Entropy Rates of Hidden Markov Processes emerge from Blackwell's Trapdoor Channel

Paul Cuff
(Trapdoor Channel work with Haim Permuter,
Benjamin Van Roy, and Tsachy Weissman)

Stanford University

BIRS
October 2, 2007

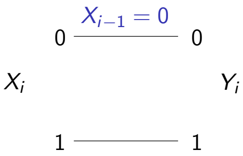
Focus of Talk

1. Blackwell's Trapdoor Channel Feedback Capacity
   - Numerical Calculations (and "chemical channel" generalization)
   - Analytic Solution
   - Zero-error Communication Scheme

2. Hidden-Markov Processes
   - Not So Hidden Markov Processes

Channel Capacity Puzzle

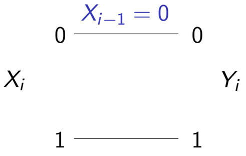
What is the capacity of this two-state channel where the state is a function of the previous input?

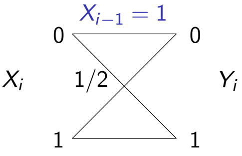


(Observation: Markov inputs lead to Hidden-Markov channel outputs.)

Channel Capacity Puzzle

What is the capacity of this two-state channel where the state is a function of the previous input?



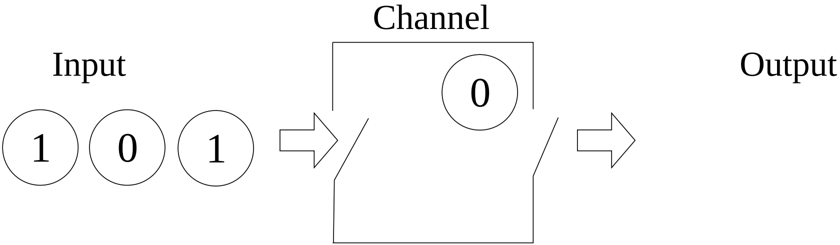


Alternative Characterization:

$$\begin{aligned} Z_i &\sim \text{i.i.d. Bern}(1/2), \\ Z^n &\perp X^n, \\ Y_i &= X_i + (X_{i-1} Z_i) \bmod 2. \end{aligned}$$

(Observation: Markov inputs lead to Hidden-Markov channel outputs.)

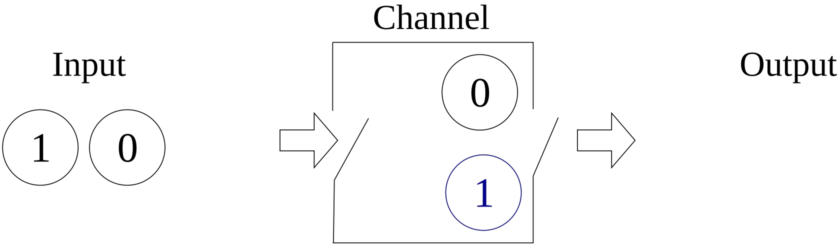
The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$s_0 = 0$

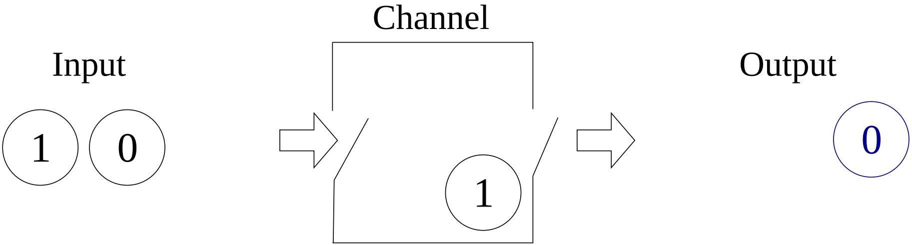
The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$s_0 = 0$
$x_1 = 1,$

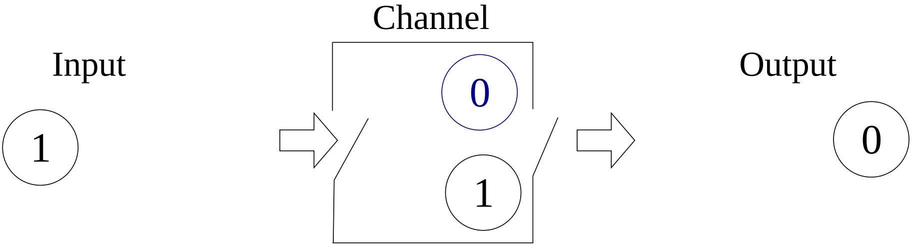
The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$s_0 = 0$
$x_1 = 1,\ s_1 = 1,\ y_1 = 0,$

