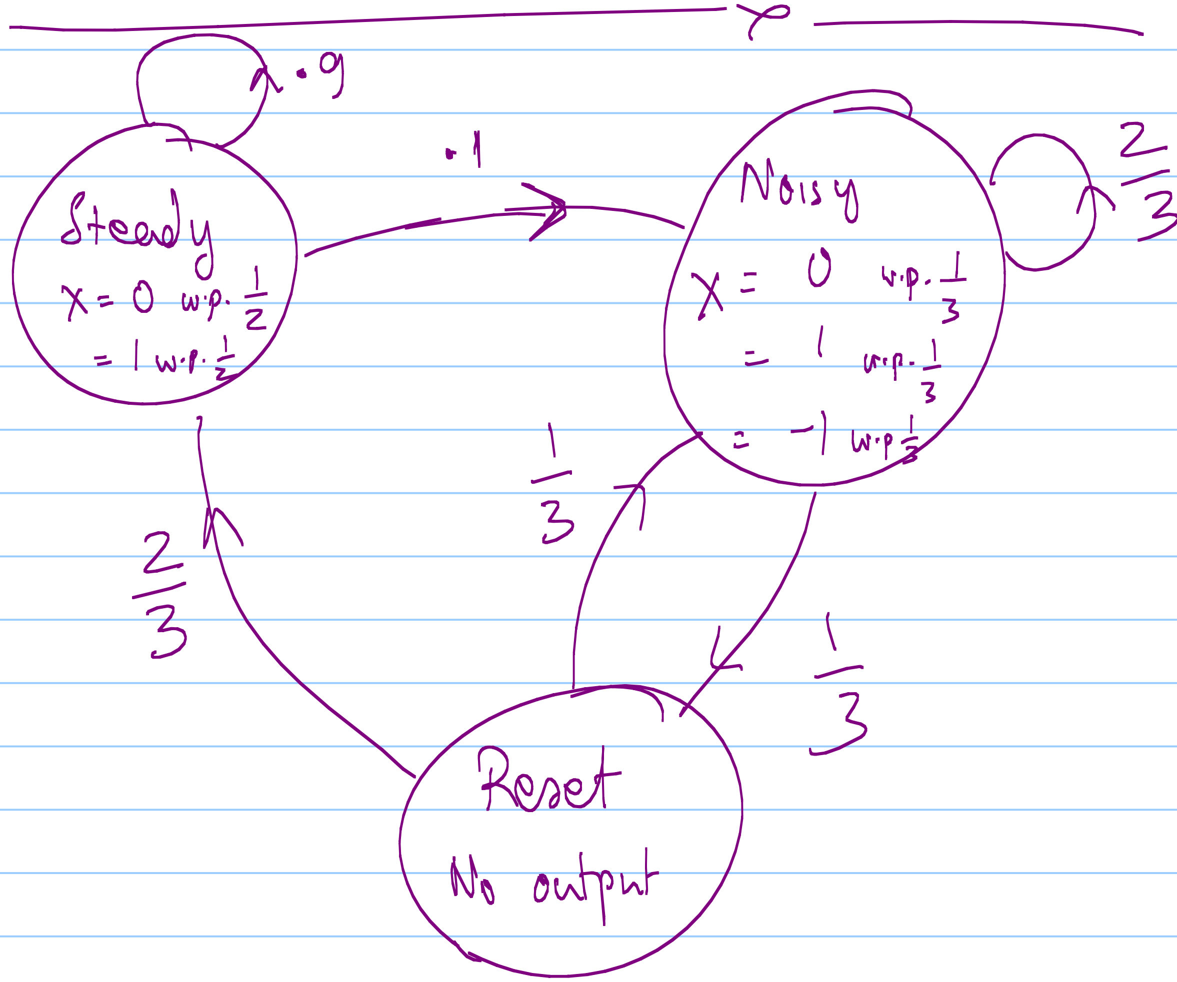
ST06 - LECTURE 05

Note Title 2/21/2006

Today

- Markov Chains / Markov processes
- Entropy rate of Markov processes.



Produce $X_1, X_2, \ldots X_t$ this way

What is $H(X_1, \dots X_t)$?

Does $\lim_{t \to \infty} \frac{1}{t} H(X_1 \dots X_t)$ exist?

