Similarity of interspike interval distributions and information gain in a stationary neuronal firing

Lubomir Kostal, Petr Lansky

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

Address for correspondence:

Lubomir Kostal

The Institute of Physiology, Academy of Sciences of The Czech Republic,

Videnska 1083,

142 20 Prague 4,

The Czech Republic

tel: +420 2 4106 2276,

fax: +420 2 4106 2488,

e-mail: kostal@biomed.cas.cz

Similarity of interspike interval distributions and information gain in a stationary neuronal firing

Abstract

The Kullback-Leibler (KL) information distance is proposed for judging similarity between two different interspike interval (ISI) distributions. The method is applied by comparison of four common ISI descriptors with an exponential model which is characterized by the highest entropy. Under the condition of equal mean ISI values, the KL distance corresponds to information gain coming from the state described by the exponential distribution to the state described by the chosen ISI model. It has been shown that information can be transmitted changing neither the spike rate nor coefficient of variation (CV). Furthermore the KL distance offers an indication of the exponentiality of the chosen ISI descriptor (or data): the distance is zero if, and only if, the ISIs are distributed exponentially. Finally an application on experimental data coming from the olfactory sensory neurons of rats is shown.

1 Introduction

The discharge activity of neurons is composed of the series of events called action potentials (or spikes). It is generally accepted that these action potentials form the dominant mode of communication in the central nervous system of living organisms. Since Shannon developed his general and rigorous theory of communication and information transmission in electro-engineering systems (Shannon, 1948), many scientists of various background (biologists, engineers, mathematicians) have tried to apply it to the study of the properties of neural systems. Probably the most fundamental questions point to a problem of neuronal coding, (Rieke et al., 1997).

The classical results in early neuroscience (Adrian, 1928) show that the number of spikes per a time period (firing rate) is related to the stimulus intensity, i.e., the firing rate increases with increasing stimulus intensity (generally non-linearly). The idea that most of the information is transferred by this rate coding is probably the oldest existing, however many related questions arise; for an overview see, e.g., Gerstner and Kistler (2002). There are examples of situations in which time averaging (counting) is hardly possible. This all leads to a different view of neural coding, where exact timing of spikes or their temporal pattern plays a key role (Buracas and Albright, 1999; Johnson and Glantz, 2004; Stevens and Zador, 1995). Whereas rate codes and temporal spike codes are shown to be compatible under many circumstances (Gerstner and Kistler, 2002), it is clear that infinitely more different spike records may have the same rate. Therefore there is a need for the quantification of the differences among these firing patterns and the information they transfer (Bhumbra et al., 2004; Buracas and Albright, 1999; Nemenman, 2004; Paninski, 2003; Rieke et al., 1997; Stein, 1967; Strong et al., 1998; Zador, 1998).

Neuronal firing under constant conditions is often described as a renewal process of interspike intervals (ISI). Then the ISIs are realizations of a positive random variable T and are fully characterized by the probability density function, $f = f(t)$, where $f(t)\,dt = \text{Prob}(T \in [t, t + dt))$ (Cox and Lewis, 1966). The renewal character of the ISIs implies stationarity of the neuronal activity. Other characteristics, for example the autocorrelation function (renewal density), can be easily computed (see Perkel et al. (1967) for details) and indicates important features of the mechanism behind neuronal firing, e.g., it is constant for the exponential model of ISI.

The basic statistical description of the ISIs can give some information related to the above mentioned possibilities of information transfer. The terms (firing) rate, mean rate, frequency or mean/instantaneous frequency are used differently by different authors, depending on the context (Lansky et al., 2004). We use the term "firing frequency", ν, for the reciprocal value of the mean interspike interval, $\nu = 1/E(T)$, where the mean value $E(T)$ is estimated by the average length of ISI. The coefficient of variation CV (a ratio of standard deviation to the mean value) gives preliminary information about the temporal dispersion of neuronal discharge within a spike train, for a review see Christodoulou and Bugmann (2001). An important characteristic of the CV is, that it gives an indication of the exponential ISI distribution – the most prominent of all ISI distributions – when $CV = 1$.

The first and second-order statistical features are easy to estimate from experimental data. From the statistical point of view, the higher moments provide information about the shape of the distribution (ISI), that cannot be obtained using only the first- and second-order moments. Thus, the higher moments help us to quantify and measure some of the features, that are visible in histograms but not in the mean and CV. The higher moments represent the step between estimate of ISI density and the mean and CV only. On the other hand the estimates of higher moments cannot be reliably determined from samples of relatively small size as is normal with the neuronal data. Thus the attempts to use higher moments are not so frequent as using mean and CV (Han et al., 1998; Lewis et al., 2001; Ruskin et al., 2002; Shinomoto et al., 2002). There is a strong demand for measuring the variablity of neuronal discharge and CV is not the only existing method. Based on the results of Holt et al. (1996), Shinomoto et al. (2003) introduce a local measure of spike train variability, L_V, which is not based on the statistical moments. Some properties of L_V are similar to those of CV, e.g., they are both zero for a regular ISI sequence and for a Poisson process $L_V = 1$. Analogously to CV, L_V is not unique in the sense, that $L_V = 1$ does not imply Poisson nature of spikes.

