Capacity of Finite State Markov Channels with General Inputs

Tim Holliday
Stanford University
e-mail: thollida@stanford.edu

Andrea Goldsmith
Stanford University
e-mail: andrea@systems.stanford.edu

Peter Glynn
Stanford University
e-mail: glynn@stanford.edu

Abstract — **We study new formulae based on Lyapunov exponents for entropy, mutual information, and capacity of finite state discrete time Markov channels. We also develop a method for directly computing mutual information and entropy using continuous state space Markov chains. We show that the entropy rate for a symbol sequence is equal to the primary Lyapunov exponent for a product of random matrices. We then develop a continuous state space Markov chain formulation that allows us to directly compute entropy rates as expectations with respect to the Markov chain's stationary distribution. We also show that the stationary distribution is a continuous function of the input symbol dynamics. This continuity allows the channel capacity to be written in terms of Lyapunov exponents.**

I. Channel Model

Let $C = (C_n : n \geq 0)$ be a stationary finite-state irreducible Markov chain living on state space $\mathcal{C}$. The random sequences of observed inputs and outputs will be denoted $X = (X_n : n \geq 0)$ and $Y = (Y_n : n \geq 0)$, and take values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. For each pair of states $(c_n, c_{n+1}) \in \mathcal{C}$ define a probability distribution on the input/output symbols $p(x_n, y_n|c_n, c_{n+1})$. Further assume the input/output sequences X and Y have a joint distribution specified by

$$p((X_0 = x_0, Y_0 = y_0), \ldots, (X_n = x_n, Y_n = y_n)|c^{n+1}) = \prod_{i=0}^{n} p((x_i, y_i)|C_i = c_i, C_{i+1} = c_{i+1}).$$

This framework for the channel incorporates a substantial number of interesting channel models. Of particular interest are finite-state Markov channels with Markov inputs, the capacities of which are currently open problems.

II. Entropies as Lyapunov Exponents

With the channel model described above, each of the entropies $H(X)$, $H(Y)$, and $H(X,Y)$ turn out to be Lyapunov exponents for products of random matrices (up to a change in sign).

Proposition 1: For $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, let $G_x^X = (G_x^X(c_0, c_1) : c_0, c_1 \in \mathcal{C})$, $G_y^Y = (G_y^Y(c_0, c_1) : c_0, c_1 \in \mathcal{C})$, and $G_{(x,y)}^{(X,Y)} = (G_{(x,y)}^{(X,Y)}(c_0, c_1) : c_0, c_1 \in \mathcal{C})$ be $|\mathcal{C}|$ x $|\mathcal{C}|$ matrices with entries given by

$$\begin{aligned} G_x^X(c_0, c_1) &= R(c_0, c_1) \sum_y q(x, y|c_0, c_1), \\ G_y^Y(c_0, c_1) &= R(c_0, c_1) \sum_x q(x, y|c_o, c_1), \\ G_{(x,y)}^{(X,Y)}(c_0, c_1) &= R(c_0, c_1) q(x, y|c_0, c_1). \end{aligned}$$

Then $H(X) = -\lambda(X)$, $H(Y) = -\lambda(Y)$, and $H(X,Y) = -\lambda(X,Y)$, where $\lambda(X)$, $\lambda(Y)$, and $\lambda(X,Y)$ are the Lyapunov exponents defined as the following a.s. limits:

$$\begin{aligned} \lambda(X) &= \lim_{n\to\infty} \frac{1}{n} \log ||G_{X_1}^X G_{X_2}^X \cdots G_{X_n}^X|| \text{ a.s.}, \\ \lambda(Y) &= \lim_{n\to\infty} \frac{1}{n} \log ||G_{Y_1}^Y G_{Y_2}^Y \cdots G_{Y_n}^Y|| \text{ a.s.}, \\ \lambda(X,Y) &= \lim_{n\to\infty} \frac{1}{n} \log ||G_{(X_1,Y_1)}^{(X,Y)} G_{(X_2,Y_2)}^{(X,Y)} \cdots G_{(X_n,Y_n)}^{(X,Y)}|| \text{ a.s.} \end{aligned}$$

From this point onward, we will focus our attention on the Lyapunov exponent $\lambda(X)$, since the conclusions for $\lambda(Y)$ and $\lambda(X,Y)$ are analogous. Define

$$\tilde{p}_n \triangleq \frac{wG_{X_1}^X \cdots G_{X_n}^X}{||wG_{X_1}^X \cdots G_{X_n}^X||} = \frac{\tilde{p}_{n-1} G_{X_n}^X}{||\tilde{p}_{n-1} G_{X_n}^X||}$$

Proposition 2: Let w be the stationary distribution of the channel C. Then, for $n \geq 0$ and $c \in \mathcal{C}$,

$$\tilde{p}_n(c) = P(C_{n+1} = c|X_1^n),$$

That is, $\tilde{p}_n$ is the standard prediction filter from the theory of hidden Markov models (HMM).

Proposition 3 The sequence $\tilde{p} = (\tilde{p}_n : n \geq 0)$ is a Markov chain taking values in the continuous state space $\mathcal{P} = \{w : w \geq 0, ||w||_1 = 1\}$. Furthermore,

$$||\tilde{p}_n G_x^X||_1 = P(X_{n+1} = x|X_1^n).$$

Proposition 4: In [1] we present conditions under which $\tilde{p}_n$ posseses a unique stationary distribution π. When those conditions hold we can state

$$H(X) = -\lambda(X) = -\sum_{x\in\mathcal{X}} \int_{\mathcal{P}} \log(||wG_x^X||_1) ||wG_x^X||_1 \pi(dw).$$

Proposition 5: In [1] we present conditions under which $\lambda(X)$ is a continuous function of the input symbol distribution and the transition probabilities for the channel. Hence we can write capacity in terms of Lyapunov exponents

$$C = \max_{p(X)} [\lambda(X) + \lambda(Y) - \lambda(X,Y)]$$

The above propositions allow us to directly compute mutual information and capacity for a new class of Markov channels with non-i.i.d. inputs. This is a significant advance over previous results that required asymptotics or simulation. Detailed proofs and computational examples are available in [1].

References

[1] T. Holliday, P. Glynn, A. Goldsmith "On Entropy and Lyapunov Exponents for Finite State Channels", *Submitted to IEEE Trans. on Information Theory.*

0-7803-7728-1/03/$17.00 ©2003 IEEE.