Terminology

- **Stochastic Process**: $X_1, X_2, \dots X_n, \dots$
  for $n = 1 \dots$
  $(X_1 \dots X_n) \sim p(x_1 \dots x_n)$ is a stochastic process.
  (just a sequence of non-independent random variables).

- Stochastic Process is **stationary** if
  $\forall\, n, l, x_1 \dots x_n$
  $$\Pr[X_1 = x_1, \dots X_n = x_n] = \Pr[X_{l+1} = x_1 \dots X_{l+n} = x_n].$$

- Stochastic Process $X_1 \dots X_n \dots$ is a Markov process (Markov chain) if

$\forall\, n,\ x_1 \dots x_n$

$$\Pr\left[X_n = x_n \mid X_1 = x_1, \dots X_{n-1} = x_{n-1}\right] = \Pr\left[X_n = x_n \mid X_{n-1} = x_{n-1}\right]$$

Examples: Model of Web-browsing

At any page i, hit forward button to page j with probability $(1-\alpha_i) P_{ij}$

& hit back button with probability α_i

X_t = page viewed at time t.

Is $X_1 \dots X_t$ a stochastic process?

Is it Markov?

Is it stationary?

Let Z_t = contents of history stack of browser.

$Z_1 \ldots Z_t$ = Markov chain?

Like finitely specified stochastic processes.

Are Markov chains finitely specified?

Not necessarily

- domain of X_n need not be <u>finite</u>
- $p(x_n | x_{n-1})$ may vary with time.

But if we restrict both, then get large class of succinctly specified stochastic processes that model many sources of info; & many sources of noise.

Today : (finite state, time invariant, discrete time)

MARKOV CHAINS.

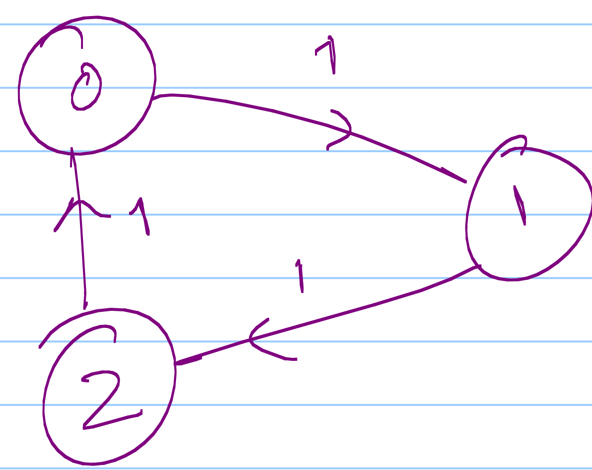
X_n : state at time n.

P_{ij} : prob $(X_n = j \mid X_{n-1} = i)$: Transition Prob. Matrix.

X_0 = initial state.

Are all Markov chains stationary?

E.g.



0, 1, 2, 0, 1, 2, 0, 1, 2, 0, 1, 2 . . .

Not stationary for sure!

But can start it of nicely so that it is a stationary process.

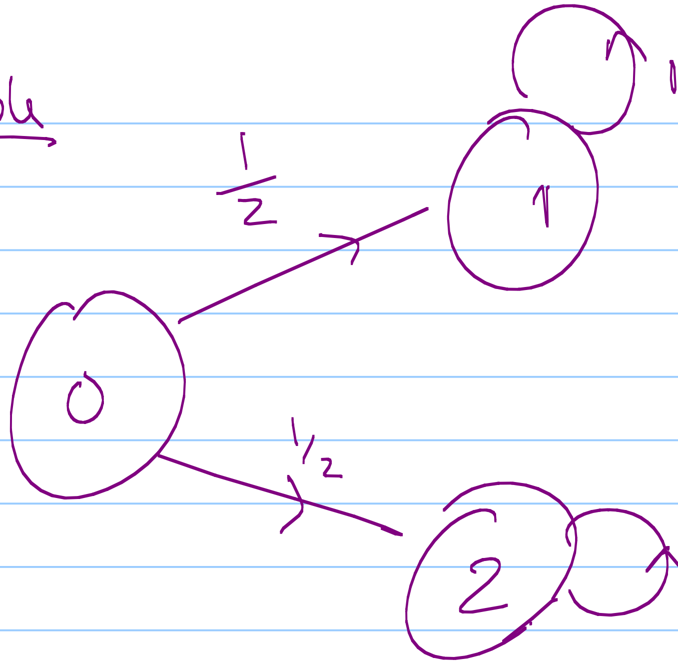
$$X_0 = \begin{cases} 0 & \text{w.p. } \frac{1}{3} \\ 1 & \text{w.p. } \frac{1}{3} \\ 2 & \text{w.p. } \frac{1}{3} \end{cases}$$