There exist many neuronal models of different types and for their mutual comparison, or their fitting to data, various methods have been applied. Most of these methods are closely related to the already mentioned coding strategies. The aim of this paper is to propose a measure of deviation between two models of ISI, respectively between a model and data and investigate its properties. We will use tools provided by information theory, employing the Kullback-Leibler (KL) distance for the purpose and show explicitly how the most common ISI descriptors differ from the exponential one using this measure. Also an example on experimental data will be presented to provide a practical illustration.

2 Theory and methods

2.1 Entropy and the Kullback-Leibler distance

The concept of entropy was introduced into statistical information theory by Shannon (1948). For a discrete probability distribution with n possible states, each with a probability $p_i, i \in \{1, \ldots, n\}$, the entropy H is defined as

$$H = -\sum_{i=1}^{n} p_i \ln p_i. \tag{1}$$

The entropy H measures the 'randomness' of the distribution, see Shannon (1948) for details. It is maximized when all p_i's are equal. In information theory the logarithm base is chosen to be 2, as it is related to 'bits' of information. For simplicity of calculation, we use a natural logarithm. The units are then commonly called 'nats'.

The differential entropy h (Cover and Thomas, 1991) of a continuous probability distribution f defined on $[0, \infty)$ is

$$h(f) = -\int_0^\infty f(t) \ln f(t)\, dt. \tag{2}$$

It is possible to rewrite the equation (2) into a form based on the cumulative distribution function, $F(t) = \int_0^t f(x)\, dx$, by using the method of Ling and He (1993), which gives

$$h(f) = -\int_0^1 \ln \frac{dF(t)}{dt}\, dF(t). \tag{3}$$

Both definitions, (2) and (3), play their role from the point of view of numerical estimation in dependency whether the density or the cumulative distribution function is employed.

Though formula (2) is analogous to equation (1), differential entropy does not have the same properties and intuitive interpretation as entropy H of a discrete random variable. Namely, h given by formula (2) cannot be used directly to measure the information content of a random variable as it may become negative and depends on the scale of the random variable (Cover and Thomas, 1991; Shannon, 1948). To overcome these difficulties we measure the information content of a random variable with a density f 'against' some reference state given by another random variable with density g (both defined on $[0, \infty)$) as a KL distance of these two distributions,

$$\mathrm{KL}(f, g) = \int_0^\infty f(t) \ln \frac{f(t)}{g(t)}\, dt. \tag{4}$$

This approach was proposed by Tarantola (1994); Tarantola and Valette (1982). The reference state described by the probability density function g is termed the 'state of null information' or the 'state of total ignorance' (Jaynes, 1968; Tarantola, 1994). The quantity $\mathrm{KL}(f, g)$ provides a measure for the information content of f (or information 'gained' from f when knowing g). Though it depends on the template distribution g, it is invariant with respect to a transformation of variable and always non-negative. This quantity can also be interpreted as a 'coding inefficiency' when using distribution g to 'encode' distribution f (more details in Cover and Thomas (1991)).

Generally, the KL distance is a measure of the 'closeness' between two probability density functions (Cover and Thomas, 1991), though it does not form a measure in the metric sense. It is not symmetrical, however one can use its symmetric extension (resistor-averaged KL distance), see e.g., Rozell et al. (2004). This proves to be useful when comparing more than two probability densities simultaneously and also when it is not clear what is the template (ideal) distribution to which the others are related.

2.2 The KL distance between the exponential and general model

The exponential distribution plays a key role in neuronal modeling. Its probability density function f_e is

$$f_e(t) = a \exp(-at), \tag{5}$$

where for the parameter $a > 0$ holds $E(T_e) = 1/a$. We see that the firing frequency ν completely characterizes the distribution, $\nu = 1/E(T_e) = a$. For the exponential distribution holds $CV = 1$, independently of the parameter a, but the reverse implication that $CV = 1$ guarantees exponentiality is not true.

One of the most important features of exponential distribution, in the context of information theory, is that among all probability distributions on the real positive half-line with fixed $E(T)$, the exponential distribution maximizes the entropy h.

The entropy $h(f_e)$ of the exponential distribution is

$$h(f_e) = 1 - \ln a. \tag{6}$$

Thus among all possible stationary neuronal discharge activities the one described by the exponential distribution of ISI is the most random. It is natural to choose it as a reference state, the state of null information.

