Hidden Markov Process: A New Representation, Entropy Rate and Estimation Entropy

Mohammad Rezaeian, *Member,IEEE*

Abstract—We consider a pair of stationary correlated processes $\{Z_n\}_{n=-\infty}^{\infty}$ and $\{S_n\}_{n=-\infty}^{\infty}$, where the former is observable and the later is hidden. The uncertainty in the estimation of Z_n upon its finite past history Z_0^{n-1} is $H(Z_n|Z_0^{n-1})$, and for estimation of S_n upon this observation is $H(S_n|Z_0^{n-1})$, which are both sequences of n. The limits of these sequences (and their existence) are of practical and theoretical interest. The first limit is the entropy rate. We call the second limit the *estimation entropy*. An example of a process jointly correlated to another one is the hidden Markov process. It is the memoryless observation of the Markov state process where state transitions are independent of past observations. We consider a new representation of hidden Markov process using iterated function system. In this representation the state transitions are deterministically related to the process. By this representation we analyze the two dynamical entropies for this process, which results in integral expressions for the limits. This analysis shows that under mild conditions the limits exist and provides a simple method for calculating the elements of the corresponding sequences.

Index Terms—entropy rate, hidden Markov process, iterated function system, estimation entropy.

I. Introduction

A stochastic process which is a noisy observation of a Markov process through a memoryless channel is called a hidden Markov process (HMP). In many applications of stochastic signal processing such as radar and speech processing, the output of the information source can be considered as an HMP. The entropy rate of HMP as the limit of compressibility of information source thus have special interest in those applications. Moreover, in the additive noise channels the noise process can be characterized as a hidden Markov process and its entropy rate is the defining factor in the capacity of channel. Finding the entropy rate of the hidden Markov process is thereby motivated by both applications in stochastic signal processing, source coding and channel capacity computation.

The study of the entropy rate of HMP started in 1957 by Blackwell [1] who obtained an integral expression for the entropy rate. This expression is defined through a measure described by an integral equation which is hard to extract from the equation in any explicit way. Bounds on the entropy rate can be computed based on the conditional entropies on sets of finite number of random variables [2]. Recent approaches for calculating the entropy rate are Monte Carlo simulation [3] and Lyapunov exponent [3],[4]. However these approaches yield indeterministic and hard to evaluate expressions. Simple

expression for the entropy rate has been recently obtained for special cases where the parameters of hidden Markov source approach zero [4],[5].

The hidden Markov process is a process defined through its stochastic relation to another process. The entropy rate of HMP thus corresponds to this relation and the dynamic of the underlying process. However this entropy rate only indicates the residual uncertainty in the symbol one step ahead of observation of the process itself. It doesn't indicate our uncertainty about the underlying process. In this paper we define estimation entropy as a variation of entropy rate to indicate this uncertainty. In general for a pair of correlated processes which one of them is hidden and the other is observable we can define estimation entropy as the long run per symbol uncertainty in the estimation of the hidden process based on the past observation. Such an entropy measure will be an important criterion for evaluating the performance of an estimator. In this paper we jointly analyze the entropy rate and estimation entropy for a hidden Markov process. This analysis is based on a mathematical model, namely the iterated function system [6], which suits the dynamics of the information state process of the HMP. This analysis results in integral expressions for these two dynamical entropies. We also derive a numerical method for iteratively calculating entropy rate and estimation entropy for HMP.

In this paper a discrete random variable is denoted by upper case and its realization by lower case. A sequence of random variables $X_0, X_1, X_2, ...X_n$ is denoted by X_0^n, whereas X^n refers to $X_{-\infty}^n$. The probability $Pr(X = x)$ is shown by $p(x)$ (similarly for conditional probabilities), whereas $\underline{p}(X)$ represents a row vector as the distribution of X, ie: the k-th element of the vector $\underline{p}(X)$ is $Pr(X = k)$. For a random variable X defined on a set $\mathcal{X}$, we denote by $\nabla_{\mathcal{X}}$ the probability simplex in $\mathbb{R}^{|\mathcal{X}|}$. A specific elements of a vector or matrix is referred to by its index in square brackets or as a subscript. The z-th row of matrix A is represented by $A^{(z)}$. The entropy of a random variable X is denoted by $H(.)$ whereas $h : \nabla_{\mathcal{X}} \rightarrow R^+$ represents the entropy function over $\nabla_{\mathcal{X}}$, i.e: $h(\underline{p}(X)) = H(X)$ for all possible random variables X on $\mathcal{X}$. Our notation does not distinguish differential entropies from ordinary entropies.

