Asymptotics for Entropy of Hidden Markov Processes and Capacity of Input-restricted Noisy Channels

Guangyue Han
University of Hong Kong

Brian Marcus
University of British Columbia

Entropy of Stationary Processes

For a stationary process $Y = \{Y_i\}$, defined by joint probability distribution p, the **entropy** of Y is:

$$H(Y) = \lim_{n \to \infty} H_n(Y),$$

where

$$H_n(Y) = H(Y_0|Y_{-1}, Y_{-2}, \cdots, Y_{-n}) = \sum_{y^0_{-n}} -p(y^0_{-n}) \log p(y_0|y^{-1}_{-n}).$$

Known as “entropy rate” in info. theory and “measure-theoretic entropy” in ergodic theory.

Entropy of Markov chains

Let Y be a stationary (finite-state, first-order) Markov chain defined by probability transition matrix Δ:

$$\Delta(i, j) = p(y_1 = j | y_0 = i).$$

and stationary vector $p(y_0 = i)$. Then,

$$H(Y) = -\sum_{i,j} p(y_0 = i)\Delta(i, j) \log \Delta(i, j).$$

Higher-order Markov chains can be recoded to first-order Markov chains.

Hidden Markov Processes (HMP)

Defn: A **Hidden Markov Process** is a stationary process which is a continuous factor of a Markov chain.

Equivalent Defn: A **Hidden Markov Process** $Z = \{Z_i\}$ is a stationary process of the form $Z_i = \Phi(Y_i)$, where $Y = \{Y_i\}$ is a Markov chain and Φ is a function on the Markov states.

Proof: Recode to 1-block factor map by enlarging state space.

Problem: compute entropy of HMP's.

Motivation:

- HMP's are tractable models of many phenomena.
- - Entropy measures compressibility^{-1} of Y.
  - Computation of entropy is a first step to compute the Shannon capacity of an input-restricted noisy channel.

Note: HMP's are typically *not* Markov (of any order).

Example 1:

Let Y be Markov chain on $\{1, 2, 3\}$ with probability transition matrix:

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

Let $Z = \Phi(Y)$ be the HMP, where $\Phi(1) = a$ and $\Phi(2) = \Phi(3) = b$,
e.g., 122213312121331222 1 is mapped to:

abbbabbababab babbba.

One computes, using the stationary vector $[2/7, 4/7, 1/7]$:

$$p(z_0 = b | z_{-1} = b, z_{-2} = b, \ldots z_{-k} = b, z_{-k-1} = b) =$$

a weighted average of 2/3 and 1/3 with relative weights:

$(4/7)(2/3)^k$ and $(1/7)(1/3)^k$. And

$$p(z_0 = b | z_{-1} = b, z_{-2} = b, \ldots z_{-k} = b, z_{-k-1} = a) =$$

a weighted average of 2/3 and 1/3 with relative weights:

$(2/7)(2/3)^k$ and $(2/7)(1/3)^k$.

- This HMP cannot be realized as an equal entropy factor of any Markov chain (Marcus-Petersen-Williams (1984)).
- This is in contrast to computation of topological entropy of sofic shift as an entropy-preserving factor of an SFT

Example 2:

$Z(\varepsilon)$ is the output process obtained when passing a stationary binary Markov chain X through a binary symmetric channel (BSC(ε)):

$$Z_n(\varepsilon) = X_n \oplus E_n(\varepsilon),$$

where $\oplus$ denotes binary addition, X_n denotes the binary input, $E_n(\varepsilon)$ denotes the i.i.d. binary noise with $p_E(0) = 1 - \varepsilon$ and $p_E(1) = \varepsilon$, and $Z_n(\varepsilon)$ denotes the corrupted output.

Let

$$\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}$$

be the probability transition matrix for X.

