Differential Entropy of Multivariate Neural Spike Trains

Nanyi Cui, Jiaying Tang, and Simon R. Schultz

Department of Bioengineering, Imperial College London, Exhibition Road
South Kensington Campus, London, SW7 2AZ, United Kingdom
{nanyi.cui07,j.tang10,s.schultz}@imperial.co.uk

Abstract. Most approaches to analysing the spatiotemporal dynamics of neural populations involve binning spike trains. This is likely to underestimate the information carried by spike timing codes, in practice, if they involve high precision patterns of inter-spike intervals (ISIs). In this paper we set out to investigate the differential entropy of multivariate neural spike trains, following the work of Victor. In our framework, the unidimensional special case corresponds to estimating the differential entropy of the ISI distribution; this is generalised to multidimensional cases including patterns across successive ISIs and across cells. We investigated the differential entropy of simulated spike trains with increasing dimensionality, and applied our approach to electrophysiological data recorded from the mouse lateral geniculate nucleus.

Keywords: Differential entropy, temporal analysis, inter-spike interval, poisson process simulation.

1 Introduction

Neurons convey sensory information using rate [17], correlational [1, 8, 13] and temporal [9] codes. A quantitative analysis of the neural code should be able to capture all of these within a single framework [10]. Most information theoretic analyses of neurophysiological data attempting to do this make use of a digitised representation of a neural spike train, by binning the spike train into segments containing 0 or 1 spikes. If the bin width is sufficiently small, this provides a complete representation of the information content of the spike train - subject to the proviso that it must be possible to adequately sample the distribution of spike trains. In practice, this limitation frequently prevents the examination of codes involving precise spike timing.

If patterns (across time and across neural ensembles) of interspike intervals (ISIs) are used to represent information, a binned strategy for neural data analysis may, due to limited sampling, be insufficiently sensitive to detect information bearing structure in the data. An alternative approach for studying multivariate neural spike trains using continuous information theory is thus of great interest. Victor [16] presented an approach for binless estimation of the information carried by a neural spike train, showing that in some circumstances bin less approaches can be significantly more efficient than binned approaches such as the

A.E.P. Villa et al. (Eds.): ICANN 2012, Part I, LNCS 7552, pp. 280–287, 2012.

direct method [15]. In the current work, we adopt the approach of Victor, examining how well it can be used to estimate the differential entropy of a sequence of interspike intervals, beginning with a one-dimensional case (the differential entropy of the ISI distribution), and extending our analysis to patterns of adjacent ISIs across time and across cells.

Our goal is to estimate the differential entropy of an ensemble of neural spike trains. We consider neural ensemble activity at a particular instant in time to be characterised by a point in R^m space, or equivalently an m-dimensional vector. In terms of analysing spike train data we can imagine moving the current instant of time throughout the spike train (see Fig. 1), updating the probability density of the ensemble ISI pattern each time a spike is encountered. Denoting t_i as the most recent ISI for cell i, and the collection of most recent ISIs as t_i, the joint ISI probability density is $p(x) = p(t_i)$. More generally, the joint density of ISI sequences going back k steps is $p(x) = p(t_i^{(j)})$ where $j = 1..k$. The total dimensionality for C cells is $m = kC$.

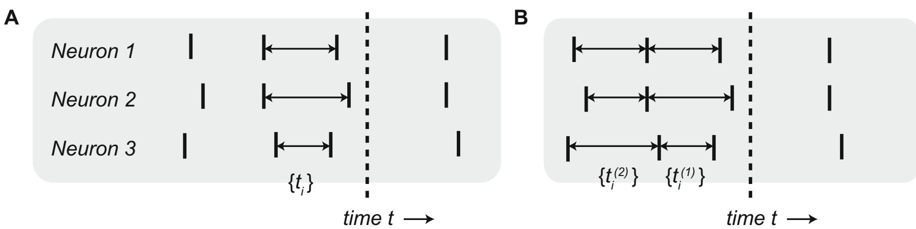

Fig. 1. Interspike interval codes depicted with rastergrams. The ISI vector is updated whenever a spike is fired by any neuron. **A** $k = 1$ defines a joint ISI distribution code. **B** For $k = 2$, a multivariate ISI pattern code is produced, which looks back at the previous two interspike intervals for each cell.

