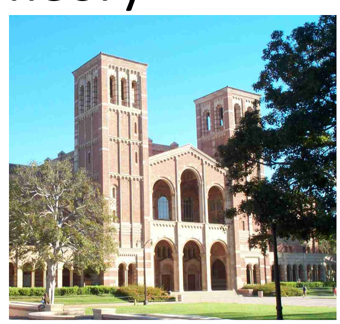
EE 231A: Information Theory Lecture 4

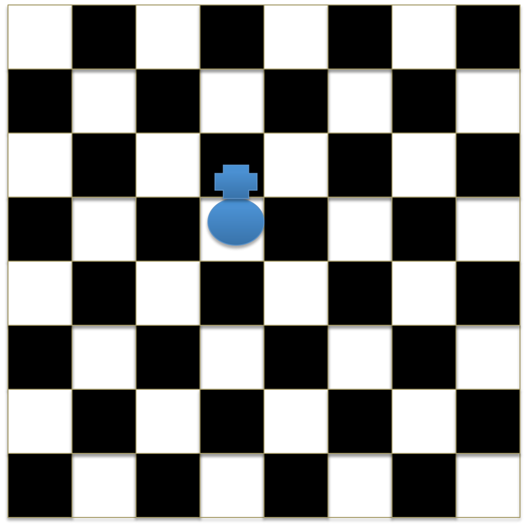
A. Entropy Rate
B. Entropy Rate of Stationary Processes
C. Stationary Markov Chains and the Stationary Distribution
D. Entropy Rate for Markov Chains Including Random Walks on a Weighted Graph
E. The General AEP

EE 231A: Information Theory Lecture 4



A. Entropy Rate
B. Entropy Rate of Stationary Processes
C. Stationary Markov Chains and the Stationary Distribution
D. Entropy Rate for Markov Chains Including Random Walks on a Weighted Graph
E. The General AEP

- How many bits does it take per move to describe the random walk of a king on the chess board?
- 8 moves = 3 bits
- But what about the corner?



Definition of Entropy Rate

- $H(\{X_i\})$ is the <u>entropy rate</u> of $\{X_i\}$.
- Define $H(\{X_i\})$ as a limit

$$H(\{X_i\}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$$

if the limit exists.

Example: entropy rate of i.i.d. sequence

- When $X_1, X_2, \ldots$ are i.i.d.,

$$H(\{X_i\}) = H(X) \qquad \text{Why?}$$

$$\begin{aligned} H(\{X_i\}) &= \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i) \\ &= \lim_{n\to\infty} H(X) \\ &= H(X) \end{aligned}$$

Entropy Rate is Entropy for i.i.d. X

- When $\{X_i\}=X_1, X_2, \ldots$ are i.i.d., we can compress to $H(X)$ bits per symbol on the average.

- What if $X_1, X_2, X_3, \ldots$ are <u>not</u> i.i.d.?

Example: independent but not identically distributed random variables

- When $X_1, X_2, \ldots$ are independent,

$$H(\{X_i\}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1)$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i)$$ IF THE LIMIT EXISTS….

Example: general case

- When $X_1, X_2, \ldots$ are any sequence,

$$H(\{X_i\}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1)$$ IF THE LIMIT EXISTS….

Closely tied to $\lim_{n\to\infty} H(X_i \mid X_{i-1}, \ldots, X_1)$

EE 231A: Information Theory
Lecture 4

A. Entropy Rate
B. Entropy Rate of Stationary Processes
C. Stationary Markov Chains and the Stationary Distribution
D. Entropy Rate for Markov Chains Including Random Walks on a Weighted Graph
E. The General AEP

Stationary Process

- Stationary

$$P\left[(X_1,X_2,X_3)=(x_1,x_2,x_3)\right]=P\left[(X_8,X_9,X_{10})=(x_1,x_2,x_3)\right]$$

  - i.e. shifting the indices of observation by a fixed constant doesn't change the distribution,
  - as long as the relative position of the indices is maintained.

Theorem 4.2.2

- For a stationary process

$$H(\{X_i\}) = \lim_{n\to\infty} H(X_n \mid X_{n-1},\ldots,X_1)$$

- Proof:
  - 1) show limit exists
  - 2) show limit equal to $H(\{X_i\})$

(1) Show that the limit exists.

- 1) $H(X_{n+1} \mid X_n, X_{n-1},\ldots,X_1) \le H(X_{n+1} \mid X_n, X_{n-1},\ldots,X_2)$

$$= H(X_n \mid X_{n-1},\ldots,X_1)$$

- Since $H(X_n \mid X_{n-1},\ldots,X_1)$ is positive and non-increasing, it has a limit.