Then $Y(\varepsilon) = \{(X_n, E_n(\varepsilon))\}$ is jointly Markov with

$$\Delta(\varepsilon) = \left[\begin{array}{c|cccc} y & (0,0) & (0,1) & (1,0) & (1,1) \\ \hline (0,0) & \pi_{00}(1-\varepsilon) & \pi_{00}\varepsilon & \pi_{01}(1-\varepsilon) & \pi_{01}\varepsilon \\ (0,1) & \pi_{00}(1-\varepsilon) & \pi_{00}\varepsilon & \pi_{01}(1-\varepsilon) & \pi_{01}\varepsilon \\ (1,0) & \pi_{10}(1-\varepsilon) & \pi_{10}\varepsilon & \pi_{11}(1-\varepsilon) & \pi_{11}\varepsilon \\ (1,1) & \pi_{10}(1-\varepsilon) & \pi_{10}\varepsilon & \pi_{11}(1-\varepsilon) & \pi_{11}\varepsilon \end{array} \right].$$

$Z(\varepsilon) = \Phi(Y(\varepsilon))$ is an HMP, where

$\Phi((0,0)) = \Phi((1,1)) = 0$ and

$\Phi((0,1)) = \Phi((1,0)) = 1$

History:

- H(Z) expressed as an integral of a simple function with respect to a complicated measure on a simplex (Blackwell (1957))
- general upper and lower bounds on H(Z) (Birch, (1962))
- Karl Petersen gets interested; MPW paper and plots of estimated entropies and capacities (1981)
- For Markov chain input over BSC(ε), exact computation of leading terms of asymptotics of entropy as noise parameters or Markov transition probabilities tend to extremes.
  - Jacquet-Seroussi-Szpankowski (2004, 2006)
  - Ordentlich-Weissmann (2004, 2005)
  - Zuk et. al (2004, 2006)
  - Peres-Quas (2007)

Decomposition of Δ

$\mathcal{M}$: set of Markov states

$\mathcal{A}$: set of hidden Markov symbols.

So, $\Phi : \mathcal{M} \rightarrow \mathcal{A}$.

For $a \in \mathcal{A}$, let Δ_a denote the matrix defined by:

$$\begin{aligned} \Delta_a(i,j) &= \Delta(i,j) \text{ for } j \text{ with } \Phi(j) = a \\ \Delta_a(i,j) &= 0 \quad \text{otherwise .} \end{aligned}$$

For Example 1:

$$\Delta_a = \begin{bmatrix} 0 & 0 & 0 \\ 1/3 & 0 & 0 \\ 2/3 & 0 & 0 \end{bmatrix} \Delta_b = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 0 & 2/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

Dynamics

Let W denote unit simplex in $R^{|\mathcal{M}|}$. For each symbol $a \in \mathcal{A}$, define $f_a : W \to W$ by:

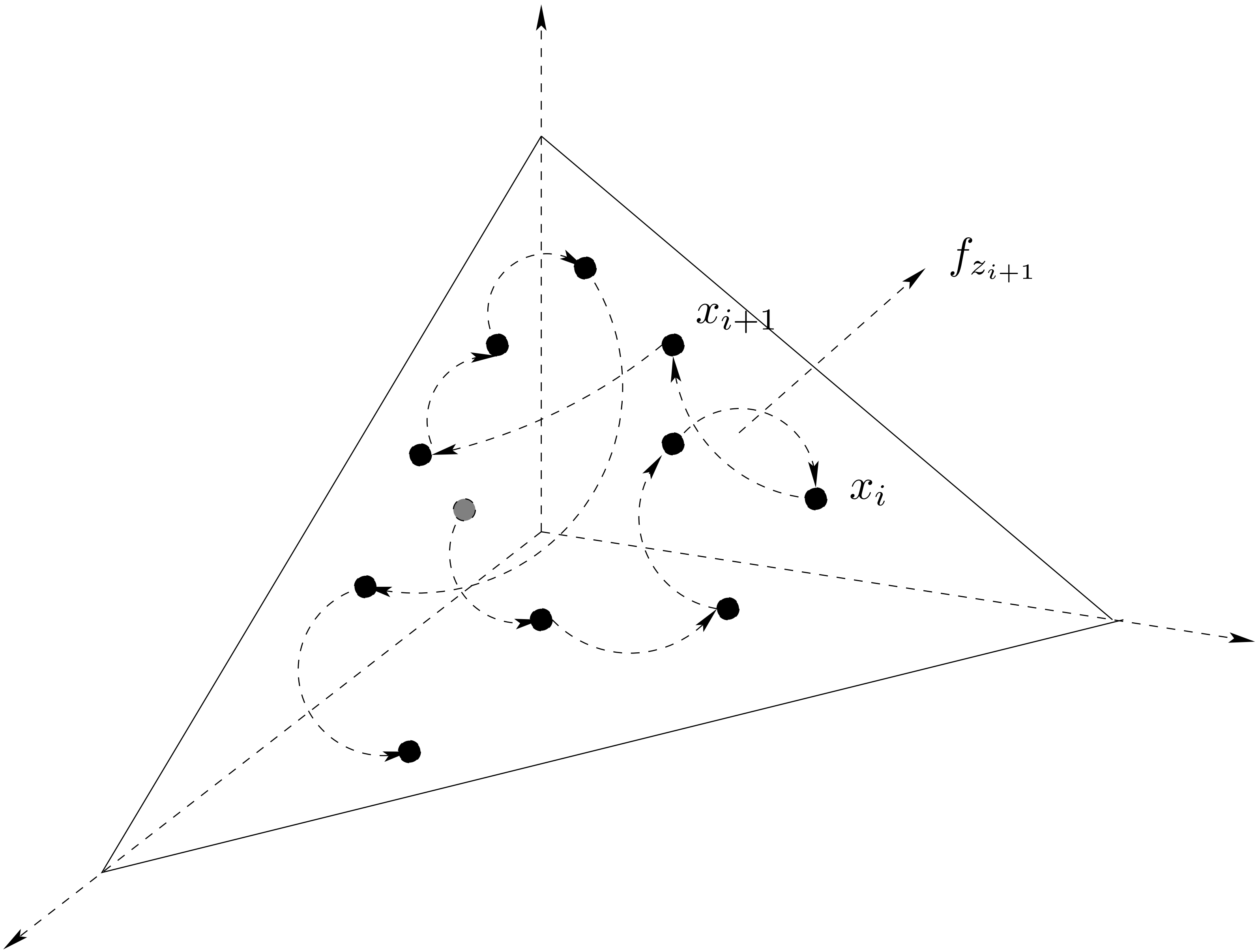
$$f_a(w) = \frac{w\Delta_a}{w\Delta_a \mathbf{1}}$$