We note that with methods for computing differential entropy established, it will be possible to compute further information theoretic quantities such as mutual information, the Kullback-Leibler divergence, and transfer entropy.

2 Methods

Entropy in Information Theory is a measure of probability distribution associated with a random variable, in this case the ISI which is a continuous random variable. Shannon defined the differential entropy,

$$H_{\text{diff}} = -\int_S p(x) ln p(x) dx \tag{1}$$

where S is the support set of the random variable x.

Although the ISI is a continuous variable, in practice we measure the ISI in discrete time values, so by making the transition from estimating differential entropy of discrete variable to estimating entropy of a continuous one, we could adopt Victor's binless method that employs the mathematical derivation by Kozachenko and Leonenko [6]. They formulated conditions for asymptotic unbiasedness of a random continuous vector built on the basis of independent observations. The continuity assumption for $p(x)$ means that within a sufficiently small neighbourhood of x_i, $p(x_i)$ is approximately a locally uniform distribution. The Euclidean Distance between any two points of a random variable x,

$$\rho(x_1, x_2) = \left\{ \sum_{j=1}^{m} \left(x_1^{(j)} - x_2^{(j)}\right)^2 \right\}^{1/2}, \qquad (2)$$

So the minimum Euclidean Distance of a point x_i is

$$\rho_i = \min \{\rho(x_i, x_j), j \in \{1, 2, ..., N\} \mid \{i\}\} \qquad (3)$$

Now with this Euclidean Distance, Victor derived the differential entropy

$$H_{\text{diff}} \approx \frac{m}{N \sum_{j=1}^{N} \log_2 \rho_j} + \log_2(\frac{c(m)(N-1)}{m}) + \frac{\gamma}{\ln 2} \qquad \text{(bits)} \qquad (4)$$

where m is the dimensionality, N is the total number of spikes, ρ is the nearest neighbour distance, $c(m)$ is a constant that depends on m, and γ is the Euler-Mascheroni constant. We numerically implemented this estimator , which was used for the results in the rest of this paper, except where otherwise described.

To test the approach, it was useful to compare against a known theoretical distribution. For a normal distribution, we have for the differential entropy

$$H_{\text{Gauss}} = \frac{1}{2} \ln(2\pi e \sigma^2) \qquad \text{(nats)} \qquad (5)$$

and for the multivariate Gaussian

$$p(x_1, ..., x_m) = \frac{1}{(2\pi)^{m/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \qquad (6)$$

$$H_{\text{Gauss}} = \frac{1}{2} \ln\left((2\pi e)^m \left|\Sigma\right|\right) \qquad \text{(nats)} \qquad (7)$$

where Σ is the covariance matrix and $|\Sigma|$ is the determinant of Σ. Note the linear relationship between differential entropy and dimensionality m.

We compared the binless estimator with the plugin estimator of the binned distribution of interspike intervals, parameterising the latter case by bin width. Figure 2 below shows the comparison between the binless and binned method. Both methods were tested on pseudo-randomly generated Gaussian distributions with sample sizes (of ISIs) from 32 to 4096. Error bars were computed by bootstrapping. The binless estimator can be seen to converge more rapidly to the true differential entropy (established analytically as described above).

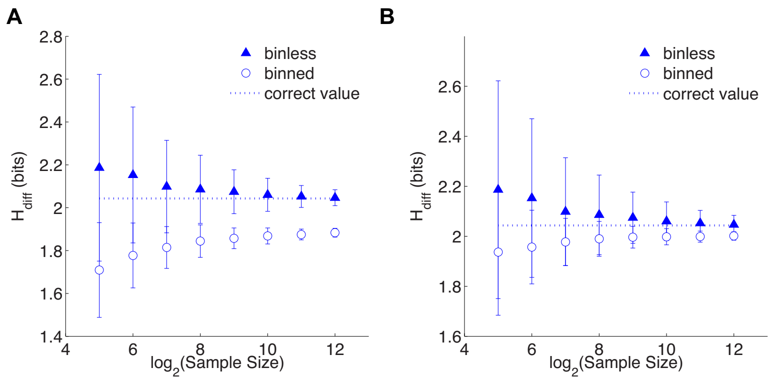

Fig. 2. Comparison between binned and binless entropy estimation for pseudo-randomly generated Gaussian distributions with unit variance, for cases with **A** substantial, and **B** smaller, quantisation loss. The true differential entropy is calculated analytically. Error bars indicate standard deviation over 200 bootstrapped samples.

