The Entropy Rate of the Hidden Markov Process

Mohammad Rezaeian

Department of Electrical and Electronic Engineering
University of Melbourne

30th Conference on Stochastic Processes and their Applications
University of California, Santa Barbara

June 28, 2005
Outline

1. The hidden Markov process.
2. The entropy rate of a process.
3. A new expression for the entropy rate of hidden Markov process.
4. Comparison with previous expressions.
5. The numerical method for computation of the entropy rate.
6. Conclusion.
Hidden Markov process

Consider the two processes

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

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

\( S_n \) is a sufficient statistics for both \( S_{n+1} \) and \( Z_n \),

\[
p(s_{n+1}|s^n, z^{n-1}) = p(s_{n+1}|s_n) \quad \text{Markovity}
\]

\[
p(z_n|s^n, z^{n-1}) = p(z_n|s_n) \quad \text{Memoryless}
\]

Then

- \( \{S_n\}_{n=0}^\infty \) is a Markov process
- \( \{Z_n\}_{n=0}^\infty \) is a hidden Markov process.
A hidden Markov process is a noisy observation of a Markov process through a memoryless channel.

The defining parameters of a HMP are $[S, P, Z, T]$,

- Each row of $P$ is a probability vector, $P^{(s)} \in \nabla S$, $p(s_{n+1}|s_n) = P[s_n, s_{n+1}]$. 
- Each row of $T$ is a probability vector, $T^{(s)} \in \nabla Z$, $p(z_n|s_n) = T[s_n, z_n]$

$\nabla S$ and $\nabla Z$ are the probability simplexes.
The Entropy of a Random Variable

\( X \) A random variable.

\( \mathcal{X} \) Its domain (a discrete set)

\( \nabla \mathcal{X} \) The probability simplex of dimension \(|\mathcal{X}| - 1\).

\( p(X) \) The distribution of \( X \), \( p(X) \in \nabla \mathcal{X} \).

\( H(X) \) The entropy of \( X \).

\[
H(X) = -E_X \log Pr(x) = -\sum_x Pr(x) \log Pr(x).
\]

The function \( h : \nabla \mathcal{X} \to R^+ \), where \( h(p(X)) = H(X) \) is the entropy function.
The Entropy Rate of a Process

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

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

Average information per symbol

\[
= \lim_{n \to \infty} H(X_n | X_{n-1}, X_{n-2}, \ldots, X_0)
\]

Steady state information per symbol

- 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 if it is a Markov process.
- **Objective:** Compute the entropy rate of the hidden Markov process.

- Blackwell, 1957, *An expression for the entropy rate by a measure as the solution of an integral equation.*
- Recently 2002-2004, *Bounds and simple expressions for special cases of hidden Markov process has been investigated.*
The Entropy Rate of a Markov process

A Markov process is identified by a triple \([S, P, \mu_0]\).

The stationary distribution is the solution \(\nu\), in

\[\nu P = \nu, \quad \nu \in \nabla_S\]

For a Markov process, the entropy rate is

\[
\hat{H}(S) = E_S h(P^{(s)}) \quad S \in S, \quad P^{(s)} \in \nabla_S
\]  

(1)

where the distribution of \(S\) is the stationary distribution of the Markov process.
The entropy rate corresponds to a probability measure

An important relation for conditional entropy $H(Y|X) \triangleq -E \log Pr(y|x)$

\[
H(Y|X) = \int_{\nabla Y} h(\omega') \psi_X(\omega') d\omega', \quad (2)
\]

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

is a measure on $\nabla Y$ which depends on the disruption of $X$. 

\[
\nabla Y \quad \quad h : \nabla Y \to \mathbb{R}^+
\]

\[
p(Y|x_1) \quad p(Y|x_2) \quad p(Y|x_3)
\]
The probability measure is in general a limiting distribution.

For a general process \( \{Z_n\}_{n=0}^{\infty} \), by defining the random variable
\[
q_n = p(Z_n|Z^{n-1}), \quad q_n \in \nabla Z.
\]

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

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

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

\[
\psi(.) = \lim_{n \to \infty} \psi_n(.)
\]
The entropy rate of the hidden Markov process

For a hidden Markov model, the entropy rate of the measurement process \( \{Z_n\}_{n=0}^{\infty} \) is

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

\[
q_n = p(Z_n | Z^{n-1})
\]

**Problem:** The limiting distribution \( \psi(.) \) is not computable.

**Resolution:** Define the random variable

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

- The distribution of \( \pi_n \) (denoted by \( \mu_n(.) \)) is computable.
- \( q_n = \pi_n \times T \)
The entropy rate of the hidden Markov process

