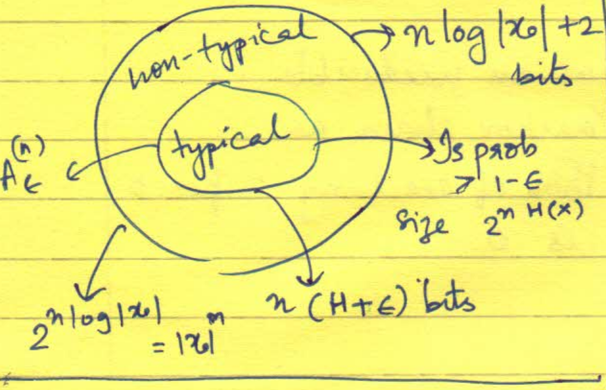
Sep 13th 2017.

lec 8: Entropy rate

Recap:

- typical sequence set



Allows us to compare a source to its entropy $H(X)$.

$\{X_n\}_{n=1}^{N} \approx NH(X)$ bits needed

provided N is large and X_i are iid.

What if they are not independant?

TODAY

- Stationarity & Markov chains
- Entropy rate
- Example

Defn: A stochastic process is an indexed sequence of RVs $\{X_i\}$

Characterized by joint pmf ← finite index set: { random vector

index set ⟺ $\mathbb{Z}$ or $\mathbb{N}$, $\mathbb{Z}_+$ { random sequence

index set ⟺ $\mathbb{R}'$, $\mathbb{N}$, $(a,b) \subseteq \mathbb{R}'$ random process

(we continue to assume that each X_i takes on values on a finite alphabet $\mathcal{X}$, $|\mathcal{X}| < \infty$)

$$\Pr(X_1 = x_1, X_2 = x_2, \dots X_n = x_n) \triangleq p(x_1, x_2, \dots x_n)$$

$$x_1, x_2, \dots x_n \in \mathcal{X}^n$$

Defn: A stochastic process $\{X_i\}$ is said to form a markov chain (MC) if $\Pr(X_n = x_n \mid X_1 = x_1 \dots X_{n-1} = x_{n-1}) = \Pr(X_n = x_n \mid X_{n-1} = x_{n-1})$ for all $n \geq 2$, $(x_1, x_2, \dots x_n) \in \mathcal{X}^n$

In this the joint pmf factors as follows

$$p(x_1, x_2, \dots x_n) = p(x_1)\prod_{i\geq 2}^{n} p(x_i \mid x_{i-1})$$

Defn: A markov chain is said to be time-invariant if

$$\Pr(X_n = b \mid X_{n-1} = a) = \Pr(X_2 = b \mid X_1 = a) \text{ for all } n \geq 2, \text{ all } a, b \in \mathcal{X}$$

We will assume throughout that the MCs we will deal with are time invariant. A time invariant MC is associated to a prob transition matrix $P = [P_{ij}]$ where

$$P_{ij} = \Pr(X_2 = j \mid X_1 = i) = \Pr(X_{n+1} = j \mid X_n = i)$$

$$1 \leq i, j \leq m, \text{ where } m = |\mathcal{X}|$$

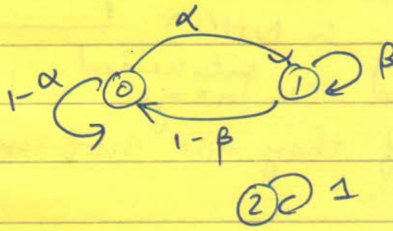
Defn: A markov chain is said to be irreducible if it is possible to go from any state of the markov chain to any other state of the markov chain with positive probability

By state of the markov chain at time n, we simply mean X_n.

<u>State diagram</u>

$\{0, 1, 2\} \equiv \mathcal{X}_0$

X_n



→ not an irreducible markov chain as Prob of reaching 2 from 0 is 0.

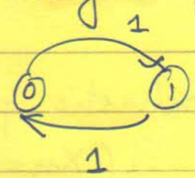
$$P_{X_{n+1}}(j) = \sum_{i \in \mathcal{X}_0} P_{X_n}(i)\, P(X_{n+1} = j \mid X_n = i)$$

$$= \sum_{i \in \mathcal{X}_0} P_{X_n}(i)\, P_{ij}$$

$$P_{ij} = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 1-\beta & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} P_{X_n}(0) & P_{X_n}(1) & P_{X_n}(2) \end{bmatrix} \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 1-\beta & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} P_{X_{n+1}}(0) & P_{X_{n+1}}(1) & P_{X_{n+1}}(2) \end{bmatrix}$$

<u>Defn</u>: A MC is said to be aperiodic if the gcd of the lengths of different the paths from a state to itself is 1



gcd 2

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \to \begin{bmatrix} 0 \\ 1 \end{bmatrix} \to \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

doesn't settle here

<u>Defn</u>: A prob distribution $\underline{p}$ is said to be stationary probability distribution for a MC if

$$\underline{p}^t P_{m \times m} = \underline{p}^t_{(m \times 1)}$$

$$p_j = \sum_{i=1}^{m} p_i P_{ij}$$

$$\begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix} \triangleq \underline{p} \triangleq \begin{bmatrix} P_r(X_{n+1} = 1) \\ \vdots \\ P_r(X_{n+1} = m) \end{bmatrix}$$