In the next section we define the iterated function system and draw some results from [6], as well as a new result. In section III we define the hidden Markov process by identifying the key properties for the probability distributions on the corresponding domain sets and show that such a process can be represented by an iterated function system . In sections IV and V we derive integral expressions for entropy rate and estimation entropy followed by a method for calculating them.

II. ITERATED FUNCTION SYSTEMS

Consider a system with a state in the space of Δ, where the state transitions depends deterministically to a correlated process taking values in a set $\mathcal{I}_m = \{1, 2, ..., m\}$ and stochastically depending on the state. The mathematical model representing such a system is an *iterated function system* (IFS) which is defined by m functions transforming a metric space Δ to itself, and m place dependent probabilities.

Definition 1: A triple $\mathcal{F} = (\Delta, F_i, q_i)_{i=1,2...,K}$ is an iterated function system if $F_i : \Delta \to \Delta$ and $q_i : \Delta \to \mathcal{R}^+$ are measurable functions and $\sum_i q_i = 1$.

Although the generality of IFS allows the functions of F_i and q_i to be measurable which is a wide range of real functions, in this paper we are only interested in a subset of those functions, the continuous functions. Such systems are referred to as continuous IFS. If the functions F_i's are only defined on $\Delta_i \subset \Delta$, where $\Delta_i = \{x \in \Delta : q(x_i) > 0\}$, then the IFS is called partial iterated function system (PIFS). Although the general application of IFS in this paper could be involved PIFS, we avoid such complexity by restricting the application.

The IFS represents the above mentioned dynamical system where the probability of event $i \in \mathcal{I}_m$ under state $x \in \Delta$ is $q_i(x)$ and the consequence of such event is the change of state to $F_i(x)$.

Consider $\mathcal{M}^1(\Delta)$ as the space of probability measures on Δ. For an $\mathcal{F}$ we define an operator $\Phi : \mathcal{M}^1(\Delta) \to \mathcal{M}^1(\Delta)$,

$$\Phi(\mu)(B) = \sum_i \int 1_B(F_i(x))q_i(x)\mu(dx), \tag{1}$$

for $\mu \in M^1(\Delta)$ and $B \subset \Delta$. The operator Φ, induced by $\mathcal{F}$, represents the evolution of probability measures under the action of $\mathcal{F}$. More specifically, if our belief on the state of system at time n is the probability measure μ_n, ($\mu_n \in \mathcal{M}^1(\Delta)$), then this belief at time $n+1$ is

$$\mu_{n+1} = \Phi\mu_n, \tag{2}$$

which can be easily verified by Equation (1) and role of functions F_i and q_i. Note that the operator Φ is deterministic and it is affine, i.e: $\Phi(\lambda\mu_1 + (1-\lambda)\mu_2) = \lambda\Phi\mu_1 + (1-\lambda)\Phi\mu_2$. By such representation Φ is a so called *Markov operator*.

For a Markov operator Φ acting on the space $\mathcal{M}^1(\Delta)$ a measure $\mu \in \mathcal{M}^1(\Delta)$ is *invariant* if $\mu = \Phi\mu$, and it is *attractive* if

$$\mu = \lim_{n\to\infty} \Phi^n \nu, \tag{3}$$

for any $\nu \in \mathcal{M}^1(\Delta)$. A Markov operator Φ (and the corresponding IFS) is called *asymptotically stable* if it admits an invariant and attractive measure. The concept of limit in Equations (3) is convergence in weak topology, meaning

$$\int f d\mu = \lim_{n\to\infty} \int f d(\Phi^n \nu), \tag{4}$$

for any continuous bounded function f. Note that the limit doesn't necessarily exist or it is not necessarily unique. The set of all attractive measures of Φ for $\mathcal{F}$ is denoted by $\mathcal{S}^{\mathcal{F}}$.

A Markov operator which is continuous in weak topology is a Feller operator. We can show that for a continuous IFS the operator Φ is a Feller operator. In this case any $\mu \in \mathcal{S}^{\mathcal{F}}$ is invariant.

