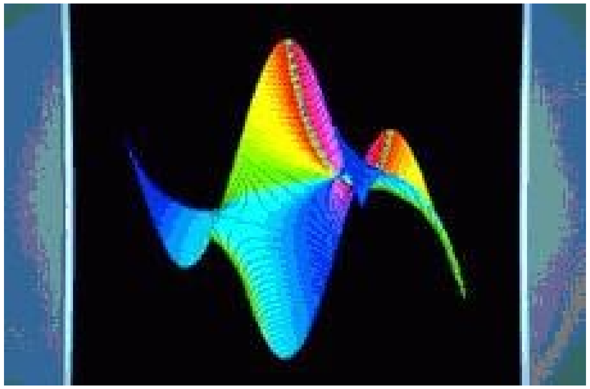
On the Entropy of a Hidden Markov Process

Gadiel Seroussi* Philippe Jacquet† W. Szpankowski‡

March 17, 2004



*HPL, Palo Alto, USA.

†INRIA, Rocquencourt, France

‡Department of Computer Science, Purdue University, USA.

Outline of the Talk

1. Hidden Markov Model and Its Applications
2. Product of Random Matrices
3. Entropy Rate as a Lyapunov Exponent
4. Asymptotic Expansion
5. Experimental Verifications
6. Sketch of the Proof
7. Rényi's Entropy

Problem Formulation

1. Let $X = \{X_k\}_{k\geq 1}$ be a first order stationary Markov process over a binary alphabet,
with transition matrix $\mathbf{P} = \{\pi_{ab}\}_{a,b\in\{01,\}}$;

$$\pi_{ab} = P_X(X_k{=}b|X_{k-1}{=}a)$$

.

2. Let $E = \{E_k\}_{k\geq 1}$ be a Bernoulli (binary i.i.d.) noise process independent of X, such that

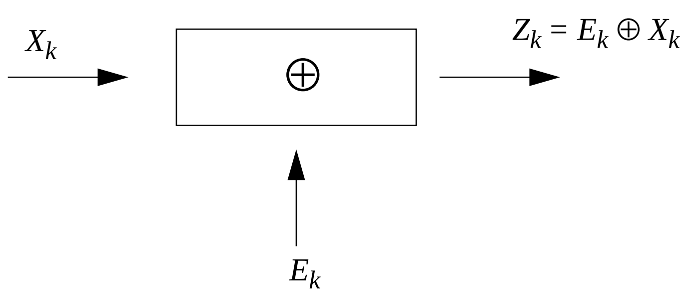
$$P(E_i = 1) = \varepsilon$$

.

3. Define $Z = \{Z_k\}_{k\geq 1}$ such that

$$\mathbf{Z_k} = X_k \oplus E_k, \quad k \geq 1,$$

where $\oplus$ denotes addition modulo 2 (exclusive-or).



Hidden Markov Process

The process **Z** is, in a sense, one of the simplest examples of a hidden Markov process (HMP).

Basic question: what is the entropy rate of such a process?

In general, a HMP is a process resulting from observing any discrete-time, finite state homogeneous Markov chain through a discrete-time memoryless channel.

In particular, $Z = f(X, E)$ for a Markov process X, i.i.d. E, and a function f.

Applications:
data compression
automatic character recognition
speech recognition
statistics
communications and information theory
DNA sequencing
denoising
performance of digital trees.

