Neural Networks 95 (2017) 57–65

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Entropy factor for randomness quantification in neuronal data

K. Rajdl *, P. Lansky, L. Kostal

Institute of Physiology, Academy of Sciences of the Czech Republic, Videnska 1083, 142 20 Prague 4, Czech Republic

ARTICLE INFO

Article history:
Received 28 March 2017
Received in revised form 27 July 2017
Accepted 28 July 2017
Available online 15 August 2017

Keywords:
Variability measure
Shannon entropy
Renewal process
Fano factor

ABSTRACT

A novel measure of neural spike train randomness, an entropy factor, is proposed. It is based on the Shannon entropy of the number of spikes in a time window and can be seen as an analogy to the Fano factor. Theoretical properties of the new measure are studied for equilibrium renewal processes and further illustrated on gamma and inverse Gaussian probability distributions of interspike intervals. Finally, the entropy factor is evaluated from the experimental records of spontaneous activity in macaque primary visual cortex and compared to its theoretical behavior deduced for the renewal process models. Both theoretical and experimental results show substantial differences between the Fano and entropy factors. Rather paradoxically, an increase in the variability of spike count is often accompanied by an increase of its predictability, as evidenced by the entropy factor.

1. Introduction

One of the most important questions in neuroscience is how information is transferred in the nervous system. Some aspects of the mechanism are clear and generally accepted—information is transferred using spikes transmitted by neurons, coded only by times of their occurrence, not by their size or shape (Gerstner & Kistler, 2002; Rieke, Warl, de Ruyter van Steveninck, & Bialek, 1999). However, more specific details of the mechanism are not obvious, mainly the exact method of coding. The simplest, and most often assumed concept, is that coding is achieved through the spike rate (defined, for example, as the number of spikes fired in a unit of time), called the rate coding. It is a natural idea, as neuronal responses are usually strongly influenced by the presynaptic spiking rate. Nevertheless, the rate reflects only a small part of the character of spike trains and thus a possibility exists that a more complex coding is employed. The codes which assume some rate-independent behavior, given by the specific position of single spikes, are called temporal codes (temporal coding).

There are many possibilities how such a temporal code could operate, with one variant focusing on the variability in spike timing (variability coding). Under this concept, information is transferred not only by the rate, but also using the variability of spike trains. The crucial question while studying the variability is how to quantify it. The two of the most often used measures are the coefficient of variation (CV) and the Fano factor, see (Ditlevsen & Lansky, 2011; Rajdl & Lansky, 2014; Stevenson, 2016) and other references within these texts. The CV is defined as the standard deviation to mean of interspike intervals (ISIs) ratio and the Fano factor is defined as the variance to mean of number of spikes in a time window ratio. The number of experimental studies where these two measures have been applied is practically countless. The definitions of CV and the Fano factor also show two main ways how to see neuronal data—as ISIs or numbers of spikes in a time window of fixed length. A real neuron more likely registers the counts of spikes than the exact lengths of ISIs, nevertheless, both of these variants are useful and provide a different perspective in evaluation of experimental data. Another feature of these two variability measures, up to our knowledge never mentioned, is that they can be seen as variance (of ISIs or counts) related to a corresponding variance of a Poisson process with the same firing intensity. As a Poisson process has a unique position among the random processes used to model spike trains, it is an important property of these measures that increases their interpretability. Although we consider CV and Fano factor to be the most common variability measures, let us note that several others have been proposed and studied. Among others, we may mention CV2 proposed by Holt, Softky, Koch, and Douglas (1996), CVlog employed in Ruigrok, Hensbroek, and Simpson (2011) or coefficient of local variation, Lv, introduced by Shinomoto and his coworkers (Aoki, Takaguchi, Kobayashi, & Lambiotte, 2016; Shimokawa & Shinomoto, 2009; Shinomoto, Miura, & Koyama, 2005). For comparison of various measures in classification of neuronal discharge patterns, see Kumbhare and Baron (2015). Generally, as seen, there are various ways of quantifying the statistical heterogeneity of a given probability law, not only variance or entropy. Among others belong the Gini index, which measures the law's egalitarianism or its alternative—the Pietra index (Eliazar & Sokolov, 2010). This latter measure is especially useful in the case of asymmetric and skewed probability laws and its future applications on neuronal data remain open.

* Corresponding author.
E-mail address: rajdlk@mail.muni.cz (K. Rajdl).

