Entropy of Multidimensional Shifts and Stochastic Processes

Nigel Burke
Summer 2012 NSERC USRA Report

September 4, 2012

This summer I worked with Prof. Brian Marcus, Prof. Andrew Rechnitzer and John Enns on methods to calculate and approximate the entropy of multidimensional shift spaces and stochastic processes. The entropy of one-dimensional shifts and stochastic processes has been extensively studied and many results have been obtained for analysis of the entropy of these systems [1]. In higher dimensions, however, analogous results are rare, and expressions for the entropy of some seemingly simple examples are not known. Our main focus was on two-dimensional shifts, and in particular on the Golden Mean shift in two-dimensions, but we were able to prove some results for general shifts satisfying certain broad restrictions. We also studied entropy in the context of stochastic processes and obtained an extension of a result for one-dimensional processes to two-dimensional processes.

We first studied entropy in the context of *stochastic processes*, which are indexed sequences of discrete random variables $\{X_i : 1 \leq i \leq n\}$. The *entropy* of a stochastic process, denoted $H(\{X_i\})$, is defined as:

$$H(\{X_i\}) = -\sum_{\{x_i\} \in \mathcal{X}^n} p(x_1, x_2, \ldots, x_n) \log p(x_1, x_2, \ldots, x_n) \tag{1}$$

where $\mathcal{X}$ is the alphabet of possible states and $p(x_1, x_2, \ldots, x_n)$ is the probability mass function for the states of the n variables $x_i \in \mathcal{X}$. The *entropy rate* of a stochastic process is given by two equivalent definitions:

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) = \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2}, \ldots, X_1) \tag{2}$$

The equivalence of the first definition to the second, which uses the entropy of a discrete random variable X_n conditioned on the previous random variables in the sequence of the stochastic process is proved in [2]. We proved an equivalent conditional entropy definition for the entropy rate of a two-dimensional stochastic process, which is a set of discrete random variables with two indices. We showed that the limit of the sequence of conditional entropies with the discrete random variables in the plane conditioned on the variables in an L-shaped set defined by three integers is equal to the two-dimensional entropy rate:

$$H(\mathcal{X}) = \lim_{n,k\to\infty} \frac{1}{nk} H(X_{1,1}, X_{1,2}, \ldots, X_{n,k}) = \lim_{n_1,n_2,n_3\to\infty} H(X_{n_1,n_2}|L_{n_1,n_2,n_3}) \tag{3}$$

where the set L_{n_1,n_2,n_3} is defined as

$$L_{n_1,n_2,n_3} = \{(x,y) \in \mathbb{N}^2 : (x = n_1 \text{ and } y < n_2) \text{ or } (x < n_1 \text{ and } y \leq n_2 + n_3)\} \tag{4}$$

Next we focused on the entropy of shift spaces. A *shift space* or *shift* in one-dimension is defined as the set of all bi-infinite sequences of symbols chosen from some finite set of symbols called an *alphabet*. Without any other restrictions, the shift is called the *full shift* on the alphabet. Other shifts can be defined as subsets of the full shift by specifying finite sequences of symbols from the alphabet and requiring that these sequences not appear in the bi-infinite sequences of a shift as subsequences of consecutive symbols. These blocks of symbols are called *forbidden blocks*. The entropy of a shift X is defined as:

$$h(X) = \lim_{n\to\infty} \left(\frac{\log N_n}{n} \right) \tag{5}$$

where N_n is the number of length n sequences of symbols with no forbidden blocks as subsequences. It is possible to construct shifts in two-dimensions from one-dimensional shifts by requiring that every row and column of symbols arranged on an infinite plane contain no blocks forbidden by the one-dimensional shift. For a shift X, we denote the two-dimensional shift constructed in this way as $X \otimes X$.

We initially focused on the one-dimensional Golden Mean shift, which is a shift on the alphabet $\{0,1\}$ defined by forbidding the block 11. This shift has entropy equal to $\log\left(\frac{1+\sqrt{5}}{2}\right)$ [1]. The entropy of the shift in two-dimensions obtained by requiring that every row and column belong to the Golden Mean shift is not exactly known, however several estimates and bounds have been obtained [3] [4] [5]. We obtained bounds on this entropy by analyzing the entropy of a one-dimensional shift with the extra condition that the sequence could be coiled up to form a valid configuration on a helix of circumference l. This shift has the advantage that the memory and processing power required to calculate its entropy increases more slowly with l than other methods of estimation. This advantage is nullified, however, by the fact that the upper and lower bounds obtained with this method are necessarily much further apart than bounds obtained through other methods, and so do not provide increased precision in estimating the two-dimensional Golden Mean shift entropy.

Finally, we examined shifts where the entropy of the one-dimensional shift is equal to the entropy of the two-dimensional shift constructed by requiring that bi-infinite sequences along rows and columns are elements of the one-dimensional shift. That is, shifts X such that $h(X) = h(X \otimes X)$. Full shifts on any alphabet have this property, however there also exist shifts with this property that forbid some blocks over their alphabet. The example of a shift like this that motivated our work was that of the odd shift, which is defined as

$$\text{ODD} = \{x \in \{0,1\}^{\mathbb{Z}} : \text{ there are an odd number of 0s between any consecutive 1s in } x\}$$

We proved necessary and sufficient conditions for a certain class of shifts to have this property. In particular, our result applies to irreducible sofic shifts. A shift is *irreducible* if for any allowed blocks u and v over the alphabet of the shift, there exists another block w so that the concatenation of blocks uwv contains no subsequence of forbidden words for the shift. A shift is *sofic* if it can be represented as the set of all labels of the edges of infinite paths through a labeled graph G. In this case, we say that the graph G *presents* the sofic shift. Many graphs may present a sofic shift, but only some presentations are of interest. A presentation is *right resolving* if no two edges joining the same pair of vertices in the graph have the same label. For a sofic shift X, the graph with the fewest vertices that presents X is called the *minimal* presentation of X. It is proved in [1] that for all irreducible sofic shifts there exists a minimal right-resolving presentation of the shift. Hence, we define the *period* of a sofic shift as the greatest common divisor of the lengths of cycles in the minimal right-resolving presentation of a shift. We proved the following theorem:

Theorem. *Let X be an irreducible sofic shift with period p. Then $h(X) = h(X \otimes X)$ if and only if the minimal right-resolving presentation of X contains p states.*

I am extremely grateful to Prof. Marcus and Prof. Rechnitzer for their help and encouragement over the course of this project and for giving me this opportunity. I am also grateful to John Enns for his collaboration and assistance and to the UBC Department of Mathematics and NSERC for their support.

References

[1] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding.* New York: Cambridge University Press, 1995.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Hoboken: John Wiley & Sons, Inc., 2006.

[3] N. J. Calkin and H. S. Wilf, "The Number of Independent Sets in a Grid Graph," *SIAM Journal on Discrete Mathematics*, vol. 11, pp. 54-60, February 1998.

[4] S. Friedland, P. H. Lundlow, and K. Markström, "The 1-vertex transfer matrix and accurate estimation of channel capacity," *IEEE Transactions on Information Theory*, vol. 56, pp. 3692-3699, August 2010.

[5] Z. Nagy and K. Zeger, "Capacity Bounds for the 3-Dimensional (0, 1) Runlength Limited Channel," *IEEE Transactions on Information Theory*, vol. 46, pp. 1030-1033, 2000.