  – Counting all possible transitions, we have
    \[ p_{00} = q = \frac{9}{13}, \quad p_{11} = p = \frac{1}{2} \]

• Q: assuming that packet loss follows this Markov chain, what is the probability to lose a burst of at least \( k \) packets starting from a random time \( n \gg 1 \)?

• A: if we know \( P(X_n = 1) \) at time \( n \), then:
  \[ P(\text{k burst}) = P(X_n = 1)p^{k-1} \]

• Next, we derive probabilities \( P(X_n = i) \) for \( n \to \infty \)
Markov Chains 15

• Recall that $a_i^{(s)}$ is the probability to find the chain in state $i$ after $s$ steps given that it started in $a$
  - Note that the limit of $a_i^{(s)}$ may not exist as $s \to \infty$
  - Instead, we are interested in the time-average of $a_i^{(s)}$ since it represents the fraction of time the chain spends in state $i$

• This can be viewed as a time-averaged integral of this curve:
• More specifically, define the fraction of time the chain spends in state $i$:

$$\pi_i = \lim_{n \to \infty} \frac{1}{n} \sum_{s=1}^{n} a_i^{(s)}$$

• Existence of this limit is not obvious and will be established next time using renewal theory
  – For now, we assume a stronger condition:
    $$\exists \lim_{s \to \infty} a_i^{(s)} = \zeta_i$$
    – Notice that if this limit in fact exists, it equals the one above:
    $$\zeta_i = \pi_i$$
Markov Chains 17

• Using this simpler definition of the probability to find the process in any state \( i \), we have:

\[
a^{(s)} = a^{(s-1)} P \Rightarrow \lim_{s \to \infty} a^{(s)} = \lim_{s \to \infty} a^{(s-1)} P
\]

• Since both limits are the same, we have

\[
\zeta = \zeta P
\]

- where \( \zeta \) is the left eigen-vector of \( P \)

\[
\zeta = (\zeta_0, \zeta_1, \ldots)
\]

• Values \( \zeta_i \) are also called *stationary probabilities*

- Also note that if the chain starts with \( a = \zeta \), it follows that \( a^{(n)} = \zeta \) for all \( n \)
• Going back to the packet loss example
  - We first construct matrix $P$

$$P = \begin{pmatrix} q & 1 - q \\ 1 - p & p \end{pmatrix}$$

  - Next solve the system of equations

$$\zeta = \zeta P = \begin{pmatrix} \zeta_0 q + \zeta_1 (1 - p) \\ \zeta_0 (1 - q) + \zeta_1 p \end{pmatrix}$$

  - Rewriting, we get a single equation:

$$\begin{cases} \zeta_0 = \zeta_0 q + \zeta_1 (1 - p) \\ \zeta_1 = \zeta_0 (1 - q) + \zeta_1 p \end{cases} \Rightarrow \frac{\zeta_0}{\zeta_1} = \frac{1 - p}{1 - q}$$
• Recalling that $\zeta_0 + \zeta_1 = 1$, we have:

\[
\begin{align*}
\zeta_0 &= \frac{1-p}{2-(p+q)} \\
\zeta_1 &= 1 - \zeta_0 = \frac{1-q}{2-(p+q)}
\end{align*}
\]

• Which leads to

\[
\begin{align*}
\zeta_0 &= 0.619 \\
\zeta_1 &= 0.381
\end{align*}
\]

- Long-term, 61.9% of all packets are transmitted uncorrupted/without loss, 38.1% are dropped by the network
Let us construct an example when $\zeta$ does not exist, but $\pi$ (time-average) does:

- Assume a deterministically alternating chain with $p_{01} = p_{10} = 1$ and $a_0 = 1 = 1 - a_1$
- Then, for all even values of $s$, the chain is always in state 0 and for all odd $s$, it is always in state 1
- Thus, the limit of $a^{(s)}$ does not exist

At the same time, the summation limit is well-defined and correctly gives the time spent in each state:

$$\pi_0 = \lim_{n \to \infty} \frac{1}{n} \sum_{s=1}^{n} a_0^{(s)} = 1/2$$
**Wrap-up**

- Under the Markov loss chain, what is the probability to receive exactly $k$ packets (no more, no less) starting at some random time $n$?

\[ P(k \text{ are good}) = \pi_0 q^{k-1} (1 - q) \]

- This is a geometric-like distribution

- Notice that burst lengths of both 0s and 1s have **exponential** tails
  - More generic models allows substantially longer (i.e., Pareto) bursts using non-Markovian dynamics