Application: Tries

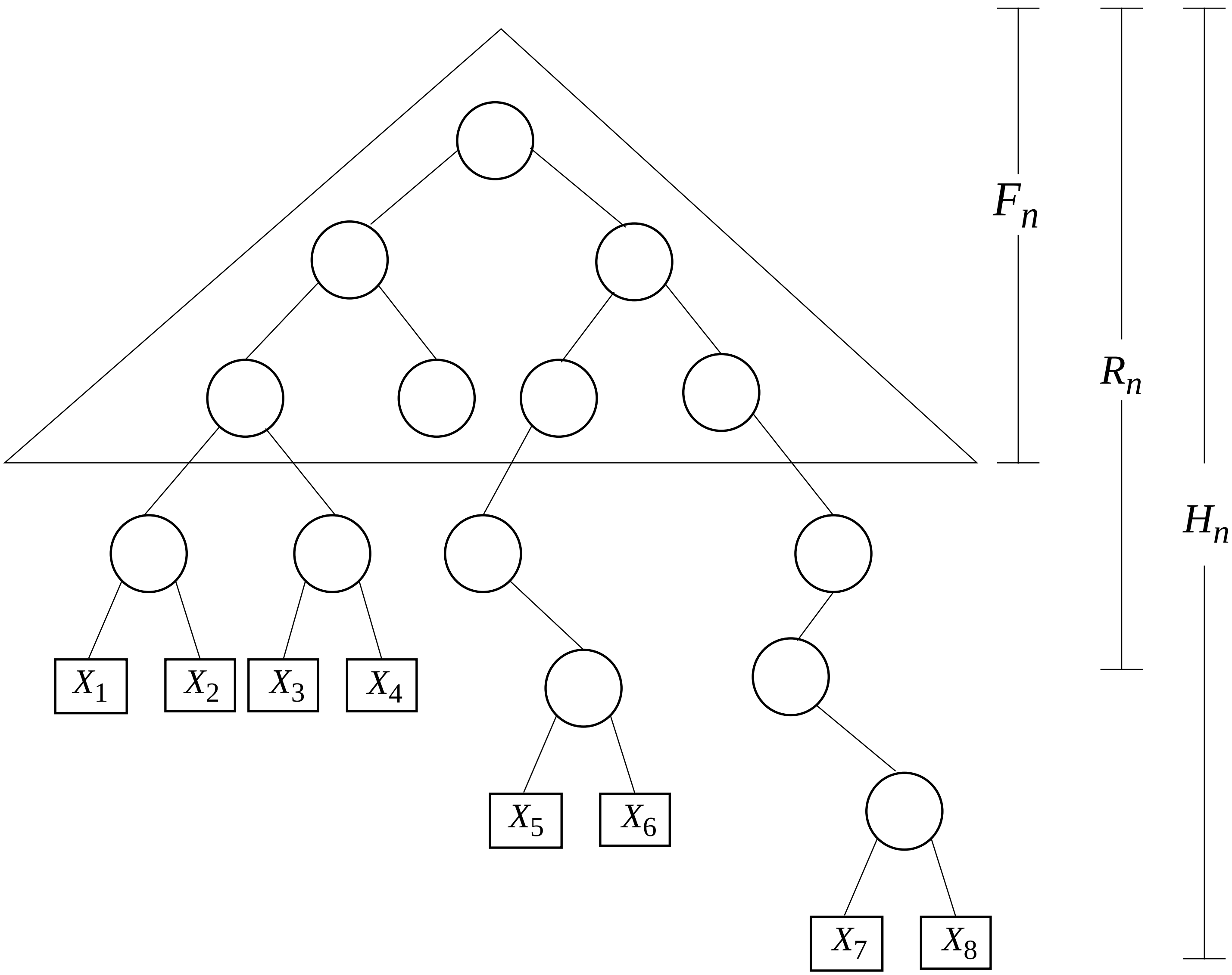


Figure 1: A trie and its parameters.

Hidden Markov Source: Binary sequences generated by a Markov source with an i.i.d. error sequence.

Some Previous Works

Blackwell derived in 1957 an expression for the entropy of HMP in terms of a measure Q, which solves an integral equation. The measure is hard to extract from the equation in any explicit way.

Ordentlich and Weissman in 2003 obtained explicit formulas for the entropy rate when $\pi_{ab} \longrightarrow 0$.

In contrast, our study focuses on the regime where the channel parameter (noise) $\varepsilon \longrightarrow 0$ is small.

Joint Distribution of $P(Z_1^n)$

For any sequence $\{Y_k\}_{k\geq 1}$, let

$$Y_i^j = Y_i Y_{i+1} \ldots Y_j.$$

Also $\bar{Y} = 1 \oplus Y$.
In particular, $Z_i = X_i$ if $E_i = 0$ and $Z_i = \bar{X}_i$ if $E_i = 1$.