We will measure the information 'gained' when changing from the response state described by the exponential distribution f_e given by equation (5) to a response state described by a probability density function f with mean value $E(T)$. For this purpose we employ formula (4), in which by substituting $g = f_e$ we obtain

$$\mathrm{KL}(f, f_e) = aE(T) - \ln a - h(f). \tag{7}$$

If the mean values of f and f_e differ, then the spike trains generally carry different information in the concept of rate coding. However, as we want to analyze possibilities for a mechanism different from rate coding, let us suppose that the means of f and f_e are equal, i.e., $E(T) = E(T_e) = 1/a$. Then for the KL distance of these two distributions we obtain the following difference in entropies (Soofi et al., 1995)

$$\mathrm{KL}(f, f_e) = h(f_e) - h(f) = 1 + \ln E(T) - h(f). \tag{8}$$

The formula (8) for the information gained when changing from the reference state is also intuitively consistent with the notion of information as a reduction in entropy (Shannon, 1948; Borst and Theunissen, 1999).

We further precise the interpretation of the KL distance as a measure of information using the notion of mutual information as it is commonly used in the neuronal context. The mutual information $I(S;R)$ (Cover and Thomas, 1991) determines the dependence between stimuli S and responses R (Borst and Theunissen, 1999). Let the set of stimuli $S = \{s_i\}_{i=1}^{n}$ be discrete and the set of responses continuous $R = \{T, T > 0\}$, where T is an ISI. Mutual information can be formally expressed as $I(S;R) = \sum_i p(s_i)\, i(R|s_i)$, where $i(R|s_i)$ is called the specific information due to the stimulus s_i. Analogously to DeWeese and Meister (1999) we express $i(R|s_i)$ as

$$i(R|s_i) = h(R) - h(R|s_i). \tag{9}$$

It follows from formula (9) that the specific information is large for those stimuli that have only a few different responses associated with them, because $h(R|s_i)$ is the uncertainty in response given a particular stimulus s_i. In the limits, if there is only a single response related to the stimulus s_i, then $h(R|s_i) = 0$. We have limited ourselves to the case in which the ISIs are described by a renewal process with probability density function f and the stimuli conditions are constant in time. Under these two assumptions we can assign to each ISI distribution with density f a chosen stimulus s_i. The uncertainty in response given stimulus s_i then becomes $h(R|s_i) = h(f)$, i.e., the smaller the value of $h(R|s_i)$, the greater the information gain due to s_i. The remaining term in formula (9), the marginal entropy $h(R)$, depends on the distribution of stimuli. This distribution affects the absolute value of $i(R|s_i)$; however, the relative encoding efficiency for different s_i's remains unchanged. It is useful to view $h(R)$ as the entropy of the spontaneous neuronal activity, see Chacron et al. (2001) for details. If this activity is described by the Poisson process, then formula (9) corresponds to the expression for the KL distance (8). Furthermore using formula (8) on two renewal processes described by distributions f_A and f_B yields

$$\mathrm{KL}(f_B, f_e) - \mathrm{KL}(f_A, f_e) = h(f_A) - h(f_B). \tag{10}$$

The KL distance thus also provides a way to determine the specific information in cases when the spontaneous activity is not a Poisson process.

It follows from the definition (4), that the KL distance may sometimes tend to infinity. This is due to the fact that a continuous random variable generally carries an infinite amount of information (van der Lubbe, 1997, p. 171). Nevertheless, this fact can be considered as merely formal and without consequences, realizing that in practice we are always working with finite precision on a finite time scale. It also follows from equation (8) that the KL distance between any density and the exponential one with the same mean is positive and equal to zero if, and only if, $f = f_e$. This statement of equivalence, in contrast to equality $CV = 1$, allows us to judge exponentiality. In statistical literature the KL distance also appears as one of approaches to exponentiality testing (Ebrahimi et al., 1992; Choi et al., 2004).

2.3 Evaluation of the KL distance from data

It follows from equation (8), that the evaluation of the KL distance from experimental data is reduced to the problem of estimating the entropy of the involved distributions. Estimation of the entropy of exponential distribution is equivalent to the estimation of the mean value $E(T_e)$. Therefore, the only open problem is estimation of the entropy $h(f)$. Two possible approaches to $h(f)$ estimation exist: the parametric one, where from a preselected model its parameters are estimated and then the entropy is calculated, and the non-parametric one, where $h(f)$ is estimated directly from data without specifying the model.

The first approach has been recently exploited, e.g., by Reeke and Coop (2004), for the case of shifted gamma distribution (with three independent parameters). To illustrate this approach we will estimate the parameters of several common ISI descriptors. The goodness of fit of the data to the models will be checked by the standard one-sample Kolmogorov-Smirnov (KS) test (Gibbons, 1971) using the estimated parameters. In the cases when the null hypothesis cannot be rejected the KL distance will be evaluated using this parametric approach. In general, however, the application of the KL distance method is not conditioned by the KS test.