Let $B(\Delta)$ be the space of all real valued continuous bounded functions on Δ. A special property of a Feller operator $\Phi : \mathcal{M}^1(\Delta) \to \mathcal{M}^1(\Delta)$ is that there exists an operator $\mathcal{U} : B(\Delta) \to B(\Delta)$ such that:

$$\int f(x)\Phi\mu(dx) = \int \mathcal{U}f(x)\mu(dx), \tag{5}$$

for all $f \in B(\Delta), \mu \in \mathcal{M}^1(\Delta)$. The operator $\mathcal{U}$ is called the operator conjugate to Φ. It can be shown [6] that for a continuous IFS the operator conjugate of Φ is $\mathcal{U}$, where

$$(\mathcal{U}f)(x) = \sum_{i\in\mathcal{I}_k} q_i(x)f(F_i(x)). \tag{6}$$

For an IFS, the concept of change of state and probability of the correlated process in each step can be extended to $n > 1$ steps. For an $\mathbf{i} = (i_1, i_2, ...i_n) \in \mathcal{I}_m^n$, we denote

$$F_{\mathbf{i}}(x) = F_{i_n}(F_{i_{n-1}}(...F_{i_1}x)...))$$

$$q_{\mathbf{i}}(x) = q_{i1}(x)q_{i2}(F_{i1}(x))...q_{in}(F_{i_{n-1}}(F_{i_{n-2}}(...F_{i1}(x))))$$

Then the probability of the sequential event $\mathbf{i}$ under state $x \in \Delta$ is $q_{\mathbf{i}}(x)$ and as a result of such sequence, the state changes from x to $F_{\mathbf{i}}(x)$ in n steps. As an extension of (6), we can show

$$(\mathcal{U}^n f)(x) = \sum_{\mathbf{i}\in\mathcal{I}_m^n} p_{\mathbf{i}}(x)f(F_{\mathbf{i}}(x)). \tag{7}$$

In this paper we define for a given continuous IFS, and for a $f \in B(\Delta)$,

$$\hat{F}(x) \triangleq \lim_{n\to\infty} (\mathcal{U}^n f)(x). \tag{8}$$

Now we state our result on IFS in the following Lemma which will be used in Section IV as the major application of IFS to the purpose of this paper.

Lemma 1: For a continuous IFS $\mathcal{F} = (\Delta, F_i, p_i)_{i=1,2...,K}$, and any function $f \in B(\Delta)$,

$$\hat{F}(x) = \int f d\mu, \tag{9}$$

where $\mu = \lim_{n\to\infty} \Phi^n \delta_x$ (if the limit exists), and $\delta_x \in \mathcal{M}^1(\Delta)$ is a distribution with all probability mass at x.

Proof: From (5) we have

$$\int f d(\Phi^2 \mu) = \int \mathcal{U}f d(\Phi\mu) = \int (\mathcal{U}^2 f)d\mu,$$

where the first equality is by substituting μ with $\Phi\mu$ in (5) and the second equality by substituting f with $\mathcal{U}f$. Therefore by repetition of (5), we have

$$\int f d(\Phi^n \mu) = \int (\mathcal{U}^n f)d\mu, \tag{10}$$

for all $f \in B(\Delta), \mu \in \mathcal{M}^1(\Delta)$. This results in

$$\hat{F}(x) = \lim_{n\to\infty} \int (\mathcal{U}^n f)d\delta_x = \lim_{n\to\infty} \int f d(\Phi^n \delta_x) = \int f d\mu,$$

where the first equality is from the definition of $\hat{F}$ in (8) and the last one is from (4).

∎

From the above Lemma we infer that for an asymptotically stable continuous IFS, the function $\hat{F}$ is a constant independent of x. Note that asymptotic stability ensures that there exists at least one μ satisfying (4) for any $\nu \in \mathcal{M}^1(\Delta)$, which is true for $\nu = \delta_x$ for any x. If there are more than one $\mu \in \mathcal{S}^{\mathcal{F}}$, all of them has to satisfy (4). So in this case the Equality of (9) independent of x is true for any $\mu \in \mathcal{S}^{\mathcal{F}}$.

We use the result of this section in the analysis of estimation entropy of hidden Markov processes by specializing Δ to be the space of information state process and f to be variations of the entropy function.

III. THE HIDDEN MARKOV PROCESS

A hidden Markov process is a process related to an underlying Markov process through a discrete memoryless channel, so it is defined by transition probability matrix P of the Markov process and the emission matrix T of the memoryless channel [7],[8]. In this paper a hidden Markov process $\{Z_n\}_{n=-\infty}^{\infty}$, $Z_n \in \mathcal{Z}$ and it underlying Markov process $\{S_n\}_{n=-\infty}^{\infty}$, $S_n \in \mathcal{S}$ are considered to be stationary. The elements of matrices $P_{|\mathcal{S}| \times |\mathcal{S}|}$ and $T_{|\mathcal{S}| \times |\mathcal{Z}|}$ are the conditional probabilities,

$$
\begin{aligned}
P[s, s'] &= p(S_{n+1} = s'|S_n = s), \\
T[s, z] &= p(Z_n = z|S_n = s).
\end{aligned}
\tag{11}
$$

A pair of matrices P and T uniquely define a hidden Markov process on the state set $\mathcal{S}$ and observation set $\mathcal{Z}$ by the following basic properties, for any n.