(2) Show limit is the entropy rate.

- Cesaro Mean: if $a_n \to a$ and $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, then $b_n \to a$
- Proof of 2):
  - Let $a_n = H(X_n \mid X_{n-1},\ldots,X_1)$

$$b_n = \frac{1}{n}\sum_{i=1}^{n} H(X_i \mid X_{i-1},\ldots,X_1)$$
$$= \frac{1}{n} H(X_1,\ldots,X_n)$$

  - So $H(\{X_i\}) = \lim_{n\to\infty} b_n = \lim_{n\to\infty} a_n$, proving the theorem.

Proof of the Cesaro mean

- Cesaro Mean: if $a_n \to a$ and $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, then $b_n \to a$

$$|b_n - a| = \left|\frac{1}{n}\sum_{i=1}^{n}(a_i - a)\right|$$
$$\le \frac{1}{n}\sum_{i=1}^{n}|a_i - a|$$
$$= \frac{1}{n}\sum_{i=1}^{N(\varepsilon)}|a_i - a| + \frac{1}{n}\sum_{i=N(\varepsilon)+1}^{n}|a_i - a|$$

Proof of the Cesaro mean

- Cesaro Mean: if $a_n \to a$ and $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, then $b_n \to a$

$$|b_n - a| = \left|\frac{1}{n}\sum_{i=1}^{n}(a_i - a)\right|$$

$$\le \frac{1}{n}\sum_{i=1}^{n}|a_i - a|$$

$$\le \frac{1}{n}\sum_{i=1}^{N(\varepsilon)}|a_i - a| + \frac{1}{n}\sum_{i=N(\varepsilon)+1}^{n}\varepsilon$$

Proof of the Cesaro mean

- Cesaro Mean: if $a_n \to a$ and $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, then $b_n \to a$

$$|b_n - a| = \left|\frac{1}{n}\sum_{i=1}^{n}(a_i - a)\right|$$

$$\le \frac{1}{n}\sum_{i=1}^{n}|a_i - a|$$

$$\le \frac{1}{n}\sum_{i=1}^{N(\varepsilon)}|a_i - a| + \frac{n - N(\varepsilon)}{n}\varepsilon$$

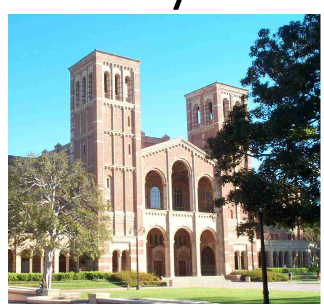
$$\le \frac{1}{n}\sum_{i=1}^{N(\varepsilon)}|a_i - a| + \varepsilon$$

Key result for Stationary Processes

- For a stationary process

$$H(\{X_i\}) = \lim_{n\to\infty} H(X_n \mid X_{n-1},\ldots,X_1)$$

EE 231A: Information Theory Lecture 4



A. Entropy Rate
B. Entropy Rate of Stationary Processes
C. Stationary Markov Chains and the Stationary Distribution
D. Entropy Rate for Markov Chains Including Random Walks on a Weighted Graph
E. The General AEP

Markov process (or Markov Chain)

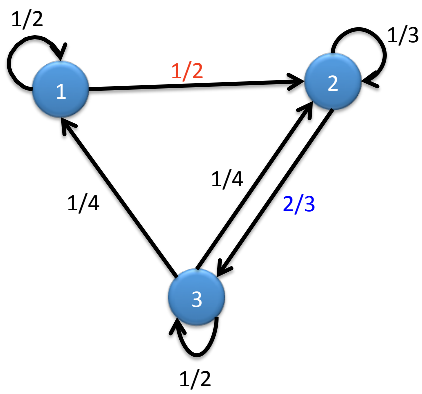
$$P\left(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \cdots, X_1 = x_1\right)$$
$$= P\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right)$$

Time invariant Markov Chain

$$P\left(X_{n+1} = b \mid X_n = a\right) = P\left(X_2 = b \mid X_1 = a\right) \quad \forall n, \forall a, b \in \chi$$

For all

Time Invariant Markov Chain



$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Row is beginning state.
Column is ending state.