It follows that a MC whose initial probability distribution function is a stationary distribution is a stationary random process.

$$\underline{p}^t = \underline{p}^t P$$

$$\therefore P_{X_1} = P_{X_2} = P_{X_3} = \dots$$

<u>Stationarity defn</u>: $P(X_1 = x_1 \dots X_n = x_n) = P_r(X_{1+l} = x_1, X_{2+l} = x_2 \dots X_{n+l} = x_n)$

$P_r(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = P_{X_1}(x_1) \prod_{j=2}^{n} P(X_j = x_j \mid X_{j-1} = x_{j-1})$

$= P_{X_{1+l}}(x_1) \prod_{j=2}^{n} P(X_{j+l} = x_j \mid X_{j-1+l} = x_{j-1})$

~~For~~ If $P_{X_1} = p_{ss} \rightarrow$ steady state or stationary distribution

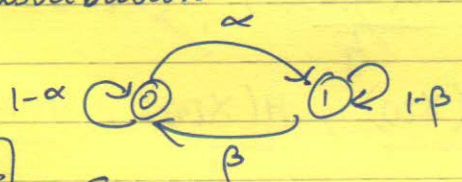
then $P_{X_{1+l}} = P_{X_1}$ $\therefore$ It will be a stationary process.

For aperiodic markov chains, let p_0 be the initial probability mass function then the sequence (p_n) converges to (p_{ss}) (p_n is pmf of X_n at time n)

$$p_0 \Rightarrow P_1 = p_0 P \Rightarrow P_2 = p_0 P^2 \Rightarrow P_n = p_0 P^n \xrightarrow[\text{Converges.}]{} p_{ss}$$ (steady state distribution)

Computing the stationary distribution

$\underline{p}^t = \underline{p}^t P$



$\mu\alpha = (1-\mu)\beta$

$$[p \quad 1-p] \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} = [p, 1-p].$$

$\Rightarrow p(1-\alpha) + \beta(1-p) = p.$

$p(1-\alpha-\beta) + \beta = p.$

$\Rightarrow p = \dfrac{\beta}{\alpha+\beta}.$ $\quad \therefore \left[\dfrac{\beta}{\alpha+\beta}, \dfrac{\alpha}{\alpha+\beta}\right]$

$\Downarrow$

Stationary distribution

If markov chain is ~~unique~~, aperiodic then this ($p_{ss} = p$) is unique.

$H(X_n) = H\left(\dfrac{\beta}{\alpha+\beta}, \dfrac{\alpha}{\alpha+\beta}\right)$ when initial distribution is the statitionary distribution.

Entropy rate:

Defn: the entropy rate of a stochastic process $\{X_i\}$ is defined by

$$H(\mathcal{X}) \triangleq \lim_{n\to\infty} \frac{H(X_1, \ldots X_n)}{n}$$ whenever the limit exists.

Eg: $\{X_i\}$ are iid

then $H(\mathcal{X}) = \lim_{n\to\infty} \dfrac{H(X_1, \ldots X_n)}{n} = \lim_{n\to\infty} \dfrac{n H(X)}{n} = H(X_1)$

What if not identically distributed

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i)$$ if the limit exists.

Eg: $\{X_i\}$ is Bernoulli with prob p_i independant.

$$p_i = \begin{cases} \frac{1}{2} & 2^{2k} < i \le 2^{2k+1}. \\ 0 & 2^{2k+1} < i \le 2^{2k+2} \quad \Leftrightarrow H(p_i) = 0 \\ 0 & i = 1, 2. \quad \Leftrightarrow H(p_i) = 0. \end{cases}$$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	..

Can verify that $\lim_{n \to \infty} \frac{H(X_1, \ldots X_n)}{n}$ doesn't converges.

$X_{[n]} = X_1 X_2 \ldots X_n$

Consider $\cancel{\frac{H(X_{[3]})}{3}}, \cancel{\frac{H(X_{[5]})}{5}}, \cancel{H(X_{[17]})}$

Consider sequences $\frac{H(X_{[1]})}{1}, \frac{H(X_{[4]})}{4}, \frac{H(X_{[16]})}{16}, \frac{H(X_{[64]})}{64} \ldots$

and $\frac{H(X_{[2]})}{2}, \frac{H(X_{[8]})}{8}, \frac{H(X_{[32]})}{32}, \ldots$

Sequence 1 $\qquad 0, \frac{1}{4}, \frac{1}{16} + \frac{1}{4}, \frac{1}{64} + \frac{1}{16} + \frac{1}{4}, \ldots \longrightarrow \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3}$

Sequence 2 $\qquad \frac{1}{2}, \frac{1}{8} + \frac{1}{2}, \frac{1}{32} + \frac{1}{8} + \frac{1}{2}, \ldots \longrightarrow \frac{\frac{1}{2}}{1 - \frac{1}{4}} = \frac{2}{3}$

There are two subsequences that converge to different values,
$\therefore$ the limit doesnot exist.