Estimation of excess entropy from spike trains

H. Motoyoshi*, Y. Katori**, and H. Suzuki*'**

** Graduate School of Information Science and Technology, The University of Tokyo,*
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
*** Institute of Industrial Science, The University of Tokyo,*
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

(motoyoshi@sat.t.u-tokyo.ac.jp)

Abstract: Entropy rate is widely used for analysis of neural data as a measure for randomness of spike trains. In addition, its convergent process also contains information on spike trains' structures or patterns. Therefore, it can be expected to be a measure for certain aspects of spike trains. In this paper, we investigate applicability of excess entropy to neural spike train data by numerical simulations of gamma process. We show that even when the spike train is not so long, the estimated excess entropy correctly reflects the shape parameter of the gamma process.

Keywords: Excess entropy, spike trains, gamma distribution, entropy rate

I. Introduction

In neural systems, spikes are thought to be a main carrier of information and are necessary to transmits e.g. sensory information and motor commands. Information representation and transmission in the neural systems are crucial issues in neuroscience. For example, in inferior olive (IO) system, prominently low firing rates of spikes of IO neurons are supposed to transmit error signals in motor control and their detailed information coding is still controversial. In order to quantify the information flow in neural system, statistical and information theoretical approaches are often used. One of the major approaches quantifying information flow is to measure the entropy rate. In addition to entropy rate, excess entropy was proposed by Clutchfield and Feldman [1]. Excess entropy captures a different element that is related with structure from entropy rate that expresses randomness. It can be a useful tool for analyzing spike trains. Excess entropy has been investigated by various researchers. Effective measure complexity proposed by Grassberger [2] is the same as excess entropy. Predictive information proposed by Bialek et al. [3] is similar to the excess entropy. However, estimating excess entropy has some problems. Excess entropy needs infinite length sample by definition. In practice, these measures need to be estimated from finite length sample. Considering fluctuation of sample data set, estimated value of the excess entropy should be corrected [4]. In the case of spike train data observed from neurons, distributions of inter-spike intervals (ISI) are known to be well approximated by the exponential distribution (Poisson process) or the gamma distribution [5]. Since the gamma distribution has two parameters: shape and scale parameters, ISI distributions of physiological data are better approximated by the gamma distribution than the exponential distribution. Here, we assume that ISI distribution follows the gamma distribution and we apply information theoretical measures to spike trains generated from gamma process.

In this paper, first, we review two information theoretic measures: entropy rate and excess entropy. Second, we review a correction method for estimated value from finite sample. Third, we apply information theoretic measures with the correction to spike trains generated from the gamma process. Finally, a brief conclusion is remarked.

II. Information theoretic measures

Estimation of excess entropy and entropy rate is based on Shannon entropy that intuitively represents randomness or uncertainty. Let X be a random variable taking value x in a finite set χ. Then Shannon entropy of the random variable X is defined by

$$H(X) = -\sum_{x \in \chi} p(x) \log_2 p(x) \,. \tag{1}$$

Entropy of a string of random variables $X^L = X_1 ... X_L$ is defined as

$$H(X^L) = -\sum_{x^L \in \chi^L} p(x^L) \log_2 p(x^L), \tag{2}$$

where $p(x^L) = p(x_1,...x_L)$ is joint probability. Then uncertainty or average amount of information per symbol is given by

$$h(L) = H(X^L)/L. \quad (3)$$

The entropy rate h is defined by $h(L)$ in the limit as L goes to infinity as follows:

$$h = \lim_{L\to\infty} h(L). \quad (4)$$

As the length L increases, $h(L)$ decreases and approaches to the entropy rate h (Fig. 1(b)). To quantify its convergence, the excess entropy E is defined by

$$E = \sum_{L=1}^{\infty} h(L) - h. \quad (5)$$

The system appears more random than it is when the length L is small. Excess entropy tells us how much additional information about the configuration is required to reveal the actual randomness h.

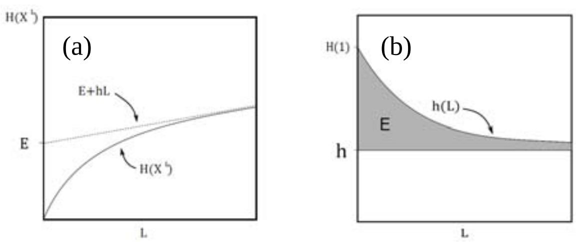

Fig.1. Image of excess entropy in relation with (a) entropy growth and (b) convergence of entropy rate.