Then for every n $X_n = 0/1/2$ w.p. $\frac{1}{3}$

<u>Fact</u>: Every Markov chain has a stationary distribution.

But is it unique?

Example



$X_0 = 1$ leads to $1, 1, 1, 1,$ as stat process

$X_0 = 2$ leads to $2, 2, \ldots$ as stat. process

- To make stationary dist. unique need more.

- M.C. is "strongly connected" or "irreducible" if $\exists$ path from state i to state j of positive prob. for every i, j.

("Reducible" if P can be drawn as

$$
\begin{array}{c|cc}
 & 1 \; \cdots \; k & k+1 \; \cdots \; n \\
\hline
1 \\ \vdots \\ k & A & B \\
k+1 \\ \vdots \\ n & 0 & C
\end{array}
$$
)

<u>Theorem</u>: Every irreducible Markov chain has a unique stationary distribution π

<u>Proof</u>: Out of scope for this class ... but a few words where it comes from.

Called the <u>Perron-Frobenius</u> Theorem

Says every irreducible non-negative matrix M has unique π, λ s.t. $\pi M = \lambda \cdot \pi$

Every other singular value of M is strictly smaller in magnitude

---

<u>Opening example:</u>

$$P = \begin{bmatrix} .9 & .1 & 0 \\ 0 & 2/3 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix}$$

$$(\pi_1, \pi_2, \pi_3) = \left(\frac{20}{32}, \frac{9}{32}, \frac{3}{32} \right)$$

So suppose we watch the states of a Markov chain. What is the entropy rate?

More generally for a Stochastic Process?

<u>Defn 1</u>: $$H(\mathcal{X}) = \lim_{t \to \infty} \frac{1}{t} H(X_1 \ldots X_t)$$
if the limit exists.

<u>Defn 2</u>: $$H'(\mathcal{X}) = \lim_{t \to \infty} H(X_t \mid X_1 \ldots X_{t-1})$$

<u>Theorem</u>: for stationary process $\mathcal{X}$, the limits $H(\mathcal{X})$ & $H'(\mathcal{X})$ exist & are equal.

Proof: Note first that if

(A) — $\lim_{t \to \infty} H(X_t | X_1 \dots X_{t-1})$ exists

Thus $\lim_{t \to \infty} \frac{1}{t} \sum H(X_t | X_1 \dots X_{t-1})$ also

exist & equals former.

So suffices to show (A) exists.

Claim:

$$H(X_t | X_1 \dots X_{t-1}) \leq H(X_{t-1} | X_1 \dots X_{t-2})$$

Why?

$$H(X_t | X_1 \dots X_{t-1}) \leq H(X_t | X_2 \dots X_{t-1})$$

(conditioning reduces entropy)

$$= H(X_{t-1} | X_1 \dots X_{t-2}) \quad [\ell = 1 \text{ shift}]$$

AEP for stationary ergodic processes

$$\lim \frac{1}{n} \log p(X_1 \dots X_n) \longrightarrow H(\mathcal{X})$$

Thm. Let $\mathcal{X}$ be a Markov chain with stationary dist. μ & p.t.m. P

Then $H(\mathcal{X}) = -\sum \mu_i P_{ij} \log P_{ij}$

Proof:

$$H(\mathcal{X}) = H(X_2 | X_1)$$

$$= \sum_i \mu_i \, H(X_2 | X_1 = i)$$

$$= -\sum_i \sum_j \mu_i P_{ij} \log P_{ij}$$

Example

$$P = \begin{bmatrix} .9 & .1 & 0 \\ 0 & 2/3 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix}$$