We have

$$P(Z_1^n, E_n) = P(Z_1^n, E_{n-1} = 0, E_n) + P(Z_1^n, E_{n-1} = 1, E_n) =$$

$$\begin{aligned}
&= P(Z_1^{n-1}, Z_n, E_{n-1} = 0, E_n) + P(Z_1^{n-1}, Z_n, E_{n-1} = 1, E_n) \\
&= P(Z_n, E_n | Z_1^{n-1}, E_{n-1} = 0) P(Z_1^{n-1}, E_{n-1} = 0) + \\
&\quad P(Z_n, E_n | Z_1^{n-1}, E_{n-1} = 1) P(Z_1^{n-1}, E_{n-1} = 1) \\
&= P(E_n) P_X(Z_n \oplus E_n | Z_{n-1}) P(Z_1^{n-1}, E_{n-1} = 0) \\
&+ P(E_n) P_X(Z_n \oplus E_n | \bar{Z}_{n-1}) P(Z_1^{n-1}, E_{n-1} = 1)
\end{aligned}$$

Entropy as a Product of Random Matrices

Let

$$\mathbf{p}_n = [P(Z_1^n, E_n = 0), P(Z_1^n, E_n = 1)]$$

and

$$\mathbf{M}(Z_{n-1}, Z_n) = \begin{bmatrix} (1-\varepsilon)P_X(Z_n|Z_{n-1}) & \varepsilon P_X(\bar{Z}_n|Z_{n-1}) \\ (1-\varepsilon)P_X(Z_n|\bar{Z}_{n-1}) & \varepsilon P_X(\bar{Z}_n|\bar{Z}_{n-1}) \end{bmatrix}$$

where the expressions $P_X(Z_i|Z_{i-1})$ are the Markov transition probabilities computed on the components of the HMP Z.

From the previous slide we conclude that

$$\mathbf{p}_n = \mathbf{p}_{n-1}\mathbf{M}(Z_{n-1}, Z_n).$$

Since $P(Z_1^n) = \mathbf{p}_n \mathbf{1}^t$ ($\mathbf{1}^t = (1, \ldots, 1)$) we finally obtain

$$P(Z_1^n) = \mathbf{p}_1 \mathbf{M}(Z_1, Z_2) \cdots \mathbf{M}(Z_{n-1}, Z_n)\mathbf{1}^t,$$

that is, product of random matrices since $P_X(Z_i|Z_{i-1})$ are random variables.

Entropy Rate as a Lyapunov Exponent

Theorem 1 (Furstenberg and Kesten, 1960). *Let* $\mathrm{M}_1, \ldots, \mathrm{M}_n$ *form a stationary ergodic sequence and* $\mathbf{E}[\log^+ ||\mathrm{M}_1||] < \infty$ *Then*

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{E}[\log ||\mathbf{M}_1 \cdots \mathbf{M}_n||] = \lim_{n\to\infty} \frac{1}{n} \log ||\mathbf{M}_1 \cdots \mathbf{M}_n|| = \mu \quad \text{a.s.}$$

where μ *is called top Lyapunov exponent.*

Corollary 1. *Consider the HMP* Z *as defined above. The entropy rate* [1]

$$\begin{aligned} h(Z) &= \lim_{n\to\infty} \mathbf{E}[\,-\frac{1}{n}\log P(Z_1^n)\,] \\ &= \lim_{n\to\infty} \frac{1}{n}\mathbf{E}[-\log\left(\mathbf{p}_1\mathbf{M}(Z_1, Z_2)\cdots\mathbf{M}(Z_{n-1}, Z_n)\mathbf{1}^t\right)] \end{aligned}$$

is a top Lyapunov exponent of $\mathbf{M}(Z_1, Z_2)\cdots\mathbf{M}(Z_{n-1}, Z_n)$*.*

Unfortunately, it is notoriously difficult to compute top Lyapunov exponents as proved in Tsitsiklis and Blondel. Therefore, in next we derive an explicit asymptotic expansion of the entropy rate $h(Z)$.