- A1: *Markovity,*

$$
p(s_n|s^{n-1}) = p_P(s_n|s_{n-1}),
\tag{12}
$$

  where $p_P(s_n|s_{n-1}) = P[s_{n-1}, s_n]$.
- A2: *Sufficient Statistics of State,*

$$
p(s_n|s_{n-1}, z^{n-1}) = p_P(s_n|s_{n-1}),
\tag{13}
$$

  where $p_P(.|.)$ is defined by P.
- A3: *Memoryless observation,*

$$
p(z^n|s^n) = \prod_i^n p_T(z_i|s_i),
\tag{14}
$$

  where $p_T(z|s) = T[s, z]$.

Property A3 implies:

$$
P(z_n|s_n, z^{n-1}) = p_T(z_n|s_n).
\tag{15}
$$

For a hidden Markov process we define two random vectors π_n and ρ_n as functions of Z^{n-1} on the domains $\nabla_{\mathcal{S}}, \nabla_{\mathcal{Z}}$, respectively,

$$
\pi_n(Z^{n-1}) = \underline{p}(S_n|Z^{n-1}).
\tag{16}
$$

$$
\rho_n(Z^{n-1}) = \underline{p}(Z_n|Z^{n-1}).
\tag{17}
$$

According to our notation, the random vector π_n has elements $\pi_n[k]$, $k = 1, 2, ..., |\mathcal{S}|$,

$$
\pi_n[k] = p(S_n = k|Z^{n-1}),
\tag{18}
$$

and similarly for ρ_n. We obtain the relation between random vectors ρ_n and π_n

$$
\begin{aligned}
&\rho_n[m](Z^{n-1}) \\
&= Pr(Z_n = m|Z^{n-1}) \\
&= \sum_k Pr(Z_n = m|Z^{n-1}, S_n = k)Pr(S_n = k|Z^{n-1}) \\
&= \sum_k Pr(Z_n = m|S_n = k)Pr(S_n = k|Z^{n-1}) \\
&= \sum_k T[k, m]\pi_n[k](Z^{n-1}),
\end{aligned}
\tag{19}
$$

which shows the matrix relation

$$
\rho_n(Z^{n-1}) = \pi_n(Z^{n-1}) \times T.
$$

Therefore ρ_n and π_n are related by

$$
\rho_n = \pi_n T.
\tag{20}
$$

More generally, we refer to $\zeta(\pi) \in \nabla_{\mathcal{Z}}$ as the projection of $\pi \in \nabla_{\mathcal{S}}$ under the mapping $T : \nabla_{\mathcal{S}} \to \nabla_{\mathcal{Z}}$, i.e:

$$
\zeta(\pi) = \pi T.
\tag{21}
$$

The random vector π_n as a function of the random sequence Z^{n-1} is called *information-state* [9] since it encapsulates all the information about state at time n that can be obtained form the observations Z^{n-1}. More specifically, the information state at time n is a sufficient statistics for the observation process at this time, ie: conditioned on π_n, Z_n is independent of Z^{n-1}, π^{n-1} [9]. As a result,

$$
\underline{p}(Z_n|\pi_n) = \underline{p}(Z_n|\pi_n, Z^{n-1}) = \underline{p}(Z_n|Z^{n-1}) = \rho_n = \zeta(\pi_n).
\tag{22}
$$

Using Bayes' rule and the law of total probability, an iterative formula for the information state can be obtained as a function of Z_n [9],

$$
\pi_{n+1} = \eta(z_n, \pi_n),
\tag{23}
$$

where

$$
\eta(z, \pi) \triangleq \frac{\pi D(z) P}{\pi D(z) \underline{1}},
\tag{24}
$$

where $D(z)$ is a diagonal matrix with $d_{k,k}(z) = T[k, z]$, $k = 1, 2, .., |\mathcal{S}|$.

Due to the sufficient statistic property of the information state, we can consider the information state process $\{\pi_n\}_{n=0}^{\infty}$ on $\nabla_{\mathcal{S}}$ as the state process of an iterated function system on $\nabla_{\mathcal{S}}$ with the hidden Markov process being its correlated process. This is because the hidden Markov process at time k is stochastically related to the information state process at that time by $Pr(Z_k = z|\pi_k = x) = \zeta(x)[k]$ (from (22)). On the other hand, $Z_k = z$ result in the deterministic change of state from $\pi_k = x$ to $\pi_{k+1} = \eta(z, x)$. Consequently, for a hidden Markov process there is a continuous iterated function systems defined by, for different values $z \in \mathcal{Z}$,

$$
\begin{aligned}
F_z(x) &= \eta(z, x), \\
q_z(x) &= \zeta(x)[z],
\end{aligned}
\tag{25}
$$

where the equality $\sum_z q_z(x) = 1, x \in \nabla_{\mathcal{S}}$ is satisfied due to $\zeta(x) \in \nabla_{\mathcal{Z}}$. These functions are in fact conditional probabilities, $F_z(x) = \underline{p}(S_{k+1}|Z_k = z, \pi_k = x)$ and $q_z(x) = Pr(Z_k = z|\pi_k = x)$ for any k.