3 Results

Spiking intervals may be more closely modelled by a Poisson process than a gaussian distribution [2], with the time intervals of a Poisson process of rate λ being exponentially distributed with a mean of $1/\lambda$. We simulated some Poisson processes with a range of rates, to examine whether the binless approach would work as well for exponential distributions as it did for the normal distribution. The differential entropy of an exponential distribution is

$$H_{\text{Exp}} = \int_0^\infty \lambda e^{-\lambda x} \log_2(\lambda e^{-\lambda x}) dx \tag{8}$$

thus

$$H_{\text{Exp}} = -\log_2 \lambda + \log_2 e \qquad \text{(bits)} \tag{9}$$

and for the m-dimensional multivariate exponential distribution the probability density is [4]

$$p(x_1, ..., x_m) = \prod_{i=1}^{m} \frac{1/\lambda + i - 1}{\theta_i} e^{\frac{x_i - k_i}{\theta_i}} \left[\sum_{i=j}^{m} e^{\frac{x_j - k_j}{\theta_j}} - m + 1 \right]^{-(1/\lambda + m)} \tag{10}$$

where λ is the rate parameter of the exponential distribution, θ and $\mathbf{k}$ are m-dimensional parameters, such that $x_i > k_i$ and $\theta > 0$. The entropy is

$$H_{\text{Exp}} = -\sum_{i=1}^{m} \ln(\frac{1/\lambda + i - 1}{\theta_i}) + (1/\lambda + m) \sum_{i=1}^{m} \frac{1}{1/\lambda + i - 1} - \frac{m}{1/\lambda} \qquad \text{(nats)} \tag{11}$$

Note that this shows that, depending on the circumstance, differential entropy can scale negatively with dimensionality m. An example is shown in Fig. 3. As the dimensionality increases, the RMS error of the bin less entropy estimate also increases, for the same sample size (see Fig. 4).

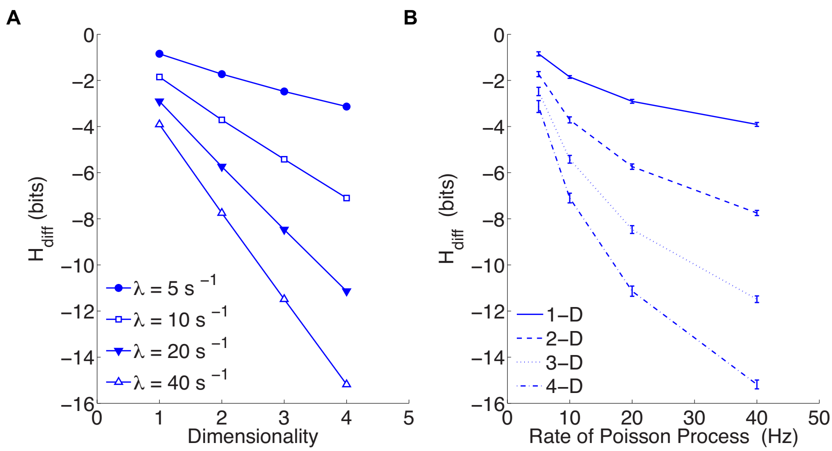

Fig. 3. A. Differential entropy estimates of Poisson process simulations as dimension increases. **B** Differential entropy estimates as the rate of the Poisson process increases.