[1] When no base is specified, logarithms are to base 2; $\ln x$ will denote the natural logarithm of x.

Main Result - Asymptotic Expansion

We now assume that $P(E_i = 1) = \varepsilon \to 0$ is small.

Theorem 2. *The entropy rate of the process Z is*

$$h(Z) = \lim_{n\to\infty} \frac{1}{n} H_n(Z^n) = h(X) + f_1(\pi_{01}, \pi_{10})\varepsilon + O(\varepsilon^2),$$

with

$$\begin{aligned} f_1(\pi_{01}, \pi_{10}) &= \mathbb{D}\left(P_X(z_1 z_2 z_3) || P_X(z_1 \bar{z}_2 z_3)\right) \\ &= \sum_{z_1 z_2 z_3} P_X(z_1 z_2 z_3) \log \frac{P_X(z_1 z_2 z_3)}{P_X(z_1 \bar{z}_2 z_3)}, \end{aligned}$$

where $h(X)$ is the entropy rate of the Markov process X, $\mathbb{D}$ denotes the Kullback-Liebler divergence, and the summation is over all binary triplets.

Example

Consider a Markov process with symmetric transition probabilities $\pi_{01} = \pi_{10} = \pi$, $\pi_{00} = \pi_{11} = 1-\pi$. This process has stationary probabilities $P_X(0) = P_X(1) = \frac{1}{2}$.

The probabilities $P_X(z_1^3)$ of binary triplets are readily computed as

$$
\begin{aligned}
P_X(000) &= P_X(111) = \frac{1}{2}(1-\pi)^2, \\
P_X(001) &= P_X(011) = P_X(100) = P_X(110) = \frac{1}{2}\pi(1-\pi), \\
P_X(010) &= P_X(101) = \pi^2.
\end{aligned}
$$

Thus we obtain from out Theorem

$$
f_1(\pi, \pi) = 2(1-2\pi) \log \frac{1-\pi}{\pi},
$$

and

$$
h(Z) = -\pi \log \pi - (1-\pi) \log(1-\pi) + \varepsilon 2(1-2\pi) \log \frac{1-\pi}{\pi} + O(\varepsilon^2).
$$

Experimental Verification

HMPs for various values of the parameters ε and $\pi_{01} = \pi_{10} = \pi$ were simulated, generating pseudo-random HMP sequences of lengths between $n = 10^8$ and $n = 4 \cdot 10^9$. For each generated sequence z_1^n, the probability $P_Z(z_1^n)$ assigned by the hidden Markov model of the given parameters was computed, and $-\frac{1}{n} \log P_Z(z_1^n)$ was taken as an estimate for the entropy rate.

Parameters			Calculated			Empirical
ε	π	n	$h(X)$	$f_1(\pi,\pi)$	$h(X)+$ $f_1(\pi,\pi)\varepsilon$	$-\frac{1}{n}\log P_Z(z_1^n)$
0.001	0.005	$4 \cdot 10^9$	0.045	15.121	0.061	0.056
0.001	0.010	$4 \cdot 10^9$	0.080	12.994	0.094	0.091
0.001	0.025	$1 \cdot 10^9$	0.168	10.042	0.179	0.177
0.01	0.050	$1 \cdot 10^8$	0.286	7.646	0.363	0.349
0.01	0.100	$1 \cdot 10^8$	0.469	5.072	0.520	0.514
0.01	0.300	$1 \cdot 10^8$	0.881	0.978	0.891	0.891

Table 1: First order approximation of $h(Z)$ according to our Theorem and empirical estimation.