If the emission matrix T has zero entries, then function $\eta(z, x)$ could be indefinite for some (z, x). This happens for

those $x \in \nabla_{\mathcal{S}}$ that the element z of vector xT is zero[1], i.e: the functions $F_z(x)$ is only defined for x that $q_z(x) > 0$. Hence for the general choice of matrix T we have a PIFS associated to the hidden Markov process. For this and other reason that will reveals later we assume that matrix T has non zero entries.

For the continuous IFS related to the hidden Markov process, we can obtain the corresponding Feller operator Φ and its conjugate operator $\mathcal{U}$. The operator $\mathcal{U}$ maps any $f \in B(\Delta)$ to $\mathcal{U}f \in B(\Delta)$ where

$$(\mathcal{U}f)(x) = \sum_z q_z(x) f(F_z(x))$$
$$= \sum_z Pr(Z_k = z | \pi_k = x) f(\underline{p}(S_{k+1} | Z_k = z, \pi_k = x)). \quad (26)$$

In general given $\pi_k = x$, the probability of a specific n-sequence $\mathbf{z} = (z_1, z_2, ..., z_n)$ for the hidden Markov process is

$$Pr(Z_k^{k+n-1} = \mathbf{z} | \pi_k = x) = $$
$$q_{z_1}(x) q_{z_2}(F_{z_1}(x)) ... q_{z_n}(F_{z_{n-1}}(F_{z_{n-2}}(...F_1(x)))), \quad (27)$$

and this sequence changes the state to

$$\pi_{k+n} = F_{z_n}(F_{z_{n-1}}(...F_{z_1}x)...)). \quad$$

Therefore we can write

$$\underline{p}(S_{k+n} | Z_k^{k+n-1} = \mathbf{z}, \pi_k = x) = F_{z_n}(F_{z_{n-1}}(...F_{z_1}x)...)). \quad (28)$$

Comparing to (7), we infer

$$(\mathcal{U}^n f)(x) = $$
$$\sum_{\mathbf{z}} Pr(Z_k^{k+n-1} = \mathbf{z} | \pi_k = x) f(\underline{p}(S_{k+n} | Z_k^{k+n-1} = \mathbf{z}, \pi_k = x)). \quad (29)$$

The IFS corresponding to a hidden Markov process under a wide range of the parameters of the process is shown to be asymptotically stable.

Definition 2: A stochastic Matrix P is primitive if there exists an n such that $(P^n)_{i,j} > 0$ for all i, j.

Lemma 2: For a primitive matrix P and an emission matrix T with strictly positive entries, the IFS defined according to (25) is asymptotically stable.

Proof: The proof follows from [6, Theorem 8.1]. The IFS $\mathcal{F}^P$ defined in [6, Theorem 8.1] by

$$(F_i^P(x))_j \triangleq \frac{\sum_{l=1}^d x_l P_{lj} T_{li}}{\sum_{l=1}^d x_l T_{li}} = \eta(i, x)[j],$$

$$q_i^P(x) \triangleq \sum_{l=1}^d x_l T_{li} = \zeta(x)[i],$$

is the same as the IFS defined by (25). It is shown in [6, Theorem 8.1] that under the conditions of this lemma $\mathcal{F}^P$ is asymptotically hyperbolic, which then has to be asymptotically stable according to [6, Theorem 3.4]. ∎
A Markov chain with primitive transition matrix P is geometrically ergodic and has a unique stationary distribution [7].

---

[1]e.g: if $T_{1,1} = T_{2,1} = 0$, then for all π that have zero components on the third elements onward, both the nominator and denominators of (24) for $z = 1$ will be zero, and for those π's the first component of πT is zero.

