
Mohammad Rezaeian

Department of Electrical and Electronic Engineering
University of Melbourne

SPA 2006
Paris

The Entropy Rate of a Process

For a stationary process \( \{X_n\}_{n=0}^{\infty} \), the entropy rate is

\[
\hat{H}(X) = \lim_{n \to \infty} \frac{1}{n} H(X_0, X_1, ..., X_n)
\]

Average information per symbol

= \lim_{n \to \infty} H(X_n \mid X_{n-1}, X_{n-2}, ..., X_0)  \quad \text{Steady state information rate}

- For an i.i.d process, the entropy rate is the entropy of any sample.
- For processes with memory the entropy rate can be computed in two cases,
  - A Markov process.
  - A hidden Markov process (HMP). Integral expressions for the entropy rate of hidden Markov process have been obtained.

A simple method for calculation of elements of the sequence of conditional entropies.

The Estimation Entropy corresponding to a pair of processes

Consider a pair of correlated processes.

\[ \{S_n\}_{n=0}^{\infty}, \quad S_n \in S, \]
\[ \{Z_n\}_{n=0}^{\infty}, \quad Z_n \in Z \]

Assume we observe \( Z_n \), and based on the observations history we estimate \( S_n \).

The uncertainty in the estimation, conditioned that we know all past observation is

\[ H(S_n|Z^{n-1}). \]

This is a function of \( n \). The limiting quantity, if exists, is a measure for per symbol estimation entropy

\[ \hat{H}(S/Z) = \lim_{n \to \infty} H(S_n|Z^{n-1}) \]

We show that for a hidden Markov process under mild conditions the limit exists.

Expected applications:

- Sensor scheduling for observation of a Markov process.
- Error probability evaluation for channel coding.

The estimation entropy is a benchmark for showing how well an estimator is working.
Comparison of Entropy Rate and Estimation Entropy for a pair of processes

\[ \hat{H}(Z) = \lim_{n \to \infty} H(Z_n | Z^{n-1}) \]  

\[ \hat{H}(S/Z) = \lim_{n \to \infty} H(S_n | Z^{n-1}) \]  

Entropy Rate

Estimation Entropy

If we have enough past observation, what is the ambiguity for the next observation?

\[ \hat{H}(Z) \simeq H(Z_1 | Z^{0}_{-\infty}) \]

If we have enough past observation, what is the ambiguity about (or for the estimation of) the current value of the unobserved process, using the best possible estimator?

\[ \hat{H}(S/Z) \simeq H(S_1 | Z^{0}_{-\infty}) \]
Example of correlated processes: The hidden Markov process (HMP)

A hidden Markov process (HMP) is a noisy observation of a Markov process through a memoryless channel. A finite alphabet HMP is defined by a transition probability $P$ and emission matrix $T$.

For a HMP the random vector

$$\pi_n = p(S_n | Z^{n-1})$$

$$\pi_n[k] = p(S_n = k | Z^{n-1})$$

is called the information-state at time $n$.

The main properties of the information-state process

- The information-state process is a Markov process itself.
- It is a sufficient statistics for the observation process $p(Z_n | \pi_n, Z^{n-1}) = p(Z_n | \pi_n)$.
- The recursion

$$\pi_{n+1} = f(z, \pi_n) \triangleq \frac{\pi_n D(z) P}{\pi_n D(z) 1}.$$  \hspace{1cm} (1)

$D(z)$ is a diagonal matrix with $d_{ii}(z) = T[i, z]$, $i = 1, \ldots, |S|$.
Integral representation of conditional entropy

The conditional entropy $H(Y|X) \triangleq -E \log Pr(y|x)$ can be written as the integral of
Entropy function with some measure

$$H(Y|X) = \int_{\nabla_y} h(\omega')\psi_X(\omega')d\omega', \quad (2)$$

$$\psi_X(\omega) = \sum_x p(x)\delta(\omega - p(Y|x))$$

$\nabla_y$ = probability simplex in $\mathcal{R}^{[Y]}$

$\psi_X(.)$ is the distribution of random variable $\omega \in \nabla_y$, $\omega(X) = p(Y|X)$. 

$h : \nabla_y \rightarrow \mathbb{R}^+$
The Entropy Rate corresponds to a probability measure.

For a general process \( \{Z_n\}_{n=0}^{\infty} \), by defining the random vector

\[ q_n = p(Z_n|Z^{n-1}) , \quad q_n \in \nabla Z . \]

\[ H(Z_n|Z^{n-1}) = \int_{\nabla Z} h(q') \psi_n(q') dq' = E h(q_n), \tag{3} \]

The measure \( \psi_n(.) \) is the distribution of the random variable \( q_n \).

\[ \hat{H}(Z) = \lim_{n \to \infty} E h(q_n) = \int_{\nabla Z} h(q) \psi(q) dq . \]

\( \psi(.) = \lim_{n \to \infty} \psi_n(.) \)

![Diagram representing an i.i.d process, a Markov process, and a general process with arrows and labels](image)
The Estimation Entropy of HMP by a limiting distribution.

For an HMP by defining the random vector $\pi_n$, we get

$$\pi_n = p(S_n|Z^{n-1}), \quad \pi_n \in \nabla S.$$ 

$$H(S_n|Z^{n-1}) = \int_{\nabla S} h(\pi) \mu_n(\pi) d\pi = Eh(\pi_n),$$

The measure $\mu_n(.)$ is the distribution of the information-state process at time $n$.

$\mu_n$ belongs to $M^1(\nabla S) \equiv$ The space of probability measures on $\nabla S$.

