Estimating the Entropy Rate of Spike Trains

Yun Gao
Div. of Applied Math.
Brown University
Providence, RI 02912
gao@dam.brown.edu

Ioannis Kontoyiannis
Div. of Applied Math. & Dept. of Computer Science
Brown University
Providence, RI 02912
yiannis@dam.brown.edu

Elie Bienenstock
Div. of Applied Math. & Dept. of Neuroscience
Brown University, Providence, RI 02912
elie@dam.brown.edu

1 Introduction

Information-theoretic methods have been widely used in neuroscience, in the broad effort to analyze and understand the fundamental information-processing tasks performed by the brain. In these studies, the entropy has been adopted as the main measure for quantifying the amount of information transmitted between neurons, via the spike trains they generate. One of the first and most important goals is to identify appropriate methods that can be used to quantify the amount of information that gets communicated by spike trains, or, in other words, to estimate the entropy of spike trains recorded from live animals.

So far, the most commonly used entropy-estimation technique has been the so-called "plug-in" (or maximum-likelihood) estimator and its various modifications. This method consists of essentially calculating the empirical frequencies of all words of a fixed length in the data, and then estimating the "true" entropy of the underlying signal as the entropy of this empirical distribution; see, e.g., [10][5][12][6][9]. For computational reasons, the plug-in estimator cannot go beyond word lengths of about 10 or 20, and hence it does not take into account the potential longer time dependencies in the signal.

Here we examine the performance of entropy estimators based on two data compression algorithms, the Lempel-Ziv algorithm (LZ) and the Context Tree Weighting method (CTW). Specifically, we consider two LZ-based entropy estimators and one based on the CTW. The first LZ-based method has been widely and very successfully used in many applications, and the other one is a new estimator with some novel and more desirable statistical properties. The CTW-based estimator is based in the work of Willems *et al* [13][14][15] and it has also been considered in [1][3].

2 Results

We demonstrate that the LZ- and CTW-based estimators naturally incorporate dependencies in the data at much larger time scales than the plug-in, and that they are consistent (in the statistical sense) for a wide class of data types generated from distributions that may posses arbitrarily long memory.

The Lempel-Ziv algorithm [17][18] is a universal data compression scheme that achieves the (optimal) entropy lower bound when applied to data generated by *any* stationary ergodic process. As the conditions of stationarity and ergodicity are very weak (and in some sense minimal), they appear well-suited for neural data, as we have no *a priori* bound on the length of the memory in the data, and in fact the very length of this memory is one of the objects we intend to study.

The main gist in the workings of the Lempel-Ziv algorithm was revealed by Wyner and Ziv in [16], where they studied the connection between the entropy of a process and the longest match-lengths along a process realization. Roughly speaking, the match-lengths measure the length of the longest string starting in a fixed position in the data which re-appears in a given window somewhere else in the same data. Intuitively, the longer the match-lengths, the more regularity there is in the data, and hence the smaller the entropy (and the more efficient the compression). Partly motivated by this connection, a number of entropy estimators have been proposed since then and have been applied to many different kinds of data; for examples see [8][2] and the references therein. Here we use two entropy estimators based on match-lengths, one described in [2], and a new one. We study their theoretical properties, we apply them to neuronal data, and we present a systematic simulation study of their statistical properties.

The CTW [13][14][15] is another universal compression algorithm for tree sources, which has the additional advantage that it also gives as its output a statistical distribution for the data it compresses. Like the Lempel-Ziv algorithm, the CTW also achieves the entropy lower bound, and in special cases it is shown to achieve the best possible redundancy rate as determined in [7]. In particular, its redundancy can be bounded above *uniformly* over all data sequences of arbitrary length, which means that the same thing can be said for the bias of the resulting entropy estimation algorithm. This is clearly very valuable information to have when this algorithm is used in practice. Here we study an entropy estimator based on the CTW, implemented in a manner similar to that described in [3].

To compare the performance of various methods, we apply these entropy estimators on simulated data, generated from homogeneous Poisson processes, Markov chains of various orders, hidden Markov models (HMMs), and renewal and Markov-renewal processes. For most of these models, the true entropy rate can be calculated in closed form. Our analysis shows that, whereas for short-memory processes the plug-in is as good as any other method, for processes with longer memory the plug-in is much worse than the both the LZ estimators and the CTW, because of undersampling problem. In fact, the CTW estimator is uniformly better than the other estimators, for both short and relatively long memory processes. Its fast convergence rate outperforms the LZ-based estimators, and its ability to allow for longer memory makes it more accurate than the plug-in.