IV. ENTROPY RATE AND ESTIMATION ENTROPY

The entropy of a random variable $Z \in \mathcal{Z}$ is a function of its distribution $\underline{p}(Z) \in \nabla_Z$,

$$H(Z) = h(\underline{p}(Z)) = \sum_z -p(z) \log p(z).$$

where

$$h(\rho) = \sum_{i=1}^{|\mathcal{Z}|} -\rho[i] \log(\rho[i]), \qquad \rho \in \nabla_{\mathcal{Z}}. \quad (30)$$

For a general process $\{Z_n\}_{n=-\infty}^{\infty}$, the entropy of any n-sequence Z_k^{k+n-1} is denoted by $H(Z_k^{k+n-1})$ which is defined by the joint probabilities $Pr(Z_k^{k+n-1} = \mathbf{z})$, for all $\mathbf{z} \in \mathcal{Z}^n$. For a stationary process these joint probabilities are invariant with k. The entropy rate of a stationary process is denoted by $\hat{H}(Z)$ and defined by

$$\hat{H}_Z = \lim_{n \to \infty} \frac{1}{n} H(Z_0^n). \quad (31)$$

It is shown that for a stationary process the limit always exists, and it is also equal to

$$\hat{H}_Z = \lim_{n \to \infty} H(Z_n | Z_0^{n-1}). \quad (32)$$

For a stationary Markov process with transition Matrix P, the entropy rate is

$$\hat{H}_Z = \lim_{n \to \infty} H(Z_n | Z_{n-1}) = H(Z_1 | Z_0) = \sum_i x[i] h(P^{(i)}), \quad (33)$$

where $x \in \nabla_Z$ is the stationary distribution of the Markov process, i.e: the solution of $xP = x$. Of special interest to this paper is the entropy rate of the hidden Markov process.

We can extend the concept of entropy rate to a pair of correlated processes. Assume we have a jointly correlated processes $\{Z_n\}_{n=-\infty}^{\infty}$ and $\{S_n\}_{n=-\infty}^{\infty}$ where we observe the first process and based on our observation estimate the state of the other process. The uncertainty in the estimation of S_n upon past observations Z_0^{n-1} is $H(S_n | Z_0^{n-1})$. The limit of this sequence which inversely measures the observability of the hidden process is of practical and theoretical interests. We call this limit *Estimation Entropy*,

$$\hat{H}_{S/Z} = \lim_{n \to \infty} H(S_n | Z_0^{n-1}), \quad (34)$$

if the limit exists. Practical application of estimation entropy is for example in sensor scheduling for observation of a Markov process [10]. The aim of such a scheduler is to find a policy for selection of sensors based on information-state which minimizes the estimation entropy, thus achieving the maximum observability for the Markov process. This entropy measure could also be related to the error probability in channel coding. The more the estimation entropy, the more uncertainty per symbol in the decoding process of the received signal, thus higher error probability. The estimation entropy can be viewed as a benchmark for indicating how well an estimator is working. It is the limit of minimum uncertainty that an estimator can achieve for estimating the current value of the unobserved process under the knowledge of enough history of observations. We consider HMP as a joint process and analysis its estimation entropy.

For a hidden Markov process the entropy rate $\hat{H}_Z$ and estimation entropy $\hat{H}_{S/Z}$ are the limiting expectations

$$\begin{aligned} \hat{H}_Z &= \lim_{n\to\infty} E[h(\rho_n)], \\ \hat{H}_{S/Z} &= \lim_{n\to\infty} E[h(\pi_n)]. \end{aligned} \tag{35}$$

However since π_n and ρ_n are functions of joint distributions of random variables Z_0^{n-1} these expectations are not directly computable. We use the IFS for a hidden Markov process to gain insight into these entropy measures and numerical method for calculating them.

Adapting Equation (1) with special functions $F_z(x)$ and $q_z(x)$ in 25, we obtain the Feller operator Φ for the IFS corresponding to a hidden Markov process.

$$\Phi(\mu)(B) = \sum_z \int_{\nabla_S} 1_B(\eta(z,x))\zeta(x)[z]\mu(dx). \tag{36}$$

To analysis the entropy measures $\hat{H}_Z$ and $\hat{H}_{S/Z}$, we define two intermediate functions

$$\begin{aligned} \hat{H}_Z(x) &= \lim_{n\to\infty} H(Z_n|Z_0^{n-1}, \pi_0 = x), \\ \hat{H}_{S/Z}(x) &= \lim_{n\to\infty} H(S_n|Z_0^{n-1}, \pi_0 = x). \end{aligned} \tag{37}$$

In comparison to (32) and (34), these functions are the corresponding per symbol entropies when it is conditioned on a specific prior distribution of state at time $n = 0$.

Now we use Lemma 1 to obtain an integral expressions for these limiting entropies.