Define

$$x_i = x_i(z_{-\infty}^i) = p(y_i = \cdot\, |z_{-\infty}^i). \tag{1}$$

Then, $\{x_i\}$ satisfies the iteration

$$x_i = f_{z_i}(x_{i-1}), \tag{2}$$



$f_{z_{i+1}}$
x_{i+1}
x_i

Defn: (**Black Hole:**) For every $a \in \mathcal{A}$,

- Δ_a has rank one – and –
- every column of Δ_a is either strictly positive or all zero.

Note: For a Black Hole, each f_a maps entire simplex to single point.

Theorem 1. *Let $\Delta(\varepsilon)$ be family of stochastic matrices and Φ a function on states. Let $Z(\varepsilon)$ denote resulting family of HMP's.*
IF

- $(\Delta(0), \Phi)$ *is a Black Hole – and –*
- $\Delta(\varepsilon)$ *is analytically parameterized around $\varepsilon = 0$,*

THEN

- $H(Z(\varepsilon))$ *is analytic around $\varepsilon = 0$.*
- $H(Z(\varepsilon))^{(N)}|_{\varepsilon=0} = H_{\lceil (N+1)/2 \rceil}(Z(\varepsilon))^{(N)}|_{\varepsilon=0}$

Example 1 Revisited

$$\Delta = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 2/3 & 0 \\ 2/3 & 0 & 1/3 \end{bmatrix},$$

with $\Phi(1) = a$ and $\Phi(2) = \Phi(3) = b$,

$$\Delta_a = \begin{bmatrix} 0 & 0 & 0 \\ 1/3 & 0 & 0 \\ 2/3 & 0 & 0 \end{bmatrix} \Delta_b = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 0 & 2/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

(Δ, Φ) is *Not* a Black Hole.

Example 2 Revisited

$$
= \left[\begin{array}{c|cccc}
y & (0,0) & (0,1) & (1,0) & (1,1) \\
\hline
(0,0) & \pi_{00}(1-\varepsilon) & \pi_{00}\varepsilon & \pi_{01}(1-\varepsilon) & \pi_{01}\varepsilon \\
(0,1) & \pi_{00}(1-\varepsilon) & \pi_{00}\varepsilon & \pi_{01}(1-\varepsilon) & \pi_{01}\varepsilon \\
(1,0) & \pi_{10}(1-\varepsilon) & \pi_{10}\varepsilon & \pi_{11}(1-\varepsilon) & \pi_{11}\varepsilon \\
(1,1) & \pi_{10}(1-\varepsilon) & \pi_{10}\varepsilon & \pi_{11}(1-\varepsilon) & \pi_{11}\varepsilon
\end{array}\right]
\Delta(0) = \left[\begin{array}{c|cccc}
y & (0,0) & (0,1) & (1,0) & (1,1) \\
\hline
(0,0) & \pi_{00} & 0 & \pi_{01} & 0 \\
(0,1) & \pi_{00} & 0 & \pi_{01} & 0 \\
(1,0) & \pi_{10} & 0 & \pi_{11} & 0 \\
(1,1) & \pi_{10} & 0 & \pi_{11} & 0
\end{array}\right]
$$