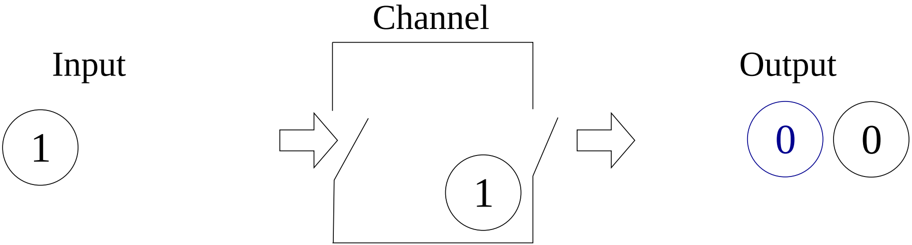
The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$s_0 = 0$
$x_1 = 1,\ s_1 = 1,\ y_1 = 0,$
$x_2 = 0,$

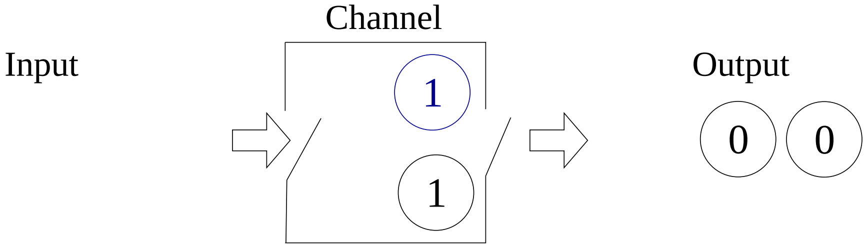
The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$s_0 = 0$
$x_1 = 1,\ s_1 = 1,\ y_1 = 0,$
$x_2 = 0,\ s_2 = 1,\ y_2 = 0,$

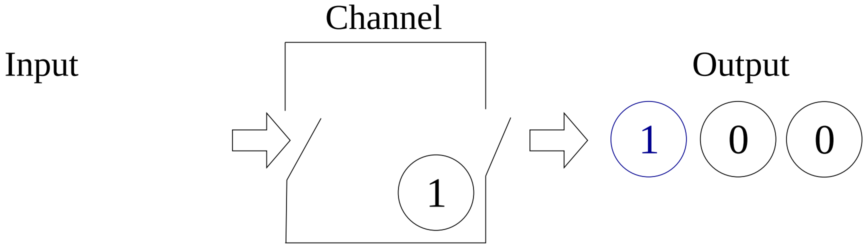
The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$s_0 = 0$
$x_1 = 1,\ s_1 = 1,\ y_1 = 0,$
$x_2 = 0,\ s_2 = 1,\ y_2 = 0,$
$x_3 = 1,$

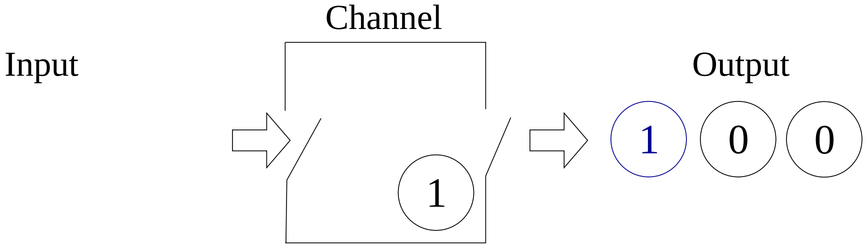
The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$s_0 = 0$
$x_1 = 1,\ s_1 = 1,\ y_1 = 0,$
$x_2 = 0,\ s_2 = 1,\ y_2 = 0,$
$x_3 = 1,\ s_3 = 1,\ y_3 = 1.$

The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$s_0 = 0$
$x_1 = 1,\ s_1 = 1,\ y_1 = 0,$
$x_2 = 0,\ s_2 = 1,\ y_2 = 0,$
$x_3 = 1,\ s_3 = 1,\ y_3 = 1.$

Biochemical Interpretation [Berger 71]

The Chemical Channel

Balls are not equally likely to exit the channel.

x_t	s_{t-1}	$p(y_t = 0 \mid x_t, s_{t-1})$	$p(y_t = 1 \mid x_t, s_{t-1})$
0	0	1	0
0	1	p_1	$1 - p_1$
1	0	p_2	$1 - p_2$
1	1	0	1

Special cases:

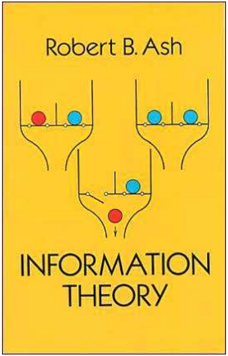
Trapdoor channel: $p_1 = p_2 = \frac{1}{2}$.