Lemma 3: For a hidden Markov process

$$\begin{aligned} \hat{H}_Z(x) &= \int_{\nabla_S} (h_1 \circ \zeta)d\mu, \\ \hat{H}_{S/Z}(x) &= \int_{\nabla_S} h_2 d\mu, \end{aligned} \tag{38}$$

where $\mu = \lim_{n\to\infty} \Phi^n \delta_x$, and $h_1 : \nabla_Z \to \mathbb{R}^+$ and $h_2 : \nabla_S \to \mathbb{R}^+$ are entropy functions.

Proof: From definition of conditional entropy we write,

$$\begin{aligned} &H(Z_n|Z_0^{n-1}, \pi_0 = x) = \\ &\sum_{\mathbf{z}} Pr(Z_0^{n-1} = \mathbf{z}|\pi_k = x)h_1(\underline{p}(Z_n|Z_0^{n-1} = \mathbf{z}, \pi_0 = x)). \end{aligned} \tag{39}$$

Now since (as in (19), using $p(z_n|s_n, z^{n-1}, \pi_0) = p(z_n|s_n)$),

$$\underline{p}(Z_n|Z_0^{n-1} = \mathbf{z}, \pi_0 = x) = \zeta(\underline{p}(S_n|Z_0^{n-1} = \mathbf{z}, \pi_0 = x)), \tag{40}$$

Equation (39) can be written as

$$\begin{aligned} &H(Z_n|Z_0^{n-1}, \pi_0 = x) = \\ &\sum_{\mathbf{z}} Pr(Z_0^{n-1} = \mathbf{z}|\pi_k = x)h_1 \circ \zeta(\underline{p}(S_n|Z_0^{n-1} = \mathbf{z}, \pi_0 = x)). \end{aligned} \tag{41}$$

Similarly from definition of conditional entropy, we can write

$$\begin{aligned} &H(S_n|Z_0^{n-1}, \pi_0 = x) = \\ &\sum_{\mathbf{z}} Pr(Z_0^{n-1} = \mathbf{z}|\pi_k = x)h_2(\underline{p}(S_n|Z_0^{n-1} = \mathbf{z}, \pi_0 = x)). \end{aligned} \tag{42}$$

Comparing Equations (41) with (29), we have

$$\hat{H}_Z(x) = \lim_{n\to\infty} (\mathcal{U}^n(h_1 \circ \zeta))(x). \tag{43}$$

Similarly by (42),

$$\hat{H}_{S/Z}(x) = \lim_{n\to\infty} (\mathcal{U}^n h_2)(x). \tag{44}$$

Now considering Equation (8) and applying Lemma 1 we obtain (38). ∎

Lemmas 2 and 3 result in integral expressions for entropy rate and estimation entropy.

Theorem 1: For a hidden Markov process with primitive matrix P and the emission matrix T with strictly positive entries,

$$\begin{aligned} \hat{H}_Z &= \int_{\nabla_S} (h_1 \circ \zeta)d\mu, \\ \hat{H}_{S/Z} &= \int_{\nabla_S} h_2 d\mu, \end{aligned} \tag{45}$$

where μ is any attractive and invariant measure of operator Φ, and h_1, h_2 are the entropy functions on ∇_Z, ∇_S, respectively.

Proof: From Lemma 2, under the condition of this Theorem, the continuous IFS corresponding to the HMP is asymptotically stable. As it is discussed after Lemma1, in this case the functions $\hat{H}_Z(x)$ and $\hat{H}_{S/Z}(x)$ (in (43) and (44)) are independent of x and the equalities of (38) are satisfied for any attractive (which is also invariant) measure of Φ. Now the independency of x for $\hat{H}_Z(x)$ and $\hat{H}_{S/Z}(x)$ in (37) establishes the equalities in (45) for $\hat{H}_Z$ and $\hat{H}_{S/Z}$. Note that for a set of random variables X, Y, Z if $H(Y|Z, X = x)$ is invariant with x, then $H(Y|Z) = H(Y|Z, X) = H(Y|Z, X = x)$. ∎

The first equality in the above theorem has been previously obtained by a different approach in [11]. However in [11], the measure μ is restricted to be $\mu = \lim_{n\to\infty} \Phi^n \delta_{x^*}$, where x^* is the stationary distribution of the underlying Markov process defined by P,

$$x^* P = x^*. \tag{46}$$

The integral expression for $\hat{H}_Z$ in Theorem 1 is also the same as the expression in [6, Proposition 8.1] for $\theta = P$. For this case the integral expression is shown to be equal to both of the following two entropy measures