3 Experimental Results

We next apply these methods to neural data. Our data come from two multi-electrode arrays implanted on a monkey's primary motor cortex (MI) and dorsal premotor cortex (PMd). The arrays simultaneously recorded neural activity from 29 different neurons. A Plexon acquisition system was used to collect the neural signal, and the units were spike-sorted using Plexon's Offline Sorter. The monkey was not engaged in any task when the data were collected, and the size of the data is approximately an hour. A detailed description of recording techniques is given in [4].

Our results on neural data show that the CTW gives somewhat lower estimates than the plug-in, despite the fact that the the bias of the plug-in estimator is negative whereas that of the CTW is positive. This suggests that the CTW estimates are more reliable, and it strongly indicates that the CTW's smaller values come from the fact that it *does indeed find longer-term dependencies in the data*. Using the two LZ-based estimators we find that one gives results systematically higher and one systematically lower that those of the plug-in.

As mentioned above, from the CTW algorithm we can also obtain an explicit statistical model for the data. In this study we looked extensively at the resulting "maximum a posteriori probability tree" models, as described in [11], which give the best (in a certain sense) tree models that can be fit to the data at hand. From the results we clearly see that the spike-train data exhibit long-range dependencies (much longer that the 10- or 20-millisecond window captured by most earlier studies). We also find that perhaps the most relevant modeling "parameter" for estimating the entropy of these spike trains is the inter-spike-interval (ISI) distribution. This offers another possible explanation for why the plug-in method and its variants may not produce satisfactory results. Furthermore, a detailed analysis of how the entropy estimates behave as the tree depth allowed in the CTW varies, suggests that it is natural to think that spike trains have, generally speaking, slowly varying firing rates, and that the firing rate is strongly related with the Fano factor, which describes the variability of the spike trains.

4 Conclusions

Overall, we find that the CTW is a significantly better estimator that either the plug-in-based or the LZ-based methods, and also that it is a more appropriate one for neuronal data. Its convergence rate is fast, and it exhibits a strong ability to model long-memory statistical properties in the data. Moreover, it offers an actual probabilistic model for the data, which can be used to read off important statistical properties of spike trains that go well beyond the entropy estimation task.

References

[1] M. Kennel and A. Mees Context tree modeling of observed symbolic dynamics. *Phys. Rev. E*, 66:056209, 2002

[2] I. Kontoyiannis, P.H. Algoet, Yu.M. Suhov, and A.J. Wyner. Nonparametric entropy estimation for stationary processes and random fields, with applications to English text. *IEEE Trans. Inform. Theory*, 44:1319–1327, 1998

[3] M. London The information efficacy of a synapse. *Nature Neurosci.*, 5(4)332-340:, 2002

[4] E. Maynard, N. Hatsopoulos, C. Ojakangas, B. Acuna, J. Sanes, R. Normann, and J. Donoghue. Neuronal interaction improve cortical population coding of movement direction. *J. of Neuroscience*, 19(18):8083–8093, 1999

[5] L. Paninski. Estimation of entropy and mutual information. *Neural Comp.*, 15: 1191–1253, 2003.

[6] P. Reinagel. Infomation theory in the brain. *Current Biology*, 10(15):R542-R544, 2000

[7] J. Rissanen. Universal Coding, information, prediction, and estimation. *IEEE Inform. Theory*, 30:629-636, 1984

[8] T. Schürmann and P.Grassberger. Entropy estimation of symbol sequences. *Chaos*, 6:414–427, 1996

[9] C.F. Stevens and A.Zador. Information through a spiking neuron *NIPS*, 8, 1996

[10] S.P. Strong, R. Koberle, R. de Ruyter van Steveninck, and W. Bialek. Entropy and information in neural spike trains. *Physical Review Letters*, 80:197–200, 1998

[11] P.A.J. Volf and F.M.J. Willems. A Study of the Context Tree maximizing method. *Proceedings of the 1995 IEEE International Symposium on Information Theory*,1995

[12] D.K. Warland, P. Reinagel and M. Meister. Decoding visual information from a population of retinal ganglion cells *J. of Neurophysiology*, 78(5):2336-2350, 1997

[13] F.M.J. Willems, Y.M. Shtarkov and T.J. Tjalkens. The Context-tree weighting method: basis properties *IEEE Trans. Inform. Theory*, 41:653-664, 1995

[14] F.M.J. Willems, Y.M. Shtarkov and T.J. Tjalkens. Context weighting for general finite-context sources *IEEE Trans. Inform. Thoery*, 42:1514-1520, 1996

[15] F.M.J. Willems. The Context-tree weighting method: extensions *IEEE Trans. Inform. Thoery*, 44:792-798, 1998

[16] A.D. Wyner and J. Ziv. Some asymptotic properties of entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inform. Theory*, 35:1250–1258, 1989

[17] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Inform. Theory*, 23:337–343, 1977

[18] J. Ziv and A. Lempel. Compression of individual sequences via variable rate coding. *IEEE Trans. Inform. Theory*, 24:530–536, 1978