New vs. Old: $p_1 = 1 - p_2 = p_{switch}$.

'0' vs. '1': $p_1 = p_2 = p_{zero}$.

The Trapdoor Channel

Introduced by David Blackwell in 1961.
[Ash 65], [Ahlswede & Kaspi 87], [Ahlswede 98], [Kobayashi 02 & 03].



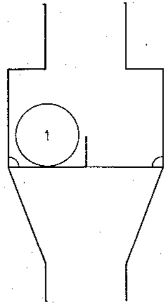


Fig. 7.1 A simple two-state channel.

A "simple two-state channel." - Blackwell

Ideas for Communication without Feedback

- Repeat each bit three time: $R = 1/3$ bit.

Ideas for Communication without Feedback

- Repeat each bit three time: $R = 1/3$ bit.
- Repeat each bit twice: $R = 1/2$ bit. [Ahlswede & Kaspi 87]

Ideas for Communication without Feedback

- Repeat each bit three time: $R = 1/3$ bit.
- Repeat each bit twice: $R = 1/2$ bit. [Ahlswede & Kaspi 87]
- $C \approx 0.572$ bits per channel use. [Kobayashi & Morita 03]

Communication Setting (with feedback)

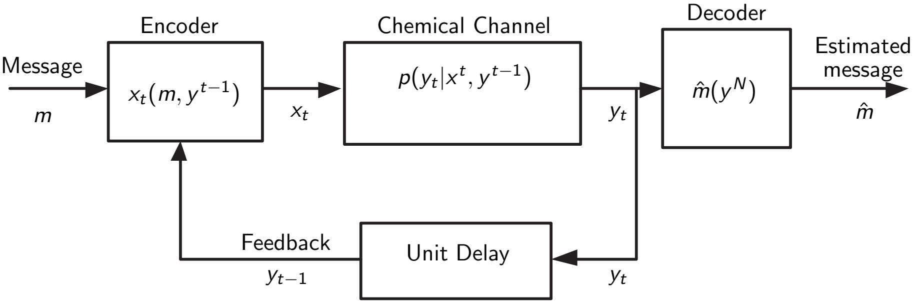


Figure: Communication with feedback

Feedback Capacity of FSC

Lower and upper bounds:

$$
\begin{aligned}
C_{FB} &\geq \lim_{N\to\infty} \frac{1}{N} \max_{\{p(x_i|x^{i-1},y^{i-1})\}_{i=1}^{N}} \min_{s_0} I(X^N \to Y^N|s_0) \\
C_{FB} &\leq \lim_{N\to\infty} \frac{1}{N} \max_{\{p(x_i|x^{i-1},y^{i-1})\}_{i=1}^{N}} \max_{s_0} I(X^N \to Y^N|s_0)
\end{aligned}
$$

[Permuter, Weissman & Goldmith ISIT06]

Directed Information

Mutual Information

$$I(X^n; Y^n) = \sum_{i=1}^{n} I(X^{n}; Y_i | Y^{i-1})$$

Directed Information was defined by Massey in 1990.

$$I(X^n \to Y^n) \triangleq \sum_{i=1}^{n} I(X^{i}; Y_i | Y^{i-1})$$

Directed Information

Mutual Information

$$I(X^n;Y^n) \;=\; \sum_{i=1}^{n} I(X^{n};Y_i|Y^{i-1})$$

Directed Information was defined by Massey in 1990.

$$I(X^n \rightarrow Y^n) \;\triangleq\; \sum_{i=1}^{n} I(X^{i};Y_i|Y^{i-1})$$

Intuition [Massey 05]:

$$I(X^n;Y^n) \;=\; I(X^n \rightarrow Y^n) + I(Y^{n-1} \rightarrow X^n)$$

Feedback Capacity of Unifilar, Strongly Connected, FSC

Chemical channel has two other properties of interest.