Another definition of the excess entropy using mutual information allows us to understand meaning and nature of the excess entropy:

$$E = I(\overleftarrow{X}, \overrightarrow{X}) = H[\overrightarrow{X}] - H[\overrightarrow{X} \mid \overleftarrow{X}], \quad (6)$$

where $\overleftarrow{X} = ...X_{-3}X_{-2}X_{-1}$ and $\overrightarrow{X} = X_0 X_1 X_2$ are the left and the right part of infinite chain of random variables $\overleftrightarrow{X}$. This formulation suggests that excess entropy captures how much information on the half part of chain affects the other side. That is, excess entropy measures memory, predictability or correlation.

III. Estimation from finite length data set

To estimate the entropy rate h from a finite sequence of length N, we first need to estimate the probabilities $p(x_1,...x_L)$ from the finite sequence. A naive estimation of $p(x_1,...x_L)$ is given by

$$\hat{p}(x_1,...x_L) = \frac{n_{x_1,...x_L}}{N}, \quad (7)$$

where $n_{x_1,...x_L}$ is the number of occurrences of the word $x_1,...x_L$. By substituting Eq. (7) into Eq. (2), we obtain an estimated value of the entropy rate as follows:

$$\hat{h} = \lim_{L\to\infty} \frac{\hat{H}(X^L)}{L}. \quad (8)$$

However, this estimation tends to underestimates the true value. This can be understood by considering an expectation value of $\hat{H}(X^L)$ [4].

$$\left\langle \hat{H}(X^L) \right\rangle = \left\langle - \sum_{x^L \in \chi^L} \frac{n_{x_1,...,x_L}}{N} \log \frac{n_{x_1,...,x_L}}{N} \right\rangle$$
$$\le \left\langle - \sum_{x^L \in \chi^L} \frac{n_{x_1,...,x_L}}{N} \log p(x_1,...,x_L) \right\rangle$$
$$= - \sum_{x^L \in \chi^L} \frac{\left\langle n_{x_1,...,x_L} \right\rangle}{N} \log p(x_1,...,x_L) = H(X^L). \quad (9)$$

Figure 2(a) shows the result of estimation with finite length sample of sequence generated by the logistic map. It indicates that $\hat{H}(X^L) \approx \hat{h}L$ in the realm of good statistic and that $\hat{H}(X^L) \approx \log N$ in the realm of bad statistics [6]. Accuracy of the estimation depends not only on the fluctuation of the sample but also on the sample length.

A correction of the estimated value was given by Schürmann and Grassberger [4]:

$$\tilde{H}(X^L) = \sum_{j=1}^{M} \frac{n_j}{N} \left(\log N - \psi(n_j) - \frac{(-1)^{n_j}}{n_j + 1} \right), \quad (10)$$

where M is the number of distinct words of length L that occurs in the given sequence, n_j is the number of occurrence of the jth word, and $\psi(x)$ is the logarithmic derivative of gamma function.

We apply the correction described in Eq. (10) to estimate the entropy of the logistic map. Figure 2(b) indicates that the estimation of the entropy of the logistic map is significantly improved with this correction. However, the entropy rate should be evaluated within the realm of good statistics.

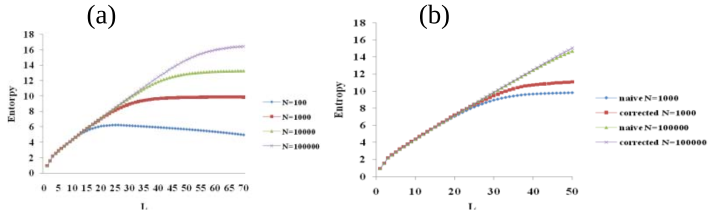

Fig. 2. Estimated entropy of the logistic map (a) without the correction and (b) with the correction.