$$\Phi((0,0)) = \Phi((1,1)) = 0; \quad \Phi((0,1)) = \Phi((1,0)) = 1$$

$$
(0) = \left[\begin{array}{c|cccc}
y & (0,0) & (0,1) & (1,0) & (1,1) \\
\hline
(0,0) & \pi_{00} & 0 & 0 & 0 \\
(0,1) & \pi_{00} & 0 & 0 & 0 \\
(1,0) & \pi_{10} & 0 & 0 & 0 \\
(1,1) & \pi_{10} & 0 & 0 & 0
\end{array}\right]
\Delta_1(0) = \left[\begin{array}{c|cccc}
y & (0,0) & (0,1) & (1,0) & (1,1) \\
\hline
(0,0) & 0 & 0 & \pi_{01} & 0 \\
(0,1) & 0 & 0 & \pi_{01} & 0 \\
(1,0) & 0 & 0 & \pi_{11} & 0 \\
(1,1) & 0 & 0 & \pi_{11} & 0
\end{array}\right]
$$

If $\Pi > 0$, then $(\Delta(0), \Phi)$ *Is* a Black Hole.

Rough Idea of Proof: Stabilization of Derivatives

Recall (1):

$$x_i = x_i(z_{-\infty}^i) = p(y_i = \cdot \; |z_{-\infty}^i).$$

and the iteration (2):

$$x_i = \frac{x_{i-1}\Delta_{z_i}}{x_{i-1}\Delta_{z_i}\mathbf{1}} \equiv g(x_{i-1}, \Delta_{z_i})$$

Since at $\varepsilon = 0$, Δ_{z_i} has rank one, g must be constant as a function of x_{i-1}. So, at $\varepsilon = 0$,

$$p(y_i = \cdot|z_{-\infty}^i) = x_i = \frac{x_{i-1}\Delta_{z_i}}{x_{i-1}\Delta_{z_i}\mathbf{1}} = \frac{p(y_{i-1} = \cdot)\Delta_{z_i}}{p(y_{i-1} = \cdot)\Delta_{z_i}\mathbf{1}} = p(y_i = \cdot|z_i). \quad (3)$$

Taking the derivative of g with respect to ε, we have:

$$x_i' = \frac{\partial g}{\partial x_{i-1}}(x_{i-1}, \Delta_{z_i})\; x_{i-1}' + \frac{\partial g}{\partial \Delta_{z_i}}(x_{i-1}, \Delta_{z_i})\; \Delta_{z_i}'.$$

Since at $\varepsilon = 0$, g is constant as a function of x_{i-1}, we have at $\varepsilon = 0$:

$$\frac{\partial g}{\partial x_{i-1}}(x_{i-1}, \Delta_{z_i}) = \frac{\partial(\text{constant vector})}{\partial x_{i-1}} = 0.$$

It then follows from (3) that

$$p'(y_i = \cdot | z_{-\infty}^i)|_{\varepsilon=0} = x_i'|_{\varepsilon=0} = p'(y_i = \cdot | z_{i-1}^i)|_{\varepsilon=0}.$$

Generalization

- **Weak Black Hole:** For every a, Δ_a has rank 0 or 1.
- $\Delta(\varepsilon)$ is **normally parameterized** by ε ($\varepsilon \geq 0$) if

  1. each entry of $\Delta(\varepsilon)$ is an analytic function around $\varepsilon = 0$,
  2. for sufficiently small $\varepsilon > 0$, $\Delta(\varepsilon)$ is irreducible.