1. Unifilar [Ziv 85]: State is deterministic function of past state, input, and output.
2. Strongly connected: Any state s_t can be reached with positive probability from any other state s_{t-1}.

Consequence

Initial state doesn't matter; upper and lower bounds become equal.

$$C_{FB} = \lim_{N\to\infty} \frac{1}{N} \max_{\{p(x_i|x^{i-1},y^{i-1})\}_{i=1}^N} I(X^N \to Y^N)$$

Feedback Capacity of Unifilar, Strongly Connected, FSC

$$
\begin{aligned}
C_{FB} &= \lim_{N\to\infty} \frac{1}{N} \max_{\{p(x_t|x^{t-1},y^{t-1})\}_{t=1}^N} I(X^N \to Y^N) \\
&= \lim_{N\to\infty} \frac{1}{N} \max_{\{p(x_t|x^{t-1},y^{t-1})\}_{t=1}^N} \sum_{t=1}^{N} I(X^t; Y_t|Y^{t-1}) \\
&= \lim_{N\to\infty} \frac{1}{N} \max_{\{p(x_t|s_{t-1},y^{t-1})\}_{t=1}^N} \sum_{t=1}^{N} I(X_t, S_{t-1}; Y_t|Y^{t-1}) \\
&= \sup_{\{p(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} I(X_t, S_{t-1}; Y_t|Y^{t-1})
\end{aligned}
$$

Dynamic Programming (infinite horizon, average reward)

Variable Assignments

State: $\beta_t = p(s_t|y^t)$

Action: $u_t = p(x_t|s_{t-1})$

Disturbance: $w_t = y_{t-1}$

Dynamic Programming Requirements

State evolution:

$$\beta_t = F(\beta_{t-1}, u_t, w_t)$$

Reward function per unit time:

$$g(\beta_{t-1}, u_t) = I(X_t, S_{t-1}; Y_t|\beta_{t-1})$$

Similar work: [Yang, Kavčić & Tatikonda 05], [Chen & Berger 05]

Dynamic Programming (infinite horizon, average reward)

Dynamic Programming Operator T

The dynamic programming operator T is given by

$$T \circ J(\beta) = \sup_{u \in \mathcal{U}} \left(g(\beta, u) + \int P_w(dw|\beta, u) J(F(\beta, u, w)) \right).$$

Bellman Equation

If there exist a function $J(\beta)$ and constant ρ that satisfy

$$J(\beta) = T \circ J(\beta) - \rho$$

then ρ is the optimal infinite horizon average reward.

Feedback Capacity of Chemical Channel (20 iterations)

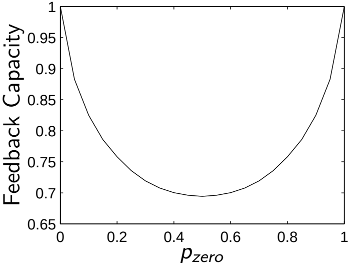


Figure: Feedback capacity of chemical channel as functions of two parameters.

Trapdoor channel feedback capacity found at $p_{zero} = 0.5$ and $p_{switch} = 0.5$.

Coincidence

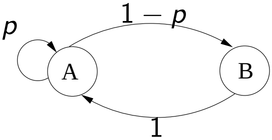
Trapdoor channel: $C_{FB} \approx 0.694$ bits

Coincidence

Trapdoor channel: $C_{FB} \approx 0.694$ bits

<u>Homework Question</u>

Entropy rate. Find the maximum entropy rate of the following two-state Markov chain:

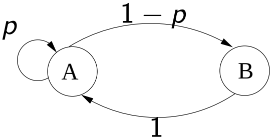


Coincidence

Trapdoor channel: $C_{FB} \approx 0.694$ bits

Homework Question

Entropy rate. Find the maximum entropy rate of the following two-state Markov chain:



Solution: (Golden Ratio: $\phi = \frac{\sqrt{5}+1}{2}$)

$$p^\star = \phi - 1 = \frac{1}{\phi}$$

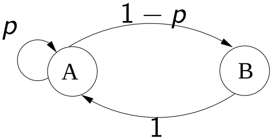
$$H(\mathcal{X}) = \log \phi = 0.6942\ldots \text{ bits}$$

Coincidence

Trapdoor channel: $C_{FB} \approx$ 0.694 bits

<u>Homework Question</u>

Entropy rate. Find the maximum entropy rate of the following two-state Markov chain:



Solution: (Golden Ratio: $\phi = \frac{\sqrt{5}+1}{2}$)

$p^\star = \phi - 1 = \frac{1}{\phi}$

$H(\mathcal{X}) = \log \phi =$ 0.6942... bits

Decodable Input

Rename channel input: $\tilde{x}_t = x_t \oplus s_{t-1}$.
Rename channel output: $\tilde{y}_t = y_t \oplus y_{t-1}$.