Stationary distribution

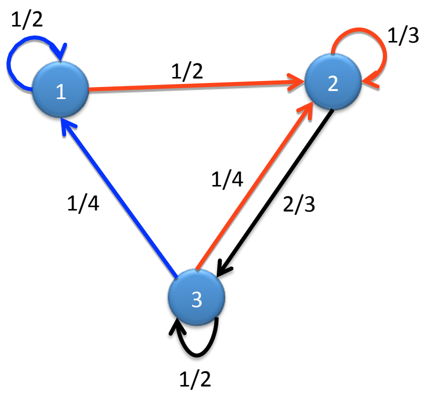
- Is a time-invariant Markov chain a stationary process?
  - It depends on the initial state.
- <u>Stationary distribution</u>

$$\mu = [\mu_1 \quad \mu_2 \quad \mu_3]$$

satisfies $\mu P = \mu$.

Stationary Distribution Equations

- For our example $\mu P = \mu$ means



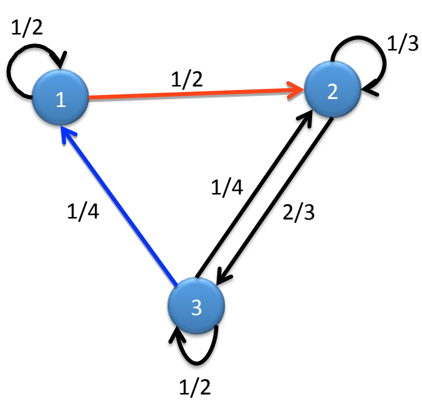
$$\mu_1 = \frac{\mu_1}{2} + \frac{\mu_3}{4}$$

$$\mu_2 = \frac{\mu_1}{2} + \frac{\mu_2}{3} + \frac{\mu_3}{4}$$

$$\mu_3 = \frac{2\mu_2}{3} + \frac{\mu_3}{2}$$

Alternatively, flow-out equals flow-in

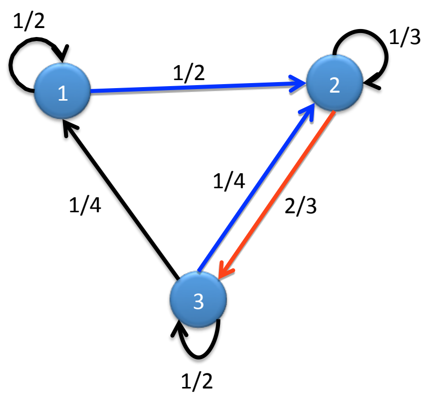
- An equivalent set of equations follows from the observation that for the stationary distribution, flow-out = flow-in.



$$\frac{\mu_1}{2} = \frac{\mu_3}{4}$$

Alternatively, , flow-out equals flow-in

- An equivalent set of equations follows from the observation that for the stationary distribution, flow-out = flow-in.

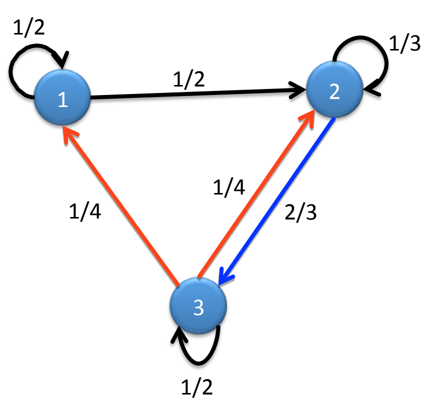


$$\frac{\mu_1}{2} = \frac{\mu_3}{4}$$

$$\frac{2\mu_2}{3} = \frac{\mu_1}{2} + \frac{\mu_3}{4}$$

Alternatively, , flow-out equals flow-in

- An equivalent set of equations follows from the observation that for the stationary distribution, flow-out = flow-in.



$$\frac{\mu_1}{2} = \frac{\mu_3}{4}$$

$$\frac{2\mu_2}{3} = \frac{\mu_1}{2} + \frac{\mu_3}{4}$$

$$\frac{\mu_3}{2} = \frac{2\mu_2}{3}$$

Finding the Stationary Distribution

$$\frac{\mu_1}{2} = \frac{\mu_3}{4}$$
$$\frac{2\mu_2}{3} = \frac{\mu_1}{2} + \frac{\mu_3}{4}$$
$$\frac{\mu_3}{2} = \frac{2\mu_2}{3}$$

$$\mu_1 + \mu_2 + \mu_3 = 1$$

$$\mu_3 = 2\mu_1$$

Finding the Stationary Distribution

$$\frac{\mu_1}{2} = \frac{\mu_3}{4}$$
$$\frac{2\mu_2}{3} = \frac{\mu_1}{2} + \frac{\mu_3}{4}$$
$$\frac{\mu_3}{2} = \frac{2\mu_2}{3}$$