As regards the second approach, many possibilities to estimate $h(f)$ non-parametrically exist and have been discussed in literature, see e.g. Beirlant et al. (1997), Tsybakov and Meulen (1996) for an overview. We may divide non-parametric entropy estimators into two groups. The so called "plug-in" estimators using formula (2) directly, thus some estimate of density has to be constructed, employing, e.g., histograms, kernel-smoothed probability densities; and the estimators where the probability density function does not appear explicitly, usually derived using formula (3) in some way. To this group belong, e.g.: the estimator of Vasicek (Vasicek, 1976), or estimators based on nearest sample spacings, see Beirlant et al. (1997). The choice of concrete estimator is strongly situation and purpose dependent. Histogram estimates or smoothed kernel densities are well known to be greatly affected by bin-size resp. bandwidth. The fact that the estimated probability density function has limited support to the real positive half-line may in some cases be the next reason why not to employ kernel-smoothed density estimates. In our case explicit evaluation of probability density function f is not needed so we avoided "plug-in" entropy estimators. To illustrate the non-parametric KL distance evaluation on the experimental data we use the simple binless entropy estimator of Vasicek (1976), which is known to converge and behave well for various types of data (Ebrahimi et al., 1992; Miller and Fisher III, 2003). The examples of estimated entropy values relevant to our calculations are included in Fig. (1). Given n ISIs $\{t_1, t_2, \ldots, t_n\}$ we sort them with respect to their length $\{t_{[1]}, t_{[2]}, \ldots, t_{[n]}\}$.

Then the estimated entropy, $h(\text{data})$, is

$$h(\text{data}) = \frac{1}{n}\sum_{i=1}^{n} \ln\left[\frac{n}{2m}(t_{[i+m]} - t_{[i-m]})\right], \tag{11}$$

with a free parameter $m < n/2$ and $t_{[j]} = t_{[1]}$ for $j < 1$ and $t_{[j]} = t_{[n]}$ for $j > n$. Note that estimator (11) is based on discretization of formula (3) using empirical cumulative distribution function. The additional parameter m allows the avoidance of possible numerical problems resulting from such discretization. The relation between n and m was determined by Ebrahimi et al. (1992). For sample sizes $n \geq 200$ holds $m = 13$, which was the value of m we used in the entropy estimation.

3 Results and discussion

This section is divided into two parts. First, we investigate the KL distances of several commonly used ISI distributions from the exponential one under the condition of equal mean values. All the tested distributions are described by two independent parameters related to their mean and variance. It follows from the formula (8), valid under equality of the mean values, that the KL distance should also depend on two parameters. Employing the scaling property of differential entropy (Cover and Thomas, 1991, p. 233) together with formula (8) yields, that in our case the KL distance does not depend on $E(T)$, as will be shown explicitly in the studied examples. To provide a unified view of the results, we chose coefficient of variation CV as an independent variable. This particular choice also seems to be the most natural for the purpose of insight into possible mechanisms of neural coding mentioned earlier. In the light of our setting, the dependence KL(CV) allows to judge possible amount of information being transmitted between two particular states of neuronal firing as a function of spike train variability (while the firing frequency is not changing). Secondly, an application of the KL distance on the experimental data is shown. The obtained results are related to those from the first part.