Decodable Input

Rename channel input: $\tilde{x}_t = x_t \oplus s_{t-1}$.
Rename channel output: $\tilde{y}_t = y_t \oplus y_{t-1}$.

Case $\tilde{x}_t = 0$

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

Decodable Input

Rename channel input: $\tilde{x}_t = x_t \oplus s_{t-1}$.
Rename channel output: $\tilde{y}_t = y_t \oplus y_{t-1}$.

Case $\tilde{x}_t = 0$

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

Proof: $x_t = s_{t-1} = y_t = s_t$

Decodable Input

Rename channel input: $\tilde{x}_t = x_t \oplus s_{t-1}$.
Rename channel output: $\tilde{y}_t = y_t \oplus y_{t-1}$.

Case $\tilde{x}_t = 0$

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

Proof: $x_t = s_{t-1} = y_t = s_t$

$$x_{t-1} \oplus s_{t-2} \quad = \quad y_{t-1} \oplus s_{t-1}$$

Decodable Input

Rename channel input: $\tilde{x}_t = x_t \oplus s_{t-1}$.
Rename channel output: $\tilde{y}_t = y_t \oplus y_{t-1}$.

Case $\tilde{x}_t = 0$

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

Proof: $x_t = s_{t-1} = y_t = s_t$

$$\begin{aligned} x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus s_{t-1} \\ x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus y_t \end{aligned}$$

Decodable Input

Rename channel input: $\tilde{x}_t = x_t \oplus s_{t-1}$.
Rename channel output: $\tilde{y}_t = y_t \oplus y_{t-1}$.

Case $\tilde{x}_t = 0$

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

Proof: $x_t = s_{t-1} = y_t = s_t$

$$\begin{aligned} x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus s_{t-1} \\ x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus y_t \end{aligned}$$

Case $\tilde{x}_t = 1$

$\tilde{x}_t = 1 \quad \Rightarrow \quad \tilde{y}_{t-1} \sim$ Bern(1/2), independent of $\tilde{x}_t$ and the past

Decodable Input

Rename channel input: $\tilde{x}_t = x_t \oplus s_{t-1}$.
Rename channel output: $\tilde{y}_t = y_t \oplus y_{t-1}$.

Case $\tilde{x}_t = 0$

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

Proof: $x_t = s_{t-1} = y_t = s_t$

$$\begin{aligned} x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus s_{t-1} \\ x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus {\color{red} y_t} \end{aligned}$$

Case $\tilde{x}_t = 1$

$$\tilde{x}_t = 1 \quad \Rightarrow \quad \tilde{y}_{t-1} \sim \text{Bern}(1/2), \text{ independent of } \tilde{x}_t \text{ and the past}$$

This is like the puzzle at the beginning (don't repeat 1's into the channel).