IV. Application to Gamma process

In this section, we consider applicability of estimation of the excess entropy of spike trains with numerical simulations. Spike trains are often assumed to be generated from a point process. Here, we suppose that ISIs t follow the gamma distribution whose probability density function is given by

$$f(t;\kappa,\lambda)=\frac{\lambda^{\kappa}}{\Gamma(\kappa)}t^{\kappa-1}e^{-\lambda t}, \qquad (11)$$

where κ is a shape parameter and λ is a scale parameter. Figure 3 shows an example of probability density function of Gamma distribution.

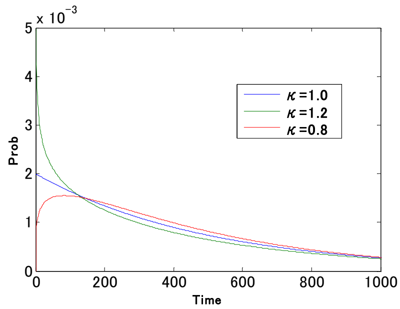

Fig. 3. Gamma distribution

The excess entropy can capture structural differences in ISI distribution from complete random process, like the Poisson process. This property of the excess entropy can be confirmed with the gamma distribution. The gamma distribution with $\kappa=1$ is equivalent to the Poisson distribution. As shown in Fig. 4, the difference of the shape parameter κ from 1 is considered to represent differences between the gamma distribution and the exponential distribution. Furthermore, the excess entropy E takes the minimum value $E=0$ when $\kappa=1$ (Poisson process). This can be understood by considering discrete time Markov process:

$$p(x_L \mid \{x_{1\le i\le L-1}\}) = p(x_L \mid x_{L-1}). \qquad (12)$$

The excess entropy of this process is written as

$$E=\sum_{\{x_i\}} p(\vec{x})\log\frac{p(x_L \mid x_{L-1})}{p(x_L)}. \qquad (13)$$

Since discrete-time Poisson process is a Bernoulli process, the excess entropy E takes zero for the gamma process with $\kappa=1$. If κ moves away from 1, spike trains have time correlations and the excess entropy increases (Fig. 4).

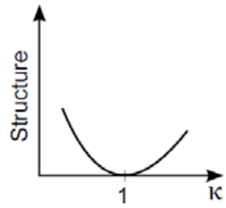

Fig. 4. Dependency of excess entropy on κ

Here, we apply information theoretic measures to spike trains generated from the gamma process. As in Fig. 5, spike train is discretized into bins and is analyzed as symbolic sequences of 0's and 1's.

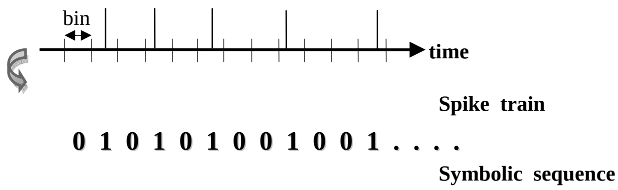

Fig. 5. Discretization of spike trains

For comparison with numerical estimation of excess entropy, we show information gain induced from Gamma distribution [7-8]. The information gain is Kullback-Leibler (KL) distance between gamma distribution and exponential distribution written as

$$\begin{aligned} G &= KL(f_g, f_e) \\ &= 1-\kappa-\log\Gamma(\kappa)+\log\kappa+(\kappa-1)\psi(\kappa)_{[bit/spike]} \\ &= \frac{\lambda}{\kappa}(1-\kappa-\log\Gamma(\kappa)+\log\kappa+(\kappa-1)\psi(\kappa)),_{[bit/time]} \end{aligned} \qquad (14)$$

where f_g is density function of gamma distribution and f_e is density function of exponential distribution. Note that we consider continuous distribution for easy calculation. This quantity also represents structural information stored in spike trains. The information gain G takes the minimum value (zero) at $\kappa=1$ (Fig. 6(a)) and this corresponds to characteristics of the excess entropy. The estimated excess entropy E is concave up around

$\kappa=1$ and has good agreement with characteristics of the information gain G. This corresponds to the above theoretical consideration and suggests that the excess entropy measures structural information (Fig. 6(b,c)).

Our method is applicable for analyses of practical neuronal data. In practical neuronal data, length of data is usually limited. As shown in section II, estimation of the entropy rate requires much long data; we hence consider estimation from experimentally practicable length of data. Figure 6(d) shows the result of estimation of the excess entropy with different length of data. As the data length gets shorter, the precision of estimation becomes worse, but its shape of curve is roughly preserved.