Figure

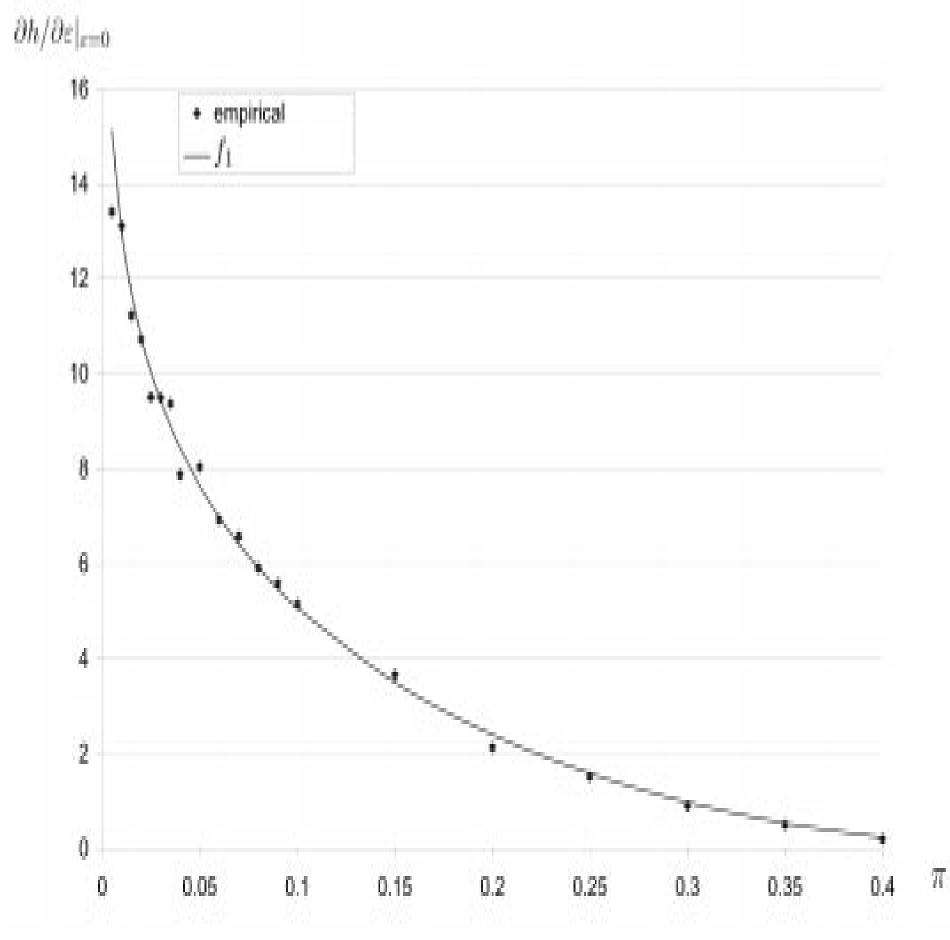


Figure 2: Values of f_1 and empirical estimation of $\partial h/\partial \varepsilon|_{\varepsilon=0}$ as a function of π.

Sketch of the Proof

1. Instead of computing entropy $H(Z_1^n)$ we evaluate the following sum

$$R(s,\varepsilon) = \sum_{z_1^n} P_Z^s(z_1^n),$$

where s is a complex variable. Observe that

$$H(Z_1^n) = \mathbf{E}\left[-\log P(Z_1^n)\right] = -(\ln 2)\frac{\partial}{\partial s} R(s,\varepsilon)\Big|_{s=1}.$$

The entropy of the underlying Markov sequence is

$$H(X_1^n) = (-\ln 2)\frac{\partial}{\partial s} R(s,0)|_{s=1}$$

and

$$R(s,0) = \sum_{z^n} P_X^s(z_1^n) = \boldsymbol{\pi}(s)\mathbf{P}^{n-1}(s)\mathbf{1}^t.$$

Proof

2. By Taylor expansion

$$R(s,\varepsilon) = R(s,0) + \varepsilon \frac{\partial}{\partial \varepsilon} R(s,\varepsilon)|_{\varepsilon=0} + O(R_{\varepsilon,\varepsilon}(s,\varepsilon')\varepsilon^2).$$