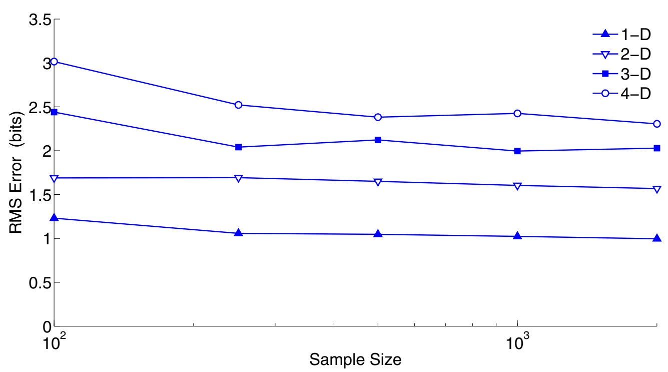

Fig. 4. Root mean squared error of binless estimates for 1-4 dimensional Poisson simulations with $\lambda = 5$ spikes per second

How does differential entropy depend upon the variability of the spike trains? We simulated spike trains with a range of variability, by (i) starting with a regular (periodic) process and adding gaussian jitter, to obtain coefficients of variation (CVs) of the interpsike intervals of between 0 and 1. We also added jitter to a Poisson simulation, in order to obtain CVs of greater than one. Fig. 5

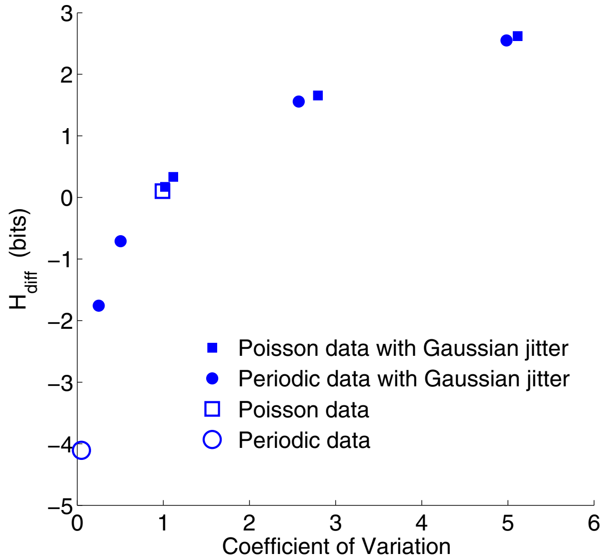

Fig. 5. Differential entropy grows with ISI variability, as described by the Coefficient of Variation, for one-dimensional (ISI distribution) entropy analysis. Simulated processes with a rate of 5 spikes/sec.

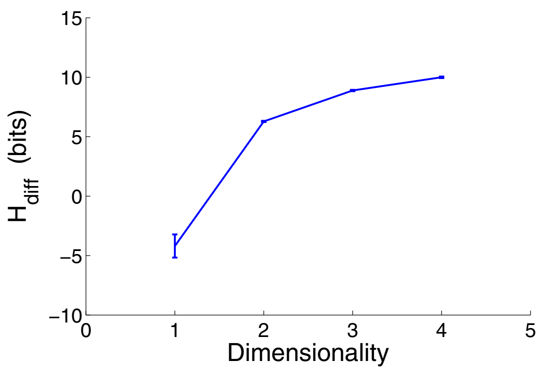

Fig. 6. Multivariate differential entropy estimation across multiple cells. Error bars indicate standard error of the mean across all selected ensembles.

shows the sub-linear relationship between differential entropy and CV that we observed with these processes. The differential entropy is negative for CV < 1, passing near zero for a Poisson process (CV=1) with rate $\lambda = 5$ spikes/sec, and becoming more positive with increasing dimensionality.

We next considered estimation of the differential entropy of spike trains from multiple cells, to demonstrate the applicability of our approach to real, multidimensional data. We examined an example of a real neurophysiological recording from the lateral geniculate nucleus of the anaesthetised mouse. Spontaneous activity was simultaneously recorded from 24 cells using a silicon micromachined electrode array (Neuronexus Technologies). We analysed the differential entropy of all single units, and of randomly selected sets of pairs, triplets and quadruplets (but analysing only a single ISI back in time, so that dimensionality is equal to the number of cells). The results are shown in Fig. 6.