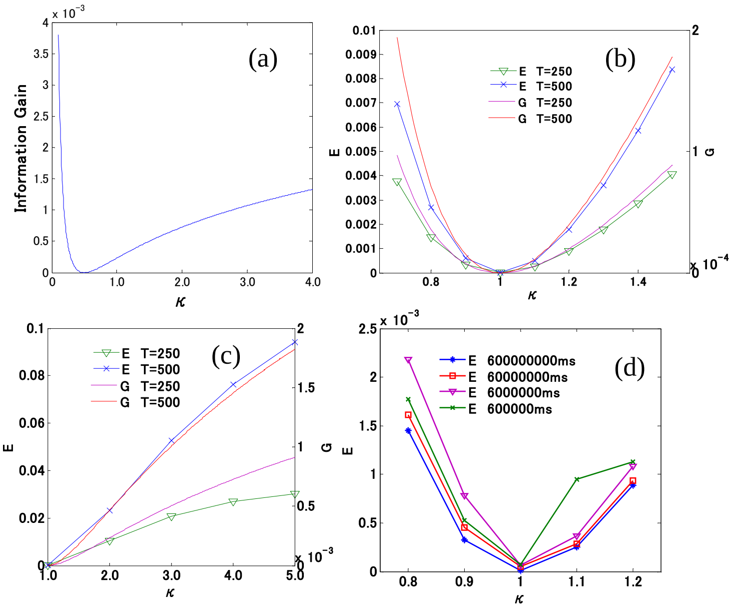

Fig.6 Estimation of the excess entropy of the gamma process and the corresponding information gain. (a) Information gain at $T=\kappa/\lambda=500$. (b,c) Result of estimation of excess entropy (Sample length: $6.0\times10^{8}ms$, bin size: $10ms$) G are fit to the scale of E. (d) Result of estimation from finite length samples.

V. Conclusion

We showed that the excess entropy can be used as a measure for capturing spike trains' structure.

In practice, there seems to be problems since infinite length sample are theoretically needed for the estimation of the excess entropy. Considering convergence, excess entropy needs much long sample. However, numerical estimation shows that we can detect structural difference from short sample to some extent.

Theoretically, excess entropy is an effective measure for spike trains like Gamma process, since spike trains have correlation or structure inside. Spike trains have more than randomness. Excess entropy captures spike trains' structure.

Also, excess entropy is a measure similar to information gain. Excess entropy directly uses spike trains and preserves structure. On the other hand, the approach to spike trains by using information gain cannot detect structural elements beyond ISI distribution. Excess entropy can detect higher order structure above ISI distribution such as temporal coding. Thus, excess entropy can be expected to be an effective measure for spike trains.

References

[1] Feldman DP, McTague CS, and Crutchfield JP. (2008), The Organization of Intrinsic Computation: Complexity-Entropy Diagrams and the Diversity of Natural Information Processing. Chaos. Vol.18, No.4: 59-73.

[2] Grassberger P. (1986) Toward a quantitative theory of self-generated complexity. Int. J. Theor. Phys. 25:907-928.

[3] Bialek W, Nemenman I, Tishby N. (2001) Predictability, Complexity, and Learning. Neural Computation. Vol.13, No.11:2409-2463.

[4] Schürmann T and Grassberger P. (1996) Entropy Estimation of Symbol Sequences. Chaos. Vol.6, No.3:414-427.

[5] Fujiwara K and Suzuki H. (2009) Evaluation of Analysis Methods on Neural Spike Trains using Stochastic Models (in Japanese). SEISANKENKYU. Vol.61, No.2:122-124.

[6]Herzel H, Schmitt A and Ebeling W. (1994) Finite sample effects in sequence analysis. Chaos, Solitons, and Fractals. Vol.4:97-113.

[7] Daley DJ and Vere-Jones D. (2005). An introduction to the theory of point processes, Volume 2. Springer.

[8] Kostal L and Lansky P. (2006). Similarity of interspike interval distributions and information gain in a stationary neuronal firing. Biol. Cybern. Vol.94:157–167.