Decoding Example

Decoding rules

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

$$\tilde{x}_t = 1 \quad \Rightarrow \quad \tilde{x}_{t-1} = 0$$

$\tilde{x}^n$: $\quad\quad\quad\quad 0$

$\tilde{y}^n$: 1 1 1 1 0 0 1

Decoding Example

Decoding rules

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

$$\tilde{x}_t = 1 \quad \Rightarrow \quad \tilde{x}_{t-1} = 0$$

$\tilde{x}^n$: 0 0

$\tilde{y}^n$: 1 1 1 1 0 0 1

Decoding Example

Decoding rules

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

$$\tilde{x}_t = 1 \quad \Rightarrow \quad \tilde{x}_{t-1} = 0$$

$\tilde{x}^n$: 1 0 0

$\tilde{y}^n$: 1 1 1 1 0 0 1

Decoding Example

Decoding rules

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$

$$\tilde{x}_t = 1 \quad \Rightarrow \quad \tilde{x}_{t-1} = 0$$

$\tilde{x}^n$: 0 1 0 0

$\tilde{y}^n$: 1 1 1 1 0 0 1

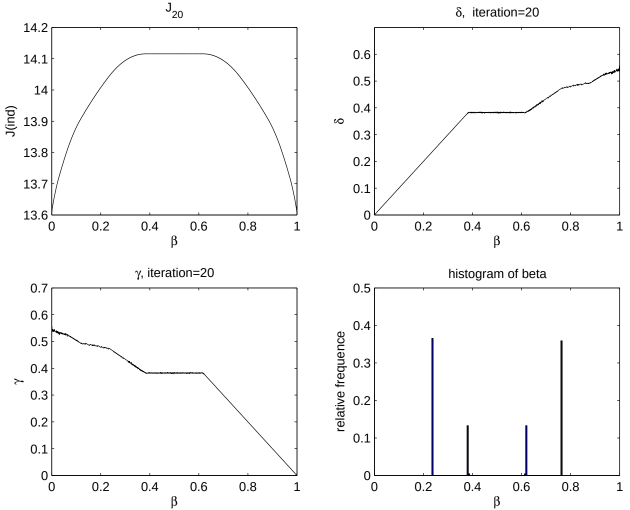
Decoding Example

Decoding rules

$$\tilde{x}_t = 0 \quad \Rightarrow \quad \tilde{x}_{t-1} = \tilde{y}_t$$
$$\tilde{x}_t = 1 \quad \Rightarrow \quad \tilde{x}_{t-1} = 0$$

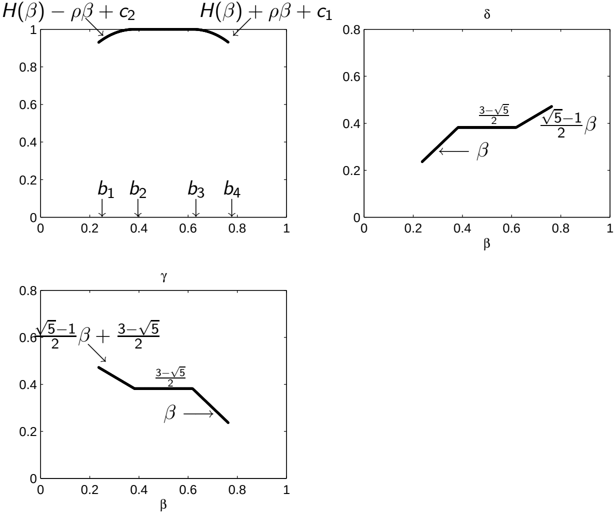
$\tilde{x}^n$: 1 0 1 0 0
$\tilde{y}^n$: 1 1 1 1 0 0 1

Dynamic Programming 20th Value iteration



Bellman Equation: $J(\beta) = T \circ J(\beta) - \rho$.

Conjectured Solution to the Bellman Equation



Bellman Equation: $J(\beta) = T \circ J(\beta) - \rho$.

Proven Feedback Capacity

Bellman Equation is satisfied.

Trapdoor channel feedback capacity:

$$C_{FB} = \log \phi = 0.6942\ldots \text{ bits}$$

A Zero-error Communication Scheme

$\tilde{x}^n$	flag	index
0010100010100101		

1. Message maps to unique sequence without repeating 1's.

A Zero-error Communication Scheme

$\tilde{x}^{n+1}$	flag	index
00101000101001010		

1. Message maps to unique sequence without repeating 1's.
2. Concatenate with 0.

A Zero-error Communication Scheme

$\tilde{x}^{n+1}$	flag	index
0010100010100101**0**	xxx	

1. Message maps to unique sequence without repeating 1's.
2. Concatenate with 0.
3. Indicate if "inconsistency" was observed.

A Zero-error Communication Scheme

$\tilde{x}^{n+1}$	flag	index
0010100010100101**0**	xxx	0100101000

1. Message maps to unique sequence without repeating 1's.
2. Concatenate with 0.
3. Indicate if "inconsistency" was observed.
4. If inconsistency exists, send index of inconsistency.

A Zero-error Communication Scheme

$\tilde{x}^{n+1}$	flag	index
00101000101001010	xxx	0100101000
n+1	3	$\log n / C_{FB}$

1. Message maps to unique sequence without repeating 1's.
2. Concatenate with 0.
3. Indicate if "inconsistency" was observed.
4. If inconsistency exists, send index of inconsistency.

A Zero-error Communication Scheme

$\tilde{x}^{n+1}$	flag	index
0010100010100101**0**	xxx	0100101000
n+1	3	$\log n / C_{FB}$

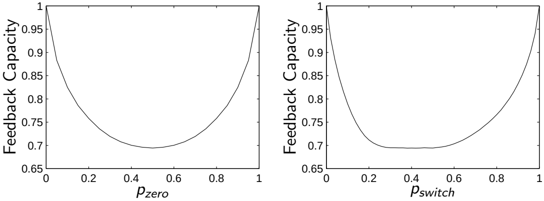
Number of messages

How many binary sequences of length n without repeating 1's?
Fibonachi sequence: $f_n \doteq \phi^n$.