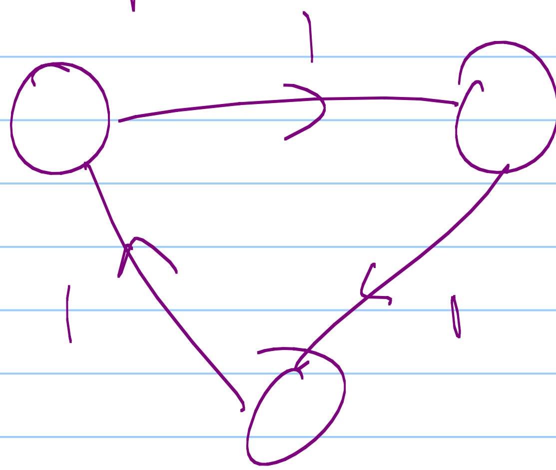
$$\mu = \left(\frac{20}{32}, \frac{9}{32}, \frac{3}{32}\right)$$

$$H(\mathfrak{X}) = -\frac{5}{8}\left(.9\log .9 + .1\log .1\right)$$

$$+ \frac{3}{8}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)$$

$$\approx$$

Another example

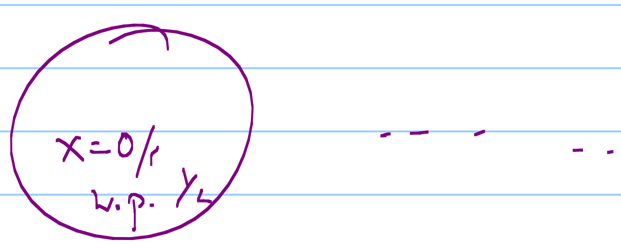


Entropy rate = ?

Back to original example:

States were just some intermediate step.

Real interest was in X's generated.



Entropy rate of $\mathcal{X}$ = ?

To study this study "Hidden Markov Chains"

if $X_1 \to X_2 \to X_3 \ldots X_n \to$

defines a Markov chain

& $Y_i = \phi(X_i)$ is some process

then we refer to $Y_1 \ldots Y_n \ldots$ as a hidden M.C.

What is the entropy rate of $\mathcal{Y}$?

Observation 1: $Y_1, Y_2 \ldots Y_n \ldots$ is stationary.

if $X_1 \ldots X_n \ldots$ is stationary.

So $H(\mathcal{Y})$ exists & equals $H'(\mathcal{Y})$ . . .

furthermore $H(\mathcal{Y}) \leq H(Y_n \mid Y_1 \ldots Y_{n-1})$

$$\leq H(Y_{n-1} \mid Y_1 \ldots Y_{n-2})$$

$$\vdots$$

But $H(\mathcal{Y}) \geq H(Y_n \mid Y_2, \ldots, Y_{n-1}, X_1)$

Proof $H(Y_n \mid Y_2 \ldots Y_{n-1}, X_1)$

$$= H(Y_n \mid Y_2 \ldots Y_{n-1}, X_1, X_0, X_{-1} \ldots X_{-k})$$

$$= H(Y_n \mid Y_2 \ldots Y_{n-1}, X_1 \ldots X_{-k}, Y_1 \ldots Y_{-k})$$

$$= H(Y_{n+k} \mid Y_2 \ldots Y_{n+k-2}, X_1, X_2 \ldots X_{k+2})$$

$$\leq H(Y_{n+k} \mid Y_2 \ldots Y_{n+k-2}) \leq H(\mathcal{Y}) + \epsilon$$

Conclude

$$H(Y_n | Y_{n-1} \dots Y_1, X_1) \le H(\mathcal{Y}) \le H(Y_n | Y_{n-1} \dots Y_1)$$

Does this ?

This we know converges

lets look at diff

$$\epsilon_n = H(Y_n | Y_{n-1} \dots Y_2, Y_1) - H(Y_n | Y_{n-1} \dots Y_1, X_1)$$

$$= I(X_1 ; Y_n | Y_{n-1}, \dots Y_1)$$

But $$\sum_{n=1}^{N} I(X_1 ; Y_n | Y_1 \dots Y_{n-1})$$

$$= I(X_1 ; (Y_1 \dots Y_N)) \le H(X_1)$$

So the sequence ϵ_n is non-negative & has finite sum & so converges to 0.

But what is Entropy rate of the Markov Chain we started with ?

I don't know ....