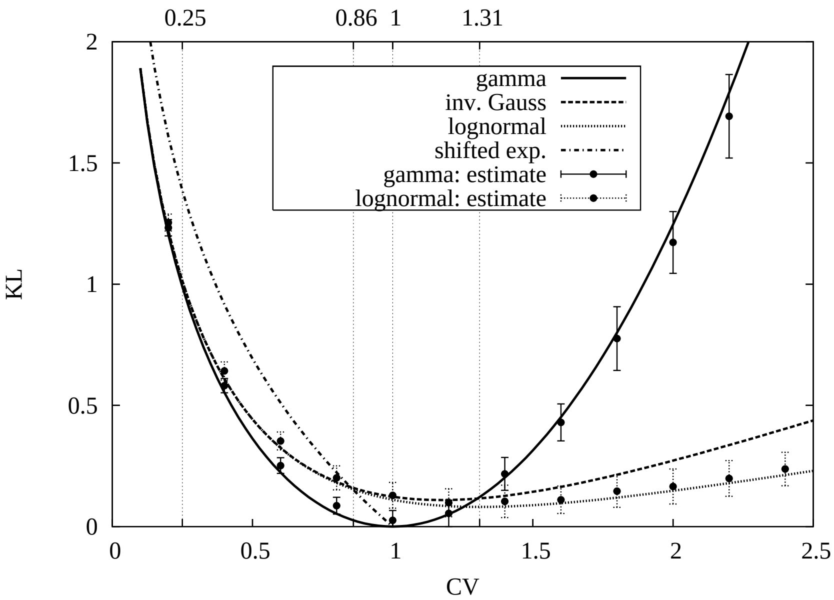

Fig. 1: The Kullback-Leibler (KL) distance as a function of CV for the tested models. Four values of CV (marked on the minor horizontal axis) are chosen to produce Fig. 2. The KL distance of gamma and shifted exponential (available only for $CV \leq 1$) distributions is zero for $CV = 1$, implying that at this point they become exponential. For $CV \to 0$ and for $CV \to \infty$ the KL distance tends to infinity for the tested distributions. With CV increasing from zero the KL distances of lognormal, inverse Gaussian and gamma distributions are initially the same. Then gamma branches off at $CV \approx 0.25$ and the lognormal and inverse Gaussian depart at $CV \approx 1$. For lognormal and inverse Gaussian distributions the distance never reaches zero and even the minima are not located at $CV = 1$ implying that for no combination of parameters these distributions become exponential. A common feature of the tested models is that near $CV = 1$ the values of KL distances are generally low, however, minimum for the lognormal, resp. inverse Gaussian distribution is located at $CV \approx 1.31$, resp. $CV \approx 1.17$. Furthermore for these two distributions, especially for the lognormal, the KL distance grows very slowly with increasing CV, compared to the gamma distribution. Estimated values from simulated data with sample sizes $n = 500$ and their standard deviations over 100 trials are shown for comparison with theoretical results.

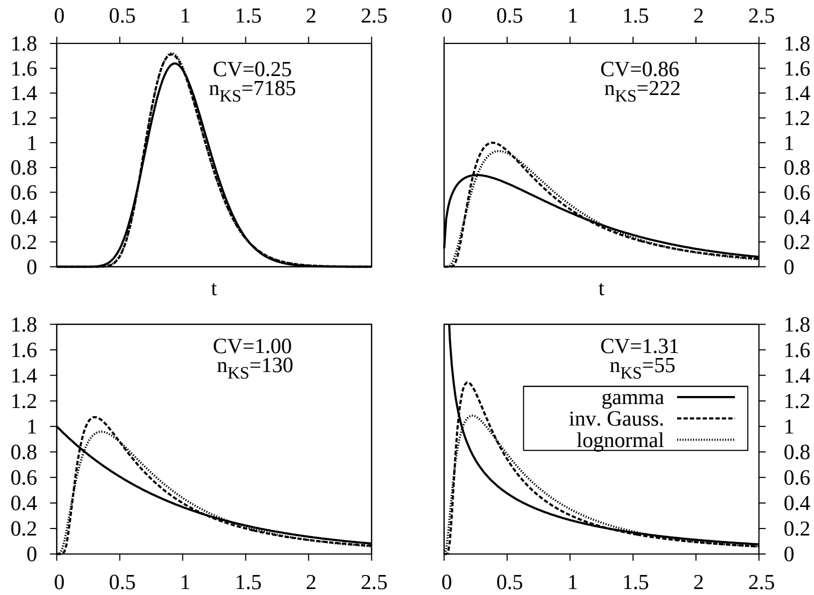

Fig. 2: Probability density functions $f(t)$ of the gamma, inverse Gaussian and lognormal distributions with $E(T) = 1\,\mathrm{s}$ for four different values of CV indicated in Fig. 1. The lowest number n_{KS} of ISIs, required for the Kolmogorov-Smirnov test to "distinguish" between gamma and inverse Gaussian distributions at 5 % significance level is given. We can see that for $CV = 0.25$ the three probability density functions are hardly distinguishable. The number n_{KS} is too large compared to usual sample sizes obtained in experiments. The probability density functions start to differ more for $CV = 0.86$ (this is the value, for which the KL distances of lognormal, inverse Gaussian and shifted exponential distribution approximately equal), n_{KS} falls into the average length of experimental records. At $CV = 1$ the gamma distribution becomes exponential (KL = 0), while the lognormal and inverse Gaussian do not, and they are even not "close" to exponential as much as possible. The minimal KL distance of lognormal distribution (corresponding to its maximum "closeness" to exponential distribution) is at $CV = 1.31$. At this point the KL distance of inverse Gaussian and gamma distributions is roughly the same, though their probability densities differ strikingly (n_{KS} is merely 55).

3.1 Gamma distribution

Gamma distribution is one of the most frequent statistical descriptors of ISIs (Hentall, 2000; Levine, 1991; McKeegan, 2002; Mandl, 1992). Its probability density function is

$$f(t) = \frac{b^a t^{a-1} e^{-bt}}{\Gamma(a)}, \tag{12}$$

where $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t)\, dt$ is the gamma function and $a > 0$, $b > 0$ are the parameters. From formula (12) these relations follow

$$E(T) = \frac{a}{b}, \quad CV = \frac{1}{\sqrt{a}}. \tag{13}$$

Using formula (2), the entropy of gamma distribution is

$$h(f) = a + (1-a)\Psi(a) - \ln b + \ln\Gamma(a), \tag{14}$$

where $\Psi(a) = \frac{d}{da} \ln\Gamma(a)$ is the digamma function. By substituting equations (13) and (14) in formula (8) we find the KL distance of the gamma distribution from the exponential one in the explicit form as a function of CV

$$\mathrm{KL}(CV) = 1 - \ln CV^2 - \ln\Gamma\left(1/CV^2\right) + \frac{\Psi(1/CV^2) - 1}{CV^2} - \Psi(1/CV^2). \tag{15}$$