4 Discussion

The binless entropy estimate is based on the Euclidean distance to the nearest neighbour. Analogous estimators can be constructed based on kth-nearest-neighbor distances, which may allow longer range temporal structure to be incorporated in a feasible way. Although the exact result of Kozachenko and Leonenko was not demonstrated beyond dimension 1, in this study we explored the validity of their mathematical model in higher dimensions, following the work of Darbellay et al [4], and applying it to real multi-electrode array neurophysiological data.

One advantage of the binless method is that it preserves the topology of spike trains that are continuous variables due to the nature of time. Another advantages of the binless method is fast convergence to the correct value as sample size increases shown in entropy estimation. We envisage that the binless estimators may prove to be a useful tool for the analysis of datasets that include both aspects of temporal precision and extended range temporal structure. They may also provide a useful alternative to mean field expansion approaches to scaling up the analysis of spatiotemporally structured neural data [12].

References

1. Averbeck, B.B., Lee, D.: Coding and transmission of information by neural ensembles. Trends in Neurosciences 27(4), 225–230 (2004)
2. Correia, M.J., Landolt, J.P.: A point process analysis of the spontaneous activity of anterior semicicular canal units in the anesthetized pigeon. Biol. Cybern. 27(4), 199–213 (1977)
3. Cover, T.M., Thomas, J.A.: Elements of information theory. Wiley, New York (1991)
4. Darbellay, G.A., Vajda, I.: Entropy expressions for multivariate continuous distributions. IEEE Transactions on Information Theory 46(2), 709–712 (2000)
5. Kandel, E.R., Schwartz, J.H., Jessell, T.M.: Principles of neural science, 4th edn. McGraw-Hill, Health Professions Division, New York (2000)
6. Kozachenko, I.F., Leonenko, N.N.: Sample estimate of the entropy of a random vector.. Probl. Inf. Transm. 23(1-2), 95–101 (1987) (English Russian original)
7. Ma, S.-K.: Calculation of entropy from data of motion. Journal of Statistical Physics 26(2), 221–240 (1981)

8. Montani, F., Kohn, A., Smith, M.A., Schultz, S.R.: The Role of Correlations in Direction and Contrast Coding in the Primary Visual Cortex. J. Neurosci. 27(9), 2338–2348 (2007)
9. Panzeri, S., Petersen, R.S., Schultz, S.R., Lebedev, M., Diamond, M.E.: The role of spike timing in the coding of stimulus location in rat somatosensory cortex. Neuron 29(3), 769–777 (2001)
10. Panzeri, S., Schultz, S.R.: A unified approach to the study of temporal, correlational, and rate coding. Neural Comput. 13(6), 1311–1349 (2001)
11. Rodieck, R.W., Kiang, N.Y., Gerstein, G.L.: Some quantitative methods for the study of spontaneous activity of single neurons. Biophys. J. 2, 351–368 (1962)
12. Schaub, M.T., Schultz, S.R.: The Ising decoder: reading out the activity of large neural ensembles. Journal of Computational Neuroscience 32(1), 101–118 (2012)
13. Schultz, S.R., Kitamura, K., Post-Uiterweer, A., Krupic, J., Hausser, M.: Spatial Pattern Coding of Sensory Information by Climbing Fiber-Evoked Calcium Signals in Networks of Neighboring Cerebellar Purkinje Cells. J. Neurosci. 29(25), 8005–8015 (2009)
14. Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal 27, 279–423, 623–656 (1948)
15. Strong, S.P., Koberle, R., de Ruyter van Steveninck, R.R., Bialek, W.: Entropy and information in neural spike trains. Physical Review Letters 80, 197–200 (1998)
16. Victor, J.D.: Binless strategies for estimation of information from neural data. Phys. Rev. E Stat. Nonlin. Soft. Matter Phys. 66(5), pt. 1, 051903 (2002)
17. Werner, G., Mountcastle, V.B.: Neural Activity in Mechanoreceptive Cutaneous Afferents: Stimulus-Response Relations, Weber Functions, and Information Transmission. J. Neurophysiol. 28(2), 359–397 (1965)