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?

\[
\begin{array}{c c c c c c}
X_{i-1} = 0 & 0 \\
X_i & Y_i \\
1 & 1 \\
\end{array}
\]

\[
\begin{array}{c c c c c c}
X_{i-1} = 1 & 0 \\
X_i & 1/2 \\
1 & 1 \\
\end{array}
\]

(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?

\[
\begin{array}{cccc}
0 & X_{i-1} = 0 & 0 & \text{0} \\
X_i & Y_i & X_i & Y_i \\
0 & 1/2 & 0 & 1 \\
1 & 1 & 1 & 1
\end{array}
\]

Alternative Characterization:

\[
Z_i \sim \text{i.i.d. Bern}(1/2),
\]
\[
Z^n \perp X^n,
\]
\[
Y_i = X_i + (X_{i-1}Z_i) \mod 2.
\]

(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

\[ s_t = s_{t-1} + x_t - y_t \]

\[ s_0 = 0 \]
\[ x_1 = 1, \quad s_1 = 1, \quad y_1 = 0, \]
\[ x_2 = 0, \]
The Trapdoor Channel

\[ s_t = s_{t-1} + x_t - y_t \]

\[ s_0 = 0 \]
\[ x_1 = 1, \quad s_1 = 1, \quad y_1 = 0, \]
\[ x_2 = 0, \quad s_2 = 1, \quad y_2 = 0, \]
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.

\[
\begin{array}{|c|c|c|c|}
\hline
x_t & s_{t-1} & p(y_t = 0|x_t, s_{t-1}) & p(y_t = 1|x_t, s_{t-1}) \\
\hline
0 & 0 & 1 & 0 \\
0 & 1 & p_1 & 1 - p_1 \\
1 & 0 & p_2 & 1 - p_2 \\
1 & 1 & 0 & 1 \\
\hline
\end{array}
\]

Special cases:

- Trapdoor channel: \( p_1 = p_2 = \frac{1}{2} \).
- New vs. Old: \( p_1 = 1 - p_2 = p_{\text{switch}} \).
- ‘0’ vs. ‘1’: \( p_1 = p_2 = p_{\text{zero}} \).
The Trapdoor Channel

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

A “simple two-state channel.” - Blackwell
Ideas for Communication without Feedback

- Repeat each bit three times: $R = 1/3$ 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 times: $R = 1/3$ 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:**

\[
C_{FB} \geq \lim_{N \to \infty} \frac{1}{N} \max_{\{p(x_i|x_{i-1},y_{i-1})\}_{i=1}^N} \min_{s_0} I(X^N \to Y^N | s_0)
\]

\[
C_{FB} \leq \lim_{N \to \infty} \frac{1}{N} \max_{\{p(x_i|x_{i-1},y_{i-1})\}_{i=1}^N} \max_{s_0} I(X^N \to Y^N | s_0)
\]

[Permuter, Weissman & Goldsmith 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.

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

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

\[ C_{FB} = \lim_{N \to \infty} \frac{1}{N} \max_{t=1}^{N} \left\{ p\left(x_t | x^{t-1}, y^{t-1}\right) \right\} \]

\[ = \lim_{N \to \infty} \frac{1}{N} \max_{t=1}^{N} \sum_{t=1}^{N} I(X^t; Y_t | Y^{t-1}) \]

\[ = \lim_{N \to \infty} \frac{1}{N} \max_{t=1}^{N} \sum_{t=1}^{N} I(X_t, S_{t-1}; Y_t | Y^{t-1}) \]

\[ = \sup_{\left\{ p\left(x_t | s_{t-1}, y^{t-1}\right) \right\} \geq 1} \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} I(X_t, S_{t-1}; Y_t | Y^{t-1}) \]
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:

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

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

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

**Bellman Equation**

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

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

then $\rho$ 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 $p_{\text{zero}} = 0.5$ and $p_{\text{switch}} = 0.5$. 
Coincidence

Trapdoor channel: \(C_{FB} \approx 0.694\) bits
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:

![Markov Chain Diagram]

\( p \quad 1 - p \)

\( A \quad B \)

\( 1 \)
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:

![Diagram of a two-state Markov chain with states A and B, transition probabilities p and 1-p.]