This dependence is shown in Fig. 1 together with the estimated values and standard deviations from simulated data. We see, that the error of estimation is relatively small and a possible positive bias with respect to true values is negligible for tested sample sizes. The density f given by formula (12) is exponential for $a = 1$ and therefore $\mathrm{KL}(CV = 1) = 0$. The KL distance tends to infinity for $CV \to 0$ and $CV \to \infty$. We can see from the figure, that $\mathrm{KL}(CV)$ increases rapidly for $CV > 1$ especially if compared to the other models presented here. For $CV < 0.25$ (approximately) the KL distances of gamma, lognormal and inverse Gaussian distributions become the same. Probability densities of investigated models for the values of CV selected with respect to results in Fig. 1 and with mean ISI equal to one are presented in Fig. 2. We can see that the gamma distribution ranges from the shape similar to the Gaussian distribution at $CV = 0.25$ via the "typical" shape of gamma distribution at $CV = 0.86$ to the exponential distribution at $CV = 1$. Finally, for $CV > 1$ the gamma distribution is characterized by a majority of very short intervals and long tail of the distribution. This feature, at least at the first approximation, looks like bursting type of neuronal activity.

The KL distance provides a different approach to the comparison of two distributions if compared with the KS test: it is 'global', i.e., it takes into account the whole shapes of the compared density functions, while the key parameter for the KS method is the extreme local 'distance' of the curves. Thus one can construct cases for which the results given by KL and KS are contradictory. The most striking example is when the template distribution is zero on some interval where the other function is not. Then the KL distance is infinite, while the KS statistics may be even an infinitely small one. Thus KS represents an alternative to KL, however, with completely different properties.

3.2 Inverse Gaussian distribution

The Inverse Gaussian distribution (Chhikhara and Folks, 1989) is often used to describe neural activity (Iynegar and Liao, 1997) and fitted to experimentally observed ISIs (Gerstein and Mandelbrot, 1964; Levine, 1991). This distribution results from the Wiener process with positive drift (the depolarization has a linear trend to the threshold) as a stochastic neuronal model (Berger et al., 1990; Berger and Pribram, 1992; Levine, 1991). The probability density of inverse Gaussian distribution can be expressed as a function of two parameters $a > 0$ and $b > 0$

$$f(t) = \sqrt{\frac{a}{2\pi b t^3}} \exp\left[-\frac{1}{2b}\frac{(t-a)^2}{at}\right], \tag{16}$$

with

$$E(T) = a, \quad CV = \sqrt{b}. \tag{17}$$

The fact that the mean and coefficient of variation can be easily related to the parameters a, b, which act as a scale and shape parameter, respectively, of the curve, is of practical advantage.

The KL distance of the inverse Gaussian distribution from the exponential one as a function of CV is

$$\mathrm{KL}(CV) = \frac{1}{2} \ln \frac{e}{2\pi} - \ln CV + \frac{3}{\sqrt{2\pi}} \frac{e^{1/CV^2}}{CV} K_{\frac{1}{2}}^{(1,0)}(1/CV^2), \tag{18}$$

where $K_\nu^{(1,0)}(z)$ is the derivative of the modified Bessel function of the second kind (Abramowitz and Stegun, 1972)

$$K_\nu^{(1,0)}(z) = \frac{\partial}{\partial \nu} K_\nu(z). \tag{19}$$

The dependence is shown in Fig. 1. Due to the fact that neither for any combination of parameters nor asymptotically the inverse Gaussian distribution is exponential, the KL distance is not zero for $CV = 1$. This fact can be interpreted in this way, following formula (8): even if there is no change in $E(T)$ and $CV = 1$ there still may be some gain of information coming from the reference state (described by exponential distribution). Following Fig. 1 we can judge the information gain under the condition of fixed CV. The minimum of $\mathrm{KL}(CV)$ for the inverse Gaussian distribution is located at $CV \approx$ 1.173. We can see a difference here, compared to the previous case of the gamma distribution. It has been already noted that the condition $CV = 1$ does not imply exponentiality, but in this case even the minimal distance of the inverse Gaussian to exponential distribution is not located at $CV = 1$. In Fig. 2 there are ISI probability densities for four selected values of CV. The important fact is that for $CV = 0.25$ the models are practically indistinguishable. Even at $CV = 0.86$ and $CV = 1$, lognormal and inverse Gaussian distributions are of very similar shape. Finally, compared to gamma, inverse Gaussian lacks very short ISIs which can be considered as positive feature of this distribution, if refractoriness should be taken into account.