1. Message maps to unique sequence without repeating 1's.
2. Concatenate with 0.
3. Indicate if "inconsistency" was observed.
4. If inconsistency exists, send index of inconsistency.

Feedback Capacity Summary

- Chemical Channel



- Trapdoor channel

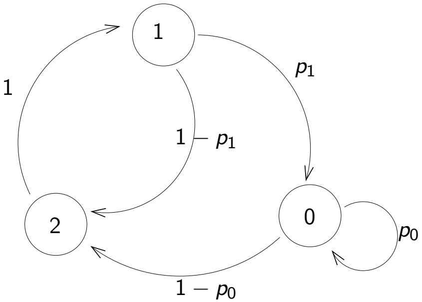
$$C_{FB} = \log \phi = 0.6942\ldots \text{ bits}$$

- Zero-error communication scheme

$\tilde{x}^{n+1}$	flag	index
0010100010100101**0**	xxx	0100101000

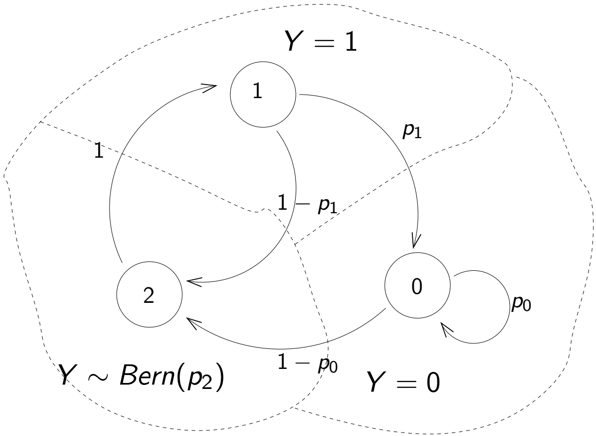
Markov Process X_n

(related to the capacity achieving input to the trapdoor channel)



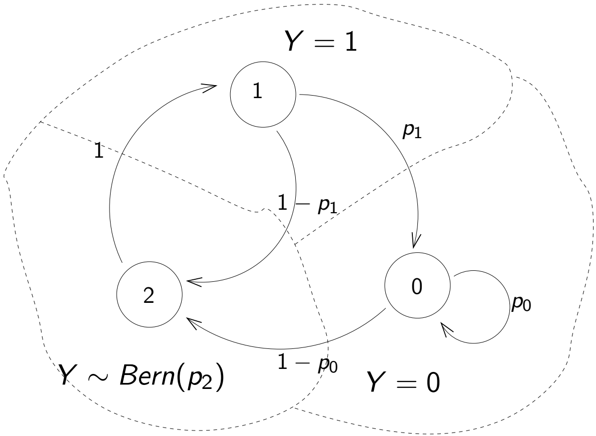
Hidden Markov Process Y_n

(related to the output of the trapdoor channel)



Hidden Markov Process Y_n

(related to the output of the trapdoor channel)



$H(\mathcal{Y}) = \mu_0 H(p_0) + \mu_1 H(p_1) + \mu_2 H(p_2)$, where μ is the stationary distribution of X_n ($\mu_1 = \mu_2 = (1 - p_0)/(2(1 - p_0) + p_1), \mu_0 = 1 - \mu_1 - \mu_2$).

Observable Underlying Markov Process

Let X_n be a discrete, time-invariant Markov process and Y_n be a memoryless, time-invariant hiding of X_n (i.e. $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$), and define the conditional entropy rate

$$H(\mathcal{X}|\mathcal{Y}) \triangleq \lim_{n\to\infty} \frac{1}{n} H(X^n|Y^n).$$

Observation 1

$$H(\mathcal{X}|\mathcal{Y}) = 0 \implies H(\mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}).$$

Observable Underlying Markov Process

Let X_n be a discrete, time-invariant Markov process and Y_n be a memoryless, time-invariant hiding of X_n (i.e. $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$), and define the conditional entropy rate

$$H(\mathcal{X}|\mathcal{Y}) \quad \triangleq \quad \lim_{n\to\infty} \frac{1}{n} H(X^n|Y^n).$$