Theorem 2. *Let* $\Delta(\varepsilon)$ *be family of stochastic matrices and* Φ *a function on states. Let* $Z(\varepsilon)$ *denote resulting family of HMP's.*
IF

- $(\Delta(0), \Phi)$ *is a* Weak Black Hole – *and* –
- $\Delta(\varepsilon)$ *is* normally parameterized

THEN there are sequences f_j, g_j *such that*

- *for* $k \geq 0$*,*

$$H(Z(\varepsilon)) = H(Z(0)) + \sum_{j=1}^{k} f_j \varepsilon^j + \sum_{j=1}^{k+1} g_j \varepsilon^j \log \varepsilon + O(\varepsilon^{k+1}) \quad (4)$$

- f_j *and* g_j *depend only on* $H_{6j+6}(Z(\varepsilon))$

Example 2 Re-Revisited

$$= \left[\begin{array}{c|cccc} y & (0,0) & (0,1) & (1,0) & (1,1) \\ \hline (0,0) & \pi_{00}(1-\varepsilon) & \pi_{00}\varepsilon & \pi_{01}(1-\varepsilon) & \pi_{01}\varepsilon \\ (0,1) & \pi_{00}(1-\varepsilon) & \pi_{00}\varepsilon & \pi_{01}(1-\varepsilon) & \pi_{01}\varepsilon \\ (1,0) & \pi_{10}(1-\varepsilon) & \pi_{10}\varepsilon & \pi_{11}(1-\varepsilon) & \pi_{11}\varepsilon \\ (1,1) & \pi_{10}(1-\varepsilon) & \pi_{10}\varepsilon & \pi_{11}(1-\varepsilon) & \pi_{11}\varepsilon \end{array}\right] \quad \Delta(0) = \left[\begin{array}{c|cccc} y & (0,0) & (0,1) & (1,0) & (1,1) \\ \hline (0,0) & \pi_{00} & 0 & \pi_{01} & 0 \\ (0,1) & \pi_{00} & 0 & \pi_{01} & 0 \\ (1,0) & \pi_{10} & 0 & \pi_{11} & 0 \\ (1,1) & \pi_{10} & 0 & \pi_{11} & 0 \end{array}\right]$$

$$(0) = \left[\begin{array}{c|cccc} y & (0,0) & (0,1) & (1,0) & (1,1) \\ \hline (0,0) & \pi_{00} & 0 & 0 & 0 \\ (0,1) & \pi_{00} & 0 & 0 & 0 \\ (1,0) & \pi_{10} & 0 & 0 & 0 \\ (1,1) & \pi_{10} & 0 & 0 & 0 \end{array}\right] \quad \Delta_1(0) = \left[\begin{array}{c|cccc} y & (0,0) & (0,1) & (1,0) & (1,1) \\ \hline (0,0) & 0 & 0 & \pi_{01} & 0 \\ (0,1) & 0 & 0 & \pi_{01} & 0 \\ (1,0) & 0 & 0 & \pi_{11} & 0 \\ (1,1) & 0 & 0 & \pi_{11} & 0 \end{array}\right]$$

- $(\Delta(0), \Phi)$ *Is* a Weak Black Hole (without any assumptions on transition matrix Π)
- If Π is irreducible, then $\Delta(\varepsilon)$ is a normal parameterization.

Example 3: Binary Erasure Channel

Inputs are either transmitted faithfully or erased with probability ε.

$$\Delta(\varepsilon) = \left[\begin{array}{c|cccc} y & (0,0) & (0,1) & (1,0) & (1,1) \\ \hline (0,0) & \pi_{00}(1-\varepsilon) & \pi_{00}\varepsilon & \pi_{01}(1-\varepsilon) & \pi_{01}\varepsilon \\ (0,1) & \pi_{00}(1-\varepsilon) & \pi_{00}\varepsilon & \pi_{01}(1-\varepsilon) & \pi_{01}\varepsilon \\ (1,0) & \pi_{10}(1-\varepsilon) & \pi_{10}\varepsilon & \pi_{11}(1-\varepsilon) & \pi_{11}\varepsilon \\ (1,1) & \pi_{10}(1-\varepsilon) & \pi_{10}\varepsilon & \pi_{11}(1-\varepsilon) & \pi_{11}\varepsilon \end{array}\right].$$

$Z(\varepsilon) = \Phi(Y(\varepsilon))$ is an HMP where:

$\Phi(0,0) = 0$,

$\Phi(1,0) = 1$,

$\Phi(0,1) = \Phi(1,1) = ?$

Thus,

$$\Delta_0 = \begin{bmatrix} \pi_{00}(1-\varepsilon) & 0 & 0 & 0 \\ \pi_{00}(1-\varepsilon) & 0 & 0 & 0 \\ \pi_{10}(1-\varepsilon) & 0 & 0 & 0 \\ \pi_{10}(1-\varepsilon) & 0 & 0 & 0 \end{bmatrix}, \Delta_1 = \begin{bmatrix} 0 & 0 & \pi_{01}(1-\varepsilon) & 0 \\ 0 & 0 & \pi_{01}(1-\varepsilon) & 0 \\ 0 & 0 & \pi_{11}(1-\varepsilon) & 0 \\ 0 & 0 & \pi_{11}(1-\varepsilon) & 0 \end{bmatrix},$$

$$\Delta_? = \begin{bmatrix} 0 & \pi_{00}\varepsilon & 0 & \pi_{01}\varepsilon \\ 0 & \pi_{00}\varepsilon & 0 & \pi_{01}\varepsilon \\ 0 & \pi_{10}\varepsilon & 0 & \pi_{11}\varepsilon \\ 0 & \pi_{10}\varepsilon & 0 & \pi_{11}\varepsilon \end{bmatrix}.$$

- $(\Delta(0), \Phi)$ *Is* a Weak Black Hole (without any assumptions on transition matrix Π).
- If Π is irreducible, then $\Delta(\varepsilon)$ is a normal parameterization.

Idea of Proof of Theorem 2

(1) Given k, let $N = N(k) = 6k + 6$. Then for any hidden Markov sequence z^0_{-N},

EITHER: $p(z^0_{-N}) = O(\varepsilon^{k+2})$

OR: for any Markov state y "that matters," and any $0 \leq i \leq 2k+1$,

$$p^{(i)}(z_0|z^{-1}_{-N})|_{\varepsilon=0} = p^{(i)}(z_0|z^{-1}_{-N}y)|_{\varepsilon=0}.$$

(2) It follows that

$$H(Z_0(\varepsilon)|Z(\varepsilon)^{-1}_{-N}) = H(Z_0(\varepsilon)|Z(\varepsilon)^{-1}_{-N}, Y_{-N-1}) + O(\varepsilon^{k+1}).$$

(3) Apply Birch bounds: Let Z be an HMP defined by Markov chain Y and function Φ. Then for all $n \geq 0$,

$$H(Z_0|Z^{-1}_{-n}, Y_{-n-1}) \leq H(Z) \leq H(Z_0|Z^{-1}_{-n}). \quad (5)$$

(In fact, upper and lower bounds agree and stabilize for $n \geq N(k)$.)

Proof of Birch bounds

- Upper bound: monotonicity.
- Lower bound: fix n;

$$H(Z) = \lim_{m\to\infty} H(Z_0|Z_{-m}^{-1}) \geq \lim_{m\to\infty} H(Z_0|Z_{-n}^{-1}, Y_{-m}^{-n-1}) =$$

$$\lim_{m\to\infty} H(Z_0|Z_{-n}^{-1}, Y_{-n-1}) = H(Z_0|Z_{-n}^{-1}, Y_{-n-1}).$$

More on proof of Theorem 2

Let $V \geq 0$ be a vector indexed by Markov states $\mathcal{M}$.

Define:

- $$p_V(z_{-n}^{-1}) = V\Delta_{z_{-n}} \cdots \Delta_{z_{-1}}\mathbf{1}$$

- $$p_V(z_0|z_{-n}^{-1}) = \frac{p_V(z_{-n}^{0})}{p_V(z_{-n}^{-1})}$$

Examples:

- If V is the stationary vector, then $p_V(z_{-n}^{0}) = p(z_{-n}^{0})$ and $p_V(z_0|z_{-n}^{-1}) = p(z_0|z_{-n}^{-1})$.
- If $V = p(y)\chi_y$, then $p_V(z_{-n}^{0}) = p(yz_{-n}^{0})$ and $p_V(z_0|z_{-n}^{-1}) = p(z_0|z_{-n}^{-1}y)$.

Let $V = V(\varepsilon) \geq 0$ analytically parameterized by ε and not identically 0.

Write:

$$p_V(z_{-n}^{-1}) = \sum_{j=0}^{\infty} b(V)_j \varepsilon^j$$

$$p_V(z_0|z_{-n}^{-1}) = \sum_{j=0}^{\infty} a(V)_j \varepsilon^j$$

Define: ord$(p_V(z_{-n}^{-1}))$ as smallest j such that $b(V)_j \neq 0$

Lemma: If ord$(p_V(z_{-n}^{-1})),$ ord$(p_{\hat{V}}(z_{-n}^{-1})) \leq k$, then

$$a(V)_j = a(\hat{V})_j, \quad \text{for all } 0 \leq j \leq n - 4k - 1$$

Proof: by induction; use rank ≤ 1 condition to build up more and more j such that this equality holds.