We can prove that

$$R_{\varepsilon,\varepsilon,s}(1,\varepsilon') = O(n) \quad \text{(IMPORTANT!)}$$

where $R_{\varepsilon,\varepsilon,s}(1,\varepsilon')$ is the first derivative with respect to s at $s = 1$ of $R_{\varepsilon,\varepsilon}(s,\varepsilon')$.
Thus

$$\begin{aligned} H(Z_1^n) &= H(X_1^n) - (\ln 2)\varepsilon \frac{\partial^2}{\partial s \partial \varepsilon} R(s,\varepsilon)\Big|_{\varepsilon=0,s=1} + O(n\varepsilon^2) \\ &= H(X_1^n) - (\ln 2)\varepsilon \sum_{z_1^n} \frac{\partial}{\partial s}\frac{\partial}{\partial \varepsilon} P_Z^s(z_1^n)\Big|_{\varepsilon=0,s=1} + O(n\varepsilon^2). \end{aligned}$$

Another Matrix Representation

3. We introduce a decomposition of $\mathbf{M}_i$ as follows

$$\begin{aligned}
\mathbf{M}_i &= \mathbf{M}(z_i, z_{i+1}) = \begin{bmatrix} (1-\varepsilon)P_X(z_{i+1}|z_i) & \varepsilon P_X(\bar{z}_{i+1}|z_i) \\ (1-\varepsilon)P_X(z_{i+1}|\bar{z}_i) & \varepsilon P_X(\bar{z}_{i+1}|\bar{z}_i) \end{bmatrix} \\
&= \begin{bmatrix} P_X(z_{i+1}|z_i) & 0 \\ P_X(z_{i+1}|\bar{z}_i) & 0 \end{bmatrix} + \varepsilon \begin{bmatrix} -P_X(z_{i+1}|z_i) & P(\bar{z}_{i+1}|z_i) \\ -P_X(z_{i+1}|\bar{z}_i) & P(\bar{z}_{i+1}|\bar{z}_i) \end{bmatrix} \\
&\stackrel{\text{def}}{=} \mathbf{M}_i^{(0)} + \varepsilon \mathbf{M}_i^{(1)},
\end{aligned}$$

Then

$$P_Z(z_1^n) = P(Z_1^n = z_1^n) = \mathbf{p}_0 \mathbf{M}_1 \mathbf{M}_2 \cdots \mathbf{M}_{n-1} \mathbf{1}^t =$$
$$(\mathbf{M}_0^{(0)} + \varepsilon \mathbf{M}_0^{(1)})(\mathbf{M}_1^{(0)} + \varepsilon \mathbf{M}_1^{(1)})(\mathbf{M}_2^{(0)} + \varepsilon \mathbf{M}_2^{(1)}) \cdots (\mathbf{M}_{n-1}^{(0)} + \varepsilon \mathbf{M}_{n-1}^{(1)}) \mathbf{1}^t.$$

Estimating Derivatives

4. To compute the derivative of $P_Z^s(z_1^n)$ at $\varepsilon = 0$, we first differentiate both sides of the above equation, obtaining

$$\frac{\partial}{\partial \varepsilon} P_Z(z_1^n)\Big|_{\varepsilon=0} = \sum_{i=0}^{n-1} \mathbf{M}_0^{(0)} \mathbf{M}_1^{(0)} \cdots \mathbf{M}_{i-1}^{(0)} \mathbf{M}_i^{(1)} \mathbf{M}_{i+1}^{(0)} \cdots \mathbf{M}_{n-1}^{(0)} \mathbf{1}.$$

And after some algebra we arrive at

$$\frac{\partial}{\partial \varepsilon} P_Z(z_1^n)\Big|_{\varepsilon=0} = P_X(z_1^n) \sum_{i=0}^{n-1} (g_i(z_1^n) - 1),$$

where

$$g_i(z_1^n) \quad = \quad \frac{P_X(\bar{z}_{i+1}|z_i) P_X(z_{i+2}|\bar{z}_{i+1})}{P_X(z_{i+1}|z_i) P_X(z_{i+2}|z_{i+1})} = \frac{P_X(z_i \bar{z}_{i+1} z_{i+2})}{P_X(z_i z_{i+1} z_{i+2})}$$