For a hidden Markov model, the entropy rate of the measurement process \( \{Z_n\}_{n=0}^{\infty} \) is

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

\[
q_n = p(Z_n|Z^{n-1})
\]

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

\[
q_n = \pi_n \times T
\]

\[
\hat{H}(Z) = \lim_{n \to \infty} Eh(\pi_n \times T) = \int_{\nabla S} h(\pi_n \times T)\mu(\pi) d\pi.
\]
For the hidden Markov process, the entropy rate can be described by a measure which is the stationary distribution of a Markov process.

\[ \hat{H}(Z) = \int_{\nabla Z} h(q)\psi(q) dq. \]

Using a mapping \( \nabla_S \rightarrow \nabla_Z \), we can define the entropy rate by a measure on \( \nabla_S \).

[Blackwell 1957]

\[ \hat{H}(Z) = \int_{\nabla S} h(\pi' \times P \times T)\varphi(\pi')d\pi'. \quad q = \pi' \times P \times T \]

\( \varphi \) is the state distribution of a stationary Markov process, and the solution of the integral equation,

\[ \varphi(E) = \sum_z \int_{f^{-1}_{zE}} r_z(w)d\varphi(w), \]

New result

\[ \hat{H}(Z) = \int_{\nabla S} h(\pi \times T)\mu(\pi)d\pi. \quad q = \pi \times T \]

\( \mu \) is the stationary distribution of a non-stationary Markov process with the transition probabilities and initial distribution explicitly defined by \( P \) and \( T \).
The information-state process

The Hidden Markov Process is a special case of the Partially Observed Markov Decision Processes (POMDP).

For a POMDP the random variable

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

is called the information-state at time \( n \).

The main properties of the information-state process

- Sufficient statistics for the observation process \( p(Z_n|\pi_n, Z^{n-1}) = p(Z_n|\pi_n) \).
- The recursion \( \pi_{n+1} = f(Z_n, \pi_n) \), using Bayes' rule and total probability.
- \( q_n = \pi_n \times T \), where \( q_n = p(Z_n|Z^{n-1}) \).
The Markovity of the information-state process

The information-state process \( \{\pi_n\}_{n=0}^{\infty} \) is a non-stationary Markov process with

- Initial distribution \( p(\pi_0) = \delta(\pi_0 - \nu) \)
- Conditional probabilities

\[
p_K(\pi_{n+1} = \alpha | \pi_n = \beta) = \begin{cases} 
q[z](\beta), & \exists z \in \mathcal{Z} : \alpha = f(z, \beta); \\
0, & \text{otherwise,}
\end{cases}
\]

This allows the computation of the distribution \( \mu_n(\cdot) \), and the measure \( \mu(\cdot) \) on \( \nabla S \).
The simple algorithm for computation of entropy rate

In iteration $n$ we generate the set $U_n = \{u_1, u_3, \ldots, u_{|Z_n|}\}$, and a probability distribution on this set $\mu_n(u_i)$, from the previous iteration, starting from $U_0 = \{\nu\}$.

The recursive generation of sets $U_n$ and the distributions $\mu_n$ is based on the transition probabilities of the information state process.

The sequence of $H_n = \sum_{i=1}^{|Z|} \mu_n(u_i) h(u_iT) = H(Z_n | Z^{n-1})$ converges monotonically from above to the entropy rate.

\[
P = \begin{pmatrix}
0.02 & 0.03 & 0.05 & 0.9 \\
0.8 & 0.06 & 0.04 & 0.1 \\
0.1 & 0.7 & 0.15 & 0.05 \\
0.9 & 0.03 & 0.03 & 0.04
\end{pmatrix}
\]

\[
T = \begin{pmatrix}
0.1 & 0.2 & 0.5 & 0.2 \\
0.6 & 0.1 & 0.2 & 0.1 \\
0.5 & 0.2 & 0.1 & 0.2 \\
0.3 & 0.2 & 0.1 & 0.4
\end{pmatrix}.
\]
Conclusion

The entropy rate for a general process is

\[ \hat{H}(Z) = E_{\pi} h(\pi), \quad \pi \in \nabla \mathcal{Z} \]

but the measure for this expectation does not have a close form or there is not an algorithm to find it.

For the hidden Markov process

[Blackwell, 1957]

\[ \hat{H}(Z) = E_{\pi} h(\pi \times P \times T), \quad \pi \in \nabla \mathcal{S} \]

The distribution of \( \pi \) needs to be found by solving an integral equation.

New expression,

\[ \hat{H}(Z) = E_{\pi} h(\pi \times T), \quad \pi \in \nabla \mathcal{S} \]

The distribution of \( \pi \) is obtained numerically, using the information-state process.