$$\mu_1 + \mu_2 + \mu_3 = 1$$

$$\mu_3 = 2\mu_1$$

$$\mu_2 = \frac{3\mu_1}{2}$$

Finding the Stationary Distribution

$$\frac{\mu_1}{2} = \frac{\mu_3}{4}$$
$$\frac{2\mu_2}{3} = \frac{\mu_1}{2} + \frac{\mu_3}{4}$$
$$\frac{\mu_3}{2} = \frac{2\mu_2}{3}$$

$$\mu_3 = 2\mu_1$$

$$\mu_2 = \frac{3\mu_1}{2}$$

$$\mu_1 + \mu_2 + \mu_3 = 1$$

$$\mu_1 + 2\mu_1 + \frac{3}{2}\mu_1 = 1$$

Finding the Stationary Distribution

$$\frac{\mu_1}{2} = \frac{\mu_3}{4}$$
$$\frac{2\mu_2}{3} = \frac{\mu_1}{2} + \frac{\mu_3}{4}$$
$$\frac{\mu_3}{2} = \frac{2\mu_2}{3}$$

$$\mu_3 = 2\mu_1$$

$$\mu_2 = \frac{3\mu_1}{2}$$

$$\mu_1 + \mu_2 + \mu_3 = 1$$

$$\mu_1 + 2\mu_1 + \frac{3}{2}\mu_1 = 1$$

$$\mu = \begin{bmatrix} \frac{2}{9} & \frac{3}{9} & \frac{4}{9} \end{bmatrix}$$

Stationary Distribution

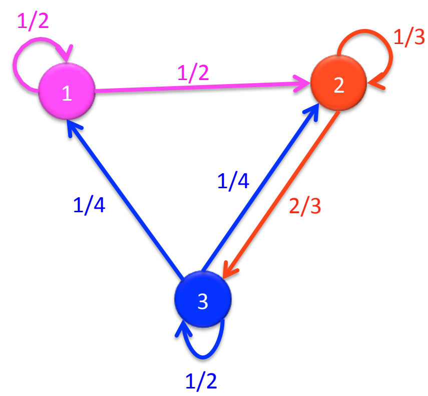
- The stationary distribution is well-defined for irreducible, aperiodic Markov chains.
- Irreducible: any state → any other state in a finite # of steps.
- Periodic state: $P(s_i \rightarrow s_i)$ in k transitions is nonzero only for $k=d$, $2d$, $3d$,… $d>1$.
- Aperiodic: no periodic states.

$H(\{X_i\})$ for stationary Markov chain

$$\begin{aligned}
H(\{X_i\}) &= \lim_{n\to\infty} H(X_n \mid X_{n-1},\dots,X_1) \\
&= \lim_{n\to\infty} H(X_n \mid X_{n-1}) \\
&= \lim_{n\to\infty} H(X_2 \mid X_1) \\
&= H(X_2 \mid X_1)
\end{aligned}$$

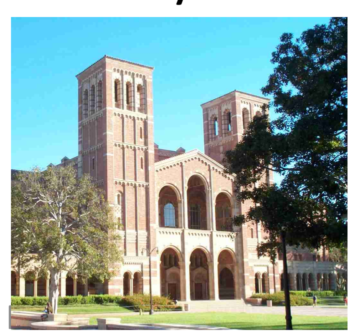
$H(\{X_i\})$ for our example

$$\mu = \begin{bmatrix} \frac{2}{9} & \frac{3}{9} & \frac{4}{9} \end{bmatrix}$$



$$\begin{aligned} H(\{X_i\}) &= H(X_2 \mid X_1) \\ &= \sum_{i=1}^{3} \mu_i H(X_2 \mid X_1 = \mu_i) \\ &= \mu_1 H(X_2 \mid X_1 = \mu_1) + \mu_2 H(X_2 \mid X_1 = \mu_2) + \mu_3 H(X_2 \mid X_1 = \mu_3) \\ &= \frac{2}{9} H\left(\frac{1}{2}\right) + \frac{3}{9} H\left(\frac{1}{3}\right) + \frac{4}{9} H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right) \end{aligned}$$

EE 231A: Information Theory Lecture 4



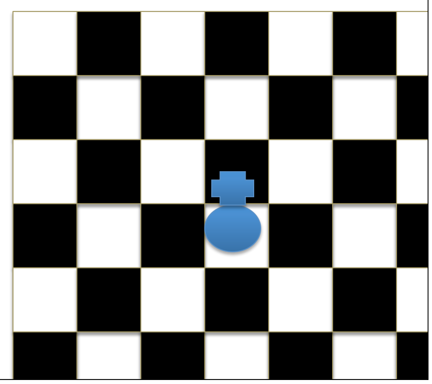
A. Entropy Rate
B. Entropy Rate of Stationary
C. Stationary Markov Chains and the Stationary Distribution
D. Entropy Rate for Markov Chains Including Random Walks on a Weighted Graph
E. The General AEP