3.3 Lognormal distribution

The lognormal distribution of ISI, with some exceptions (Bershadskii et al., 2001), is rarely presented as a result of a neuronal model. However, it represents quite a common descriptor in ISI data analysis (Levine, 1991), e.g., a mixture of two lognormal distributions has been used recently (Bhumbra et al., 2004).

The lognormal distribution is given by the probability density function

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln t - m)^2}{2\sigma^2}\right], \tag{20}$$

where m and $\sigma > 0$ are the parameters. Because the variable $\ln T$ is normally distributed it follows

$$E(T) = \exp(m + \sigma^2/2), \quad CV = \sqrt{\exp \sigma^2 - 1}. \tag{21}$$

We use formula (8) to compute the KL distance of lognormal distribution from the exponential one. Expressing the KL distance as a function of CV we come to a formula

$$\mathrm{KL}(CV) = \frac{1}{2}\left[\ln \frac{CV^2 + 1}{\ln(CV^2 + 1)} + \ln \frac{e}{2\pi}\right]. \tag{22}$$

From there it follows that

$$\frac{d\mathrm{KL}}{d(CV)} = \frac{CV[\ln(CV^2+1)-1]}{(CV^2+1)\ln(CV^2+1)} \tag{23}$$

and the minimum is at $CV = \sqrt{e-1} \approx 1.311$. The dependence is shown in Fig. 1 together with the estimated values and standard deviations from simulated data, which shows good correspondence between theoretical and numerical approaches. The estimate is not systematically biased and the relative error is very small. Again, as in the case of inverse Gaussian distribution, we see that even if $CV = 1$ the distribution is not exponential. Yet again – the minimal possible deviation of lognormal distribution from exponential one is not at $CV = 1$. It is interesting that for $CV < 1$ (approximately) there is no difference in lognormal and inverse Gaussian distributions from the perspective of the KL distance. The equality in the KL distance, of course, does not imply that these distributions are identical, see Fig. 2. Nevertheless, their similarity is very high.

3.4 Distributions involving a refractory period

The refractory phase is such a state of a neuron, coming immediately after a spike was generated, when it is impossible for another spike to be emitted. In more detail, one can distinguish the absolute refractory phase, when the generation of the next spike is absolutely impossible and the relative refractory phase, when it is merely not probable. The typical duration of the absolute refractory phase is around 2–4 ms, while the relative one may last around 10–20 ms, depending on the definition of "not probable", Gerstner and Kistler (2002). The discussion on the topic of refractory phases and their importance is still ongoing (Berry II and Meister, 1998).

Recently a shifted gamma distribution was used as a generally suitable ISI probability distribution for parametric entropy estimation by Reeke and Coop (2004). The absolute refractory phase is, in this model, described by a shift in time for $\tau > 0$, while parameters a, b are kept the same as in equation (12). Correspondingly, $E(T)$ and CV change in a simple way but the entropy of such distribution is independent of τ as follows from formula (2). In the case of shifted gamma distribution we have three independent parameters and it is no longer possible to describe the KL distance from the exponential model just as a function of one parameter (CV). On the other hand comparing the shifted gamma model to the exponential distribution shifted by the same value τ gives naturally the same results for KL(CV) as given by equation (15). The same is true for any shifted distribution. However, it might be interesting to compare two exponentials, one with and one without refractory phase, as follows.

The probability density function of the shifted exponential distribution with parameter $a > 0$ and a shift $\tau \geq 0$ is

$$f(t) = \begin{cases} 0, & t \leq \tau \\ ae^{-a(t-\tau)}, & t > \tau. \end{cases} \tag{24}$$

Then it follows

$$E(T) = \frac{1 + a\tau}{a}, \quad CV = \frac{1}{1 + a\tau}. \tag{25}$$

It is obvious that in this case it is always $CV < 1$ for $\tau > 0$. Evaluation of the KL distance of the shifted exponential distribution from the exponential one under the condition of the same means using equations (8) and (25) gives

$$\mathrm{KL}(CV) = -\ln(CV). \tag{26}$$

For $\tau = 0$ is $CV = 1$ and f given by formula (24) is exponential which is confirmed by $\mathrm{KL} = 0$. The function $\mathrm{KL}(CV)$ is shown in Fig. 1.