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

\[
p^* = \phi - 1 = \frac{1}{\phi}
\]

\[
H(\mathcal{X}) = \log \phi = 0.6942... \text{ bits}
\]
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:

\[
\begin{align*}
 p &\quad 1 - p \\
 A &\quad B
\end{align*}
\]

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

\[
p^* = \phi - 1 = \frac{1}{\phi}
\]

\[
H(X) = \log \phi = 0.6942... \text{ bits}
\]
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 $\tilde{x}_t = 0$

$\tilde{x}_t = 0 \implies \tilde{x}_{t-1} = \tilde{y}_t$
Decodable Input

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

<table>
<thead>
<tr>
<th>Case ( \tilde{x}_t = 0 )</th>
<th>Proof: ( x_t = s_{t-1} = y_t = s_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tilde{x}<em>t = 0 ) ( \Rightarrow ) ( \tilde{x}</em>{t-1} = \tilde{y}_t )</td>
<td></td>
</tr>
</tbody>
</table>
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 $\tilde{x}_t = 0$

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

Proof:

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

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

\[
\begin{align*}
x_t &= s_{t-1} = y_t = s_t \\
x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus s_{t-1} \\
x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus y_t
\end{align*}
\]
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 $\tilde{x}_t = 0$

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

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

<table>
<thead>
<tr>
<th>$x_{t-1} \oplus s_{t-2}$</th>
<th>$y_{t-1} \oplus s_{t-1}$</th>
</tr>
</thead>
</table>

Case $\tilde{x}_t = 1$

$$\tilde{x}_t = 1 \implies \tilde{y}_{t-1} \sim \text{Bern}(1/2), \text{ independent of } \tilde{x}_t \text{ and the past}$$
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:
\[
\begin{align*}
    x_t = s_{t-1} &= y_t = s_t \\
    x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus s_{t-1} \\
    x_{t-1} \oplus s_{t-2} &= y_{t-1} \oplus y_t
\end{align*}
\]

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

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

\[ \tilde{x}^n: 0 \]
\[ \tilde{y}^n: 1 1 1 1 0 0 1 \]
Decoding Example

Decoding rules

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

\[ \tilde{x}^n: \quad 0 \quad 0 \]
\[ \tilde{y}^n: \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 1 \]
Decoding Example

Decoding rules

\[ \tilde{x}_t = 0 \implies \tilde{x}_{t-1} = \tilde{y}_t \]
\[ \tilde{x}_t = 1 \implies \tilde{x}_{t-1} = 0 \]

\[ \tilde{x}^n: \quad 1 \ 0 \ 0 \]
\[ \tilde{y}^n: \quad 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \]
Decoding Example

Decoding rules

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

\[\tilde{x}^n: \quad 0 \ 1 \ 0 \ 0\]
\[\tilde{y}^n: \quad 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1\]
Decoding Example

Decoding rules

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

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

\( \tilde{x}^n: 1 \ 0 \ 1 \ 0 \ 0 \)

\( \tilde{y}^n: 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \)
Bellman Equation: \( J(\beta) = T \circ J(\beta) - \rho \).
Conjectured Solution to the Bellman Equation

\[ H(\beta) - \rho\beta + c_2 \]

\[ H(\beta) + \rho\beta + c_1 \]

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

<table>
<thead>
<tr>
<th>$\tilde{x}^n$</th>
<th>flag</th>
<th>index</th>
</tr>
</thead>
<tbody>
<tr>
<td>0010100010100101</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Message maps to unique sequence without repeating 1’s.
A Zero-error Communication Scheme

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

\[ \tilde{x}^{n+1} \text{ flag index} \]

00101000101001010
A Zero-error Communication Scheme

\[ \tilde{x}^{n+1} \quad \text{flag} \quad \text{index} \]