If $\mu_n$ converges (in weak topology) to a $\mu \in M^1(\nabla S)$, then

$$\hat{H}(S/Z) = \lim_{n \to \infty} \ H(S_n|Z^{n-1}) = \int_{\nabla S} h(\pi) \mu(\pi) d\pi.$$ 

A sequence $\{\mu_n\}_{n=0}^\infty$ converges to $\mu$ in weak topology if for any bounded continuous $f$,

$$\lim_{n \to \infty} \int f d\mu_n = \int f d\mu.$$
Proof of convergence for HMP

We show that there is a deterministic continuous operator $\Phi$ on $M^1(\nabla S)$ such that

$$\mu_{n+1} = \Phi \mu_n.$$ 

We show that such an operator has a fixed point, and $\Phi^n \nu$ for any $\nu \in M^1(\nabla S)$ converges to it.

Iterated Function System (IFS), $\mathcal{F}(\Delta, F_i, p_i)_{i=1, 2, \ldots, k}$ is defined by the set of functions: $F_i : \Delta \rightarrow \Delta$ and $p_i : \Delta \rightarrow R^+$, such that $\sum p_i = 1$.

A continuous IFS induces a Feller operator $\Phi : M^1(\Delta) \rightarrow M^1(\Delta)$.

$$\Phi(\nu)(B) = \sum_i \int_{F_i^{-1}(B)} p_i d\nu.$$  

for $\nu \in M^1(\Delta)$

An HMP corresponds to an IFS,

The $|S| \times |Z|$ matrix $T$ defines the set of continuous functions $p_i : \nabla S \rightarrow R$, $i = 1, 2, \ldots, |Z|$, such that $\sum p_i = 1$.

The recursion (1) defines the set of continuous functions $F_i : \nabla S \rightarrow \nabla S$.

We have

$$\mu_{n+1} = \Phi \mu_n.$$
Definitions:
The measure $\mu$ is **attractive** if $\Phi^n \nu \to \mu$ for any $\nu \in M^1(\nabla_S)$.
The measure $\mu$ is **invariant** if $\Phi \mu = \mu$.

An IFS is **asymptotically stable** if it admits an attractive and invariant measure.

**Lemma 1** The IFS defined by

$$ (F_i(\pi))_j = \frac{\sum_{l=1}^{d} x_l T_{li} P_{lj}}{\sum_{l=1}^{d} x_l T_{li}} $$

$$ p_i(\pi) = (\pi T)[i], $$

where $P$ is primitive and $T$ has nonzero entries, is asymptotically stable.

This is the IFS representing HMP. Therefore:

Under above conditions, there exists an attractive and invariant measure $\mu$.

$$ \mu = \lim_{n \to \infty} \Phi^n \nu, \quad \forall \nu \in M^1(\nabla_S) $$

The sequence $\{\mu_n\}_{n=0}^{\infty}$ where $\mu_{n+1} = \Phi \mu_n$ converges to $\mu$ in weak topology.

The sequence of conditional entropy $H(S_n|Z^{n-1})$ converges to a number.
Basic properties of the Feller operator $\Phi$

- $\Phi$ is continuous in the weak topology.
- Any attractive measure is invariant.
- For any measurable function $h$ on $\nabla_S$,

$$\int hd(\Phi\nu) = \int (Uh)d\nu$$  \hspace{1cm} (4)

where operator $U$ maps a measurable function $h$ to the measurable function $Uh$ defined by

$$Uh(x) = \sum_i p_i(x)h(F_i(x)),$$

for $x \in \nabla_S$

$U$ is the operator conjugate to $\Phi$.

By repetition of (4), we get,

$$\int hd(\Phi^n\nu) = \int (U^nh)d\nu,$$

We show for $h$ being the entropy function,

Therefore if $\nu = \delta_x$,

Now if $x = x^*$ where $x^* = x^*P$,

then conditioning will be ineffective, so

$$H(S_n|Z_0^{n-1}, \pi_0 = x) = U^nh(x).$$

$$H(S_n|Z_0^{n-1}, \pi_0 = x) = \int hd(\Phi^n\delta_x).$$

$$H(S_n|Z_0^{n-1}) = \int hd(\Phi^n\delta_{x^*})$$

comparing to previous equality for $H(S_n|Z_0^{n-1})$,

$$\mu_n = \Phi^n\delta_{x^*} \quad \text{i.e.:} \quad \mu_0 = \delta_{x^*}.$$
From \( \mu_0 = \delta_{x^*} \), \( \mu_n = \Phi^n \delta_{x^*} \) is a probability mass function for any \( n \).

This results in a simple numerical method for calculating \( \mu_n \), and therefore \( H(S_n|Z^{n-1}) \).

Considering the definition of \( \Phi \), we can generate iteratively sets \( U_n = \{ u_1, u_3, ..., u_{|Z^n|} \} \), and a probability distribution \( \mu_n(\cdot) \) on these sets, starting from \( U_0 = \{ x^* \} \).

- \( \mu_n(u) = \mu_{n-1}(v)(vT)[z] \) for \( u = f(z, v), v \in U_{n-1}, z = 1, 2, ..., |Z| \).
- calculate (\( h_1 \) and \( h_2 \) are entropy functions over \( \nabla_Z \) and \( \nabla_S \) respectively)

\[
\begin{align*}
H^n_Z &= \sum_{i=1}^{|Z|^n} \mu_n(u_i) h_1(u_i T), \quad u_i \in U_n, \\
H^n_{S/Z} &= \sum_{i=1}^{|Z|^n} \mu_n(u_i) h_2(u_i), \quad u_i \in U_n,
\end{align*}
\]

- Entropy rate and estimation entropy are the limit of the above sequences.