Random Walk on a Chessboard

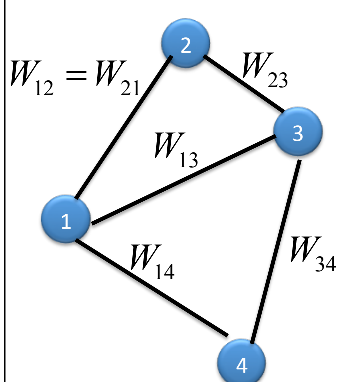
- A king is equally likely to move to any adjacent square at each transition.
- $H(X_2 \mid X_1) = \sum_x p(x) H(X_2 \mid X_1 = x)$
- $H(X_2 \mid X_1 \in \text{interior}) = \log 8$

  $H(X_2 \mid X_1 \in \text{corner}) = \log 3$

  $H(X_2 \mid X_1 \in \text{edge}) = \log 5$
- How do we determine $p(x)$?



Weighted Graphs



- $W_i = \sum_k W_{ik}$ is the sum of all weight leaving i .
- $2W = \sum_i W_i$ is the sum of all weight counting each edge twice
- If $P_{ij} = \dfrac{W_{ij}}{W_i}$
- then $\mu_i = \dfrac{W_i}{2W}$ is the stationary distribution.

Proof of Stationary Distribution

$$\mu P = \mu$$

$$\sum_i \mu_i P_{ij} = \sum_i \frac{W_i}{2W} \frac{W_{ij}}{W_i}$$

$$= \sum_i \frac{W_{ij}}{2W} = \frac{\sum_i W_{ij}}{2W} = \frac{W_j}{2W} = \mu_j$$

Back to our chess problem

- 4 corners with 3 edges
- 24 sides with 5 edges
- 36 interior with 8 edges

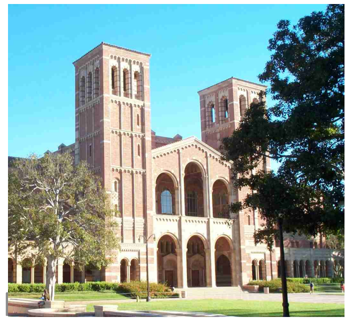
$$2W = 4 \cdot 3 + 24 \cdot 5 + 36 \cdot 8 = 420$$

$$\mu_{\text{corner}} = \frac{3}{420} \qquad \mu_{\text{side}} = \frac{5}{420} \qquad \mu_{\text{interior}} = \frac{8}{420}$$

- $H(X_2 \mid X_1) = \frac{4 \cdot 3}{420} \log 3 + \frac{24 \cdot 5}{420} \log 5 + \frac{36 \cdot 8}{420} \log 8$

  $= 2.77 \quad \text{or} \quad 0.92 \log 8$

EE 231A: Information Theory Lecture 4



A. Entropy Rate
B. Entropy Rate of Stationary
C. Stationary Markov Chains and the Stationary Distribution
D. Entropy Rate for Markov Chains Including Random Walks on a Weighted Graph
E. The General AEP

General AEP

- Theorem: for any stationary ergodic process,

$$-\frac{1}{n}\log p(X_1, X_2, ..., X_n) \to H(\{X_i\})$$

with probability 1.

- General AEP is better known as the Shannon, McMillan, Breiman theorem.

AEP properties carry over

- All the regular AEP properties follow with $H(X)$ replaced by $H(\{X_i\})$.
  - $|A_\varepsilon^{(n)}| \approx 2^{nH(\{X_i\})}$
  - $p(x^n) \approx 2^{-nH(\{X_i\})}$ for $x^n \in A_\varepsilon^{(n)}$
  - can compress x^n to $nH(\{Xi\})$ bits.

Stationary Process

- Stationary Necessary for $H(\{X_i\})$

$$P\left[(X_1, X_2, X_3) = (x_1, x_2, x_3)\right] = P\left[(X_8, X_9, X_{10}) = (x_1, x_2, x_3)\right]$$

  - i.e. shifting the indices of observation by a fixed constant doesn't change the distribution,
  - as long as the relative position of the indices is maintained.

Ergodic Process

- Ergodic Necessary for AEP
  - Time averages (summation over indices) equal ensemble averages (expectation) in the limit.
  - i.e. it satisfies a l.l.n.
- The ergodicity assumption makes the general AEP seem reasonable.
- The proof (see section 16.8) is rather long and not required.