http://dx.doi.org/10.1016/j.neunet.2017.07.016

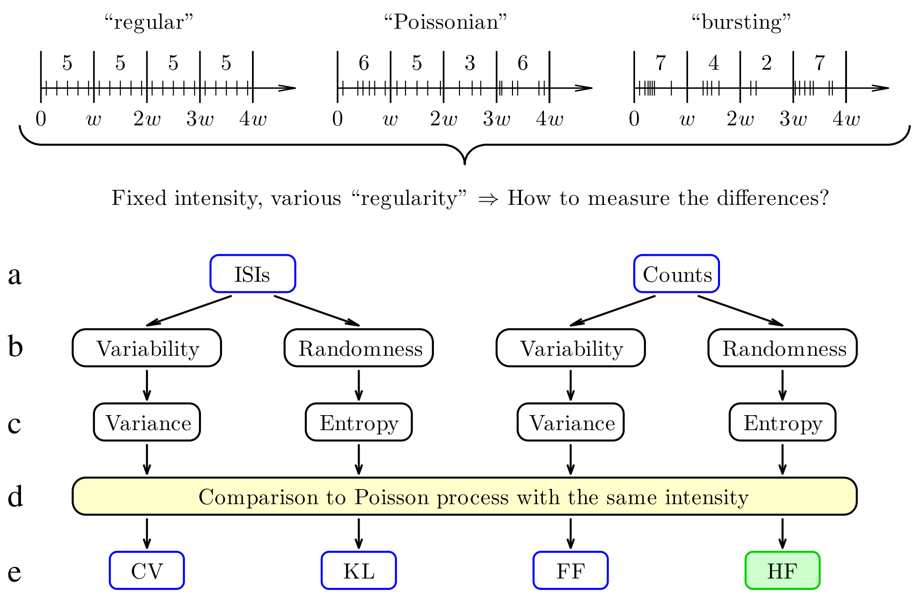

Fig. 1. An overview of selected measures of intensity independent behavior of spike trains. (a) Two main ways how to describe spike trains—using ISIs or counts of spikes in a time window of length w. (b) Two concepts how to understand variability and randomness. (c) Specific characteristics representing variability and randomness—variance and (Shannon) entropy. (d) Relating of given characteristics to a Poisson process of the same intensity. (e) Resulting measures—coefficient of variation (CV), Kullback–Leibler distance (KL), Fano factor (FF) and entropy factor (HF).

Although the variability is commonly used as a general term, there is a similar concept, but not equivalent one, which should be distinguished—the randomness or predictability. Both, the variability and randomness describe the character of spike trains which is not fully determined by the intensity, but there is a clear difference between them. It can be seen, for example, from the fact that even a variable process can be non-random. A very suitable quantity to measure randomness is Shannon entropy (Shannon & Weaver, 1998), which has been widely applied in neuroscience (among many others, Borst and Haag (2001); Chacron, Longtin, and Maler (2001); Ince et al. (2010); Kostal, Lansky, and Rospars (2007); McDonnell, Ikeda, and Manton (2011); Steuer et al. (2001); Strong, Koberle, de Ruyter van Steveninck, and Bialek (1998); Watters and Reeke (2014)). Some randomness measures based on entropy have been proposed in neural context and thoroughly studied (Kostal & Lansky, 2006; Kostal, Lansky, & Pokora, 2011; Kostal et al., 2007). Nevertheless, they focus only on ISIs, creating an alternative to CV. The most suitable way of measuring the randomness of ISIs appears to be through using the Kullback–Leibler (KL) distance of probability density of ISIs to density of ISIs in a Poisson process with the same intensity, thus to density of an exponential distribution.

Quantities based on the KL distance can be seen as analogies to CV representing randomness instead of variance (Kostal, Lansky, & Pokora, 2013). It would be thus natural to use entropy to measure the randomness of spike-counts analogously to the Fano factor. We propose such a measure in this paper. It is defined as the ratio of the Shannon entropy of numbers of spikes in a time interval to the Shannon entropy of a Poisson counting process with the same intensity. Due to similarity to the Fano factor, we call the measure an entropy factor (HF). It creates a natural complement to existing measures. For an overview of selected measures of variability and randomness, see Fig. 1. The measures are classified into two groups according to whether they focus on ISIs or spike counts, the crucial difference between these two approaches being in the presence of an additional parameter (the observation window length) in the distribution of counts. Hence, the variability and randomness measures of spike-counts provide a more complete and extensive characterization of the underlying statistical spiking model, as reported in this paper. As Fig. 1 also shows, all the measures compare a characteristic of given spike train to a Poisson process, which is a simple, but efficient way how to improve interpretation and comparison of various experiments.

The aim of this paper is to explore properties of entropy factor and compare them with properties of Fano factor, to show that there are some fundamental differences between the measures. Behavior of the Fano factor has been studied in various papers, mostly its dependence on the length of the observation window or CV (Nawrot et al., 2008; Pipa, Gruen, & Vreeswijk, 2013; Rajdl & Lansky, 2014) or its statistical properties (Eden & Kramer, 2010). Here, the theoretical properties of the entropy factor are studied and explicit formulas describing its behavior are derived. Firstly, the equilibrium renewal process as a model of spike train is defined. Based on this model, the standard variability measures and their basic properties are summarized. Secondly, the new measure is defined and investigated in the next section. To illustrate its properties, two of the most often assumed models of the lengths of ISIs are used, gamma and inverse Gaussian distributions (Fisch, Schwalger, Lindner, Herz, & Benda, 2012; Koyama & Kostal, 2014; Lansky, Sacerdote, & Zucca, 2016; Nawrot et al., 2008; Omi & Shinomoto, 2011; Ostojic, 2011; Pipa et al., 2013; Shimokawa, Koyama, & Shinomoto, 2010). Finally, the entropy factor is compared to the Fano factor estimated from experimental data.

2. Spike train model

In a formal description, a spike train can be represented as a sequence of times of occurrence of the individual spikes, $X_1, \ldots, X_n$, $n \in \mathbb{N}$, and modeled using a (stochastic) point process. Probably the most often used are renewal processes, which assume that all the ISIs, $T_i = X_{i+1} - X_i$, $i = 1, \ldots, n-1$, are mutually independent and identically distributed random variables. This model is also used here, however, we are well aware that the true character of spike trains can be more complex (Avila-Akerberg & Chacron, 2011; Chacron et al., 2001; Farkhooi, Strube-Bloss, & Nawrot, 2009; Schwalger, Droste, & Lindner, 2015).

A renewal process is defined by a continuous positive random variable T, representing the lengths of ISIs, with a probability density function $f(t)$, cumulative distribution function $F(t)$, and mean $\mu = \mathrm{E}(T)$. To fully specify the renewal process, it is also necessary to state the relationship between the sequence of spikes