00101000101001010 \quad 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

\[ \tilde{x}^{n+1} \quad \text{flag} \quad \text{index} \]

\[
\begin{array}{ccc}
0010100010100101 & xxxx & 0100101000
\end{array}
\]

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

<table>
<thead>
<tr>
<th>$\tilde{x}^{n+1}$</th>
<th>flag</th>
<th>index</th>
</tr>
</thead>
<tbody>
<tr>
<td>00101000101001010</td>
<td>xxx</td>
<td>0100101000</td>
</tr>
<tr>
<td>$n+1$</td>
<td>3</td>
<td>$\log n/C_{FB}$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\tilde{x}^{n+1}$</th>
<th>flag</th>
<th>index</th>
</tr>
</thead>
<tbody>
<tr>
<td>00101000101001010</td>
<td>xxx</td>
<td>0100101000</td>
</tr>
</tbody>
</table>

$n+1$ 3 \log n/C_{FB}

**Number of messages**

How many binary sequences of length $n$ without repeating 1’s?

Fibonacci sequence: $f_n = \phi^n$.

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**

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

- **Trapdoor channel**

- **Zero-error communication scheme**

\[
\begin{array}{ccc}
\tilde{x}^{n+1} & \text{flag} & \text{index} \\
0010100010100101 & \text{xxx} & 0100101000
\end{array}
\]
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)

$$Y \sim Bern(p_2)$$

$$H(Y) = \mu_0 H(p_0) + \mu_1 H(p_1) + \mu_2 H(p_2),$$

where $\mu$ 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

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

Observation 1

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

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

**Observation 1**

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

**Theorem 1**

$$H(X|Y) = 0 \iff H(X_1^k|Y_1^k, X_0, X_{k+1}) = 0 \text{ for all } k \in \mathcal{N}.$$
Proof (part a) of Theorem 1

"\[\Rightarrow\]"

\[
H(\mathcal{X}|\mathcal{Y}) = \lim_{n \to \infty} \frac{1}{n} H(X^n|Y^n)
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \left[ H(X_1^n|Y_1^n, X_0, X_{n+1}) + I(X_1^n; X_0, X_{n+1}|Y_1^n) \right]
\]

\[
= \lim_{n \to \infty} \frac{1}{n} I(X_1^n; X_0, X_{n+1}|Y_1^n)
\]

\[
\leq \lim_{n \to \infty} \frac{1}{n} \left[ H(X_0) + H(X_{n+1}) \right]
\]

\[
= 0.
\]
Proof (part b) of Theorem 1

“⇒”

\[ H(X^n | Y^n) \]

Figure: Hidden Markov Process
Proof (part b) of Theorem 1

“⇒”

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

“⇒”

Stationarity allows time shifts.

Figure: Hidden Markov Process
Proof (part b) of Theorem 1

“⇒”

\[
H(X_1^k | Y_1^k, X_0, X_{k+1}) = \frac{1}{n} \sum_{i=1}^{n} H(X_{(i-1)k'+1}^{ik'-1} | Y_{(i-1)k'+1}^{ik'-1}, X_{(i-1)k'}, X_{ik'}) \\
= \frac{1}{n} H(X_1^{nk'} | Y_1^{nk'}, \{X_{ik'}\}_i^{n=0}) \\
= \lim_{n \to \infty} \frac{1}{n} H(X_1^{nk'} | Y_1^{nk'}, \{X_{ik'}\}_i^{n=0}) \\
\leq k' \lim_{n \to \infty} \frac{1}{nk'} H(X_1^{nk'} | Y_1^{nk'}) \\
= k' H(X | Y) \\
= 0,
\]

where $k' = k + 1$. 
Other Questions

- How many states are necessary to construct a non-trivial HMP that satisfies \( H(X^k_1|Y^k_1, X_0, X_{k+1}) \) for all \( k \)?

- Does \( H(X^k_1|Y^k_1, X_0, X_{k+1}) \) 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