Observation 1

$$H(\mathcal{X}|\mathcal{Y}) = 0 \quad \Longrightarrow \quad H(\mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}).$$

Theorem 1

$$H(\mathcal{X}|\mathcal{Y}) = 0 \quad \Longleftrightarrow \quad H(X_1^k|Y_1^k, X_0, X_{k+1}) = 0 \text{ for all } k \in \mathcal{N}.$$

Proof (part a) of Theorem 1

"$\Leftarrow$"

$$
\begin{aligned}
H(\mathcal{X}|\mathcal{Y}) &= \lim_{n\to\infty} \frac{1}{n} H(X^n|Y^n) \\
&= \lim_{n\to\infty} \frac{1}{n} \left[H(X_1^n|Y_1^n, X_0, X_{n+1}) + I(X_1^n; X_0, X_{n+1}|Y_1^n)\right] \\
&= \lim_{n\to\infty} \frac{1}{n} I(X_1^n; X_0, X_{n+1}|Y_1^n) \\
&\leq \lim_{n\to\infty} \frac{1}{n} \left[H(X_0) + H(X_{n+1})\right] \\
&= 0.
\end{aligned}
$$

Proof (part b) of Theorem 1

"$\Rightarrow$"

$$H(X^n|Y^n)$$

X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y
X	Y

Figure: Hidden Markov Process

Proof (part b) of Theorem 1

"$\Rightarrow$"

Conditioning on more decreases entropy.

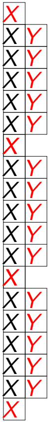


Figure: Hidden Markov Process

Proof (part b) of Theorem 1

"⇒"

Markovity allows us to separate and add the blocks.

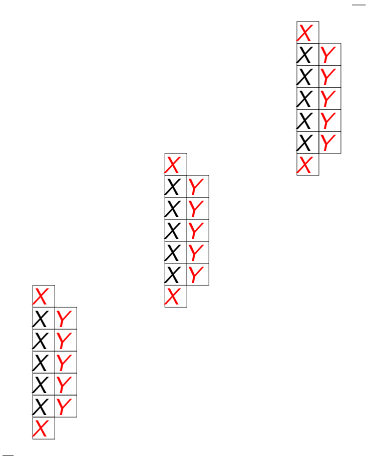


Figure: Hidden Markov Process

Proof (part b) of Theorem 1

"$\Rightarrow$"

Stationarity allows time shifts.

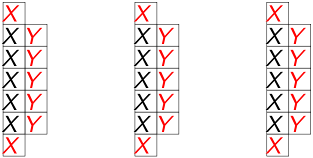


Figure: Hidden Markov Process

Proof (part b) of Theorem 1

"$\Rightarrow$"

$$
\begin{aligned}
H(X_1^k|Y_1^k, X_0, X_{k+1}) &= \frac{1}{n}\sum_{i=1}^{n} H(X_{(i-1)k'+1)}^{ik'-1}|Y_{(i-1)k'+1}^{ik'-1}, X_{(i-1)k'}, X_{ik'}) \\
&= \frac{1}{n} H(X_1^{nk'}|Y_1^{nk'}, \{X_{ik'}\}_{i=0}^{n}) \\
&= \lim_{n\to\infty} \frac{1}{n} H(X_1^{nk'}|Y_1^{nk'}, \{X_{ik'}\}_{i=0}^{n}) \\
&\leq k' \lim_{n\to\infty} \frac{1}{nk'} H(X_1^{nk'}|Y_1^{nk'}) \\
&= k' H(\mathcal{X}|\mathcal{Y}) \\
&= 0,
\end{aligned}
$$

where $k' = k + 1$.

Other Questions

- How many states are necessary to construct a non-trivial HMP that satisfies $H(X_1^k|Y_1^k, X_0, X_{k+1})$ for all k?

- Does $H(X_1^k|Y_1^k, X_0, X_{k+1})$ for all k imply that either $H(X_1|Y_1, X_0) = 0$ or $H(X_1|Y_1, X_2) = 0$?

Other Questions

- How many states are necessary to construct a non-trivial HMP that satisfies $H(X_1^k|Y_1^k, X_0, X_{k+1})$ for all k?
  - 3
- Does $H(X_1^k|Y_1^k, X_0, X_{k+1})$ for all k imply that either $H(X_1|Y_1, X_0) = 0$ or $H(X_1|Y_1, X_2) = 0$?
  - No