5. Thus

$$\frac{\partial}{\partial \varepsilon} P_Z^s(z_1^n)\bigg|_{\varepsilon=0} = \left[s P_Z^{s-1}(z_1^n) P_X(z_1^n) \sum_{i=0}^{n-1} (g_i(z_1^n) - 1) \right]_{\varepsilon=0} .$$

thus, in a matrix form, we obtain

$$\frac{\partial}{\partial \varepsilon} R(s, \varepsilon)\bigg|_{\varepsilon=0} = s\boldsymbol{\pi}(s) \sum_{i=1}^{n-1} \mathbf{P}^{i-1}(s) \left(\mathbf{Q}_1(s)\mathbf{Q}_2(s) - \mathbf{P}^2(s) \right) \mathbf{P}^{n-i-2}(s)\mathbf{1}^t$$

where

$$\mathbf{Q}_1(s) = \begin{bmatrix} \pi_{00}\pi_{01}^{s-1} & \pi_{01}\pi_{00}^{s-1} \\ \pi_{10}\pi_{11}^{s-1} & \pi_{11}\pi_{10}^{s-1} \end{bmatrix}, \quad \mathbf{Q}_2(s) = \begin{bmatrix} \pi_{00}\pi_{10}^{s-1} & \pi_{01}\pi_{11}^{s-1} \\ \pi_{10}\pi_{00}^{s-1} & \pi_{11}\pi_{01}^{s-1} \end{bmatrix}.$$

and

$$\mathbf{P}(s) = \begin{bmatrix} \pi_{00}^s & \pi_{01}^s \\ \pi_{10}^s & \pi_{11}^s \end{bmatrix},$$

Observe that

$$Q_1(1)Q_2(1) = P^2(1)$$

.

Finishing Up ...

6. To find the linear term in the Taylor expansion for entropy, we use the spectral representation of the matrix $\mathbf{P}(s)$. Let

$\lambda(s)$ – be the main eigenvalue of $\mathbf{P}(s)$

$\mathbf{r}_1^t(s)$, $\mathbf{l}_1(s)$ – the corresponding right and left main eigenvectors,

$\mu(s)$ – be the second eigenvalue,

$\mathbf{r}_2^t(s)$, $\mathbf{l}_2(s)$ – the respective right and left eigenvectors.

The matrix spectral representation yields

$$\mathbf{P}^k(s) = \lambda^k(s)\,\mathbf{r}_1^t(s)\mathbf{l}_1(s) + \mu^k(s)\,\mathbf{r}_2^t(s)\mathbf{l}_2(s).$$

Using this we finally obtain

$$\begin{aligned}\left.\frac{\partial^2}{\partial\varepsilon\partial s}R(s,\varepsilon)\right|_{\substack{\varepsilon=0,\\ s=1}} &= n\,\boldsymbol{\pi}(1)\mathbf{r}_1^t(1)\;\mathbf{l}_1(1)\mathbf{1}^t\;\mathbf{l}_1(1)\\ &\times\left.\frac{\partial}{\partial s}\Big(\mathbf{Q}_1(s)\mathbf{Q}_2(s)-\mathbf{P}^2(s)\Big)\right|_{s=1}\mathbf{r}_1^t(1),\end{aligned}$$

since $Q_1(1)Q_2(1) = P^2(1)$.

Renyi's Entropy

Let $H_s(Z_1^n)$ denote the Rényi's entropy of order s, that is,

$$H_s(Z_1^n) = \frac{\log \sum_{z_1^n} P^s(z_1^n)}{1-s}.$$

Then the entropy rate is

$$h_s(Z) = h_s(X) + \frac{\varepsilon}{(1-s)\lambda(s)} \mathbf{l}_1(s) \left(\mathbf{Q}(s) - \mathbf{P}^2(s) \right) \mathbf{r}_1(s) + O(\varepsilon^2),$$

where the Markov Renyi's entropy rate is

$$h_s(X) = \frac{1}{1-s} \log \lambda(s).$$