$$\begin{aligned} \mathcal{H}(x^*) &\triangleq \lim_{n\to\infty} \tfrac{1}{n} \sum_{\mathbf{z}\in\mathcal{Z}^n} q_{\mathbf{z}}(x^*) \log(q_{\mathbf{z}}(x^*)), \\ \mathcal{H}(\mu) &\triangleq \lim_{n\to\infty} \tfrac{1}{n} \sum_{\mathbf{z}\in\mathcal{Z}^n} \int q_{\mathbf{z}}(x)\mu(dx). \log(\int q_{\mathbf{z}}(x)\mu(dx)), \end{aligned} \tag{47}$$

where μ is the attractive measure of Φ for the IFS defined by (25). Considering $q_{\mathbf{z}}(x) = p(Z_0^{n-1} = \mathbf{z}|\pi_0 = x)$ for HMP, (27), the two equalities match with Lemma 3 and Theorem 1.

The integral expression for $\hat{H}_Z$ is also equivalent to the original Blackwell's formulation [1] by a change of variable x to xP. This is because the expression in [1] is derived based on $\alpha_{n-1} = \underline{p}(S_{n-1}|Z^{n-1})$ instead of $\pi_n = \alpha_{n-1}P$ in (16) (cf. (13)). The measure of integral also corresponds to this change of variable. Note that the measure μ in (45) satisfies

$$\mu(B) = \Phi(\mu)(B) = \sum_z \int_{F_z^{-1}(B)} (xT)[z]\mu(dx), \tag{48}$$

(cf. (36)) which is the same as the integral equation for the measure in [1] if we change the integrand of (48) to $r_z(x) = (xPT)[z]$ and instead of $F_z(x)$ use the function $f_z(x) = xPD(z)/r_z(x)$ (derived from (24) by $\pi = \alpha P$, satisfying $\alpha_{n+1} = f_z(\alpha_n)$).

VI. Conclusion

HMP is a process described by its relation to a Markov state process which has stochastic transition to the next state independent of the current realization of the process. In this paper we showed that HMP can be better described and more rigorously analyzed by iterated function systems whose state transitions are deterministically related to the process. In both descriptions the state is hidden and the process at any time is stochastically related to the state at that time.

In this paper we also introduced the concept of estimation entropy for a pair of joint processes which has practical applications. The entropy rate for a process, like HMP, which is correlated to another process can be viewed as the self estimation entropy. Both entropy rate and estimation entropy for the hidden Markov process can be analyzed using the iterated function system description of the process. This analysis results in integral expressions for these dynamical entropies. The integral expressions are based on an attractive and invariant measure of the Markov operator induced by the iterated function system. These integrals can be evaluated numerically as the limit of special numerical sequences.

VII. Acknowledgment

References

[1] D.Blackwell, "The entropy of functions of finite-state Markov chains", Trans. First Prague Conf. Inf. Th., Statistical Decision Functions, Random Processes, page 13-20, 1957.

[2] T.M.Cover and J.A.Thomas. "Elements of Information Theory", Wiley, New York, 1991.

[3] T. Holiday, P. Glynn, and A. Goldsmith," Capacity of finite state Markov channels with general inputs", *Int. Symp. Inf. Th.*, p. 289, Yokohama, Japan, July 2003.

[4] P.Jacquet, G.Seroussi, and W. Szpankowski,"On the entropy rate of a hidden Markov process. *Int. Symp. Inf. Th.*, p.10, Chicago, IL, July 2004."

[5] E. Ordentlich and T. Weissman" New Bounds on the Entropy Rate of Hidden Markov Processes", San Antonio Information Theory Workshop, October 2004.

[6] W. Slomczynski. "Dynamical Entropy, Markov Operators and Iterated Function Systems", Wydawnictwo Uniwersytetu Jagiellonskiego, ISBN 83-233-1769-0, Krakow, 2003.

[7] Y. Ephraim and N. Merhav, "Hidden Markov Processes," *IEEE Trans. Inform. Theory*, vol. IT-48 No.6 , pp. 1518-1569, June 2002.

[8] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of the IEEE, vol 77, No 2, February 1989, pp. 257-286.

[9] R.D. Smallwood and E.J. Sondik, "Optimal control of partially observed Markov processes over a finite horizon," *Operation Research*, vol.21, pp. 1071-1088, 1973.

[10] —, "Minimum entropy scheduling for hidden Markov procesess," *Raytheon Systems Company* internal report, Integrated Sensing Processor Phase II. March 2006.

[11] M. Rezaeian, "The entropy rate of the hidden Markov process," submitted to *IEEE Trans. Inform. Theory*, May 2005.

[12] M. Rezaeian, "Symmetric characterization of finite state Markov channels," *IEEE Int. Symp. Inform. Theory*, July 2006, Seattle, USA.