Say $k = 0, n = 1$.

$$\sum_{j=0}^{\infty} a(V)_j \varepsilon^j = p_V(z_0|z_{-n}^{-1}) = \frac{p_V(z_{-1}z_0)}{p_V(z_{-1})} = \frac{V\Delta_{z_{-1}}\Delta_{z_0}\mathbf{1}}{V\Delta_{z_{-1}}\mathbf{1}} = \frac{V\Delta_{z_{-1}}}{V\Delta_{z_{-1}}\mathbf{1}}\Delta_{z_0}\mathbf{1}$$

By assumption that $k = 0$, we have $V(0)\Delta_{z_{-1}}(0)\mathbf{1} \neq \mathbf{0}$. Thus,

$$a(V)_0 = \frac{V(0)\Delta_{z_{-1}}(0)}{V(0)\Delta_{z_{-1}}(0)\mathbf{1}}\Delta_{z_0}(0)\mathbf{1}$$

Similarly,

$$a(\hat{V})_0 = \frac{\hat{V}(0)\Delta_{z_{-1}}(0)}{\hat{V}(0)\Delta_{z_{-1}}(0)\mathbf{1}}\Delta_{z_0}(0)\mathbf{1}$$

But since $\Delta_{z_{-1}}(0)$ has rank 1, these two expressions are equal.

Explicit formula in special case

Let X be a Markov chain of order m (recode to first order Markov chain in order to fit our framework).

Let $\mathcal{A}(X)$ denote set of words of positive probability for X.

Let $Z(\varepsilon)$ be the output of X passed through BSC (ε). Then for $H(Z(\varepsilon))$, g_1 depends only on X and

$$g_1 = g_1(X) = -\sum_{w\in\mathcal{A}(X), wv\notin\mathcal{A}(X), |w|=2m, |v|=1} d(wv)$$

where

$$d(u_{-n}^{-1}) = \sum_{j=1}^{n} p_X(u_{-n}^{-j-1}\bar{u}_{-j}u_{-j+1}^{-1})$$

Asymptotics of Input-Restricted Noisy Channel Capacity

Consider a binary irreducible Shift of Finite Type S. For BSC(ε) with input sequences restricted to S, capacity is defined:

$$C(S,\varepsilon) = \sup_{\text{stationary } X \text{ supported on } S} H(Z(\varepsilon)) - H(\varepsilon),$$

where $Z(\varepsilon)$ is the output process corresponding to X and $H(\varepsilon)$ is the binary entropy function.

Theorem 3. *(JSS, HM (2006))*

$$C(S,\varepsilon) = H(S) + (g_1(X_{max}) + 1)\varepsilon \log(\varepsilon) + O(\varepsilon),$$

where $H(S)$ is the topological entropy of $\mathcal{S}$, X_{max} is the maximum entropy process associated with S, and g_1 is as in Theorem 2.

Current work: higher order asymptotics and other channels.

Example: Let X be a first order input Markov chain supported on $S =$ Golden Mean Shift (i.e., 11 is forbidden), transmitted over BSC(ε) with corresponding output HMP $Z(\varepsilon)$. Theorem 2 yields:

$$H(Z(\varepsilon)) = H(X) + \left(\frac{\pi_{01}(\pi_{01} - 2)}{1 + \pi_{01}} \right) \varepsilon \log(\varepsilon) + O(\varepsilon).$$

(originally due to Ordentlich-Weissman (2005))

The maximum entropy Markov chain is defined by the transition matrix:

$$\begin{bmatrix} 1/\lambda & 1/\lambda^2 \\ 1 & 0 \end{bmatrix}$$

and

$$H(\mathcal{S}) = H(X_{max}) = \log \lambda,$$

where λ is the golden mean. Thus, in this case $\pi_{01} = 1/\lambda^2$, and from Theorem 3, we obtain:

$$C(S, \varepsilon) = \log \lambda + \left(\frac{2\lambda + 2}{4\lambda + 3} \right) \varepsilon \log(\varepsilon) + O(\varepsilon).$$

Dear Karl,

1. Thanks for all the wonderful mathematics: past, present and future.

2. HAPPY BIRTHDAY

3. I will save the embarrassing stories for Friday night.

Best wishes - Brian