The shape of $\mathrm{KL}(CV)$ given by equation (26) differs from the KL distances of gamma, inverse Gaussian and lognormal models with the corresponding CV. We see that this curve, despite taking the value of zero at $CV = 1$, is the steepest among all of them. Thus we can ask, what is the critical value of CV such that for smaller values the KL distance of the shifted exponential distribution is greater than that of any other tested model? Following Fig. 1 we estimate the critical value of CV as the one for which the KL distances of the lognormal and shifted exponential distributions equal. This results in (a very nice) equation

$$\frac{(CV^2+1)CV^2}{\ln(CV^2+1)} = \frac{2\pi}{e}, \tag{27}$$

which yields $CV \approx 0.86$. Realizing that

$$CV = 1 - \frac{\tau}{E(T)} \tag{28}$$

and using the critical value of CV we receive a critical ratio of the refractory phase to the mean value as $\tau/E(T) \approx 14.4\,\%$. For ratios greater than the critical the KL distance of the shifted exponential distribution is greater than that of other models tested here. For example, if the average absolute refractory phase is 3 ms, the corresponding critical mean value is approximately 21 ms.

3.5 Experimental data

The data comes from extracellular recordings made from olfactory receptor neurons of freely breathing and tracheotomized rats using glass insulated tungsten microelectrodes. Spontaneous, single-unit action potentials were recorded using metal-filled glass micropipettes filled with an alloy of Wood's metal (80%) and indium (20%). The single unit nature of the recorded spikes was controlled during the experiment by triggering the recorded neuron near the background-noise on a storage oscilloscope. More details of the data acquisition is described in Duchamp-Viret et al. (2003). Here we do not distinguish between the data that comes from freely breathing or tracheotomized rats. A comparison of these two conditions are published elsewhere (Duchamp-Viret et al., 2005). The sample sizes range from (circa) $n = 100$ to $n = 2000$ and all records have been tested for nonstationarity (the Wald-Wolfowitz test, serial correlation, periodogram). We use these recordings to illustrate the estimation of the KL distance. Both parametric and non-parametric methods were applied. The results are summarized in Fig. 3, where estimated KL distances are plotted in dependency on the CV along with the theoretical curves of KL distances re-plotted from Fig. 1. Two different categories of data are distinguished, exponentiality rejected or not rejected, based on the Kolmogorov-Smirnov test at 5 % significance level. We can see that even if $CV \approx 1$ the exponentiality may be rejected. We note that the data obeys the general feature of the tested models: the small values of the KL distance are distributed around $CV = 1$. Of particular interest is the asymmetrical distribution of possibly exponential data around $CV = 1$. Except for one spike train all records for which exponentiality is not rejected have $CV < 1$. Furthermore these points (▲) closely follow the theoretical curve for the shifted exponential distribution. Of course from this fact we cannot conclude that these ISI follow this distribution.

For $CV > 1$ the calculated KL distances are far above all considered curves. Though the KL distance does not tell us which particular distribution to use, the difference between theoretical and estimated KL values suggest that the gamma distribution cannot describe the data well. We can expect that analogously to

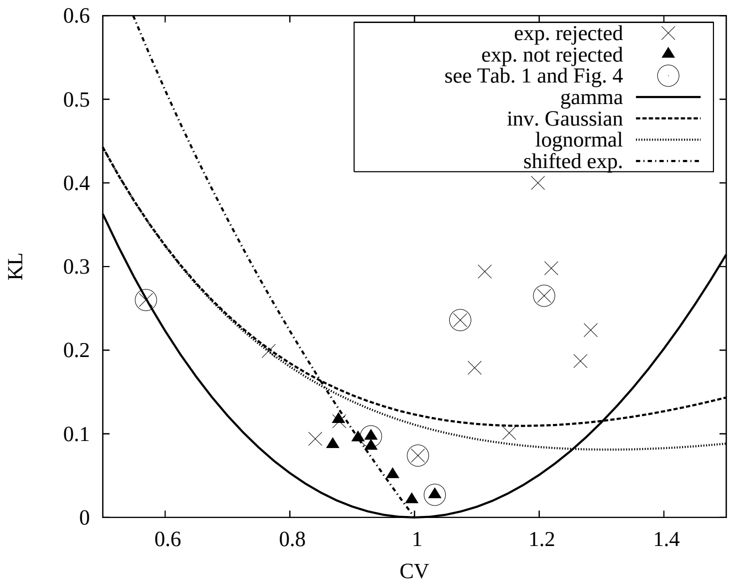

Fig. 3: The Kullback-Leibler (KL) distance as a function of CV for the experimental data. Theoretical curves of the tested models are re-plotted from Fig. 1. Two different categories of data are distinguished based on the Kolmogorov-Smirnov test of exponentiality at 5 % level of significance. Spike trains and ISI histograms of the encircled data are shown in Fig. 4 and those with CV close to unity are given a more detailed treatment in Tab. 1. We see, that even for $CV \approx 1$ there exist data that are not exponential. The points around the value $CV = 1$, where the hypothesis of exponentiality was not rejected, are asymmetric, closely following the theoretical curve for the shifted exponential distribution. Generally we see that the data obey the rule indicated in Fig. 1 for the common ISI descriptors, i.e., the smallest values of the KL distance are around $CV = 1$. For $CV > 1$ the general "course" of the data is even steeper than that of gamma distribution.

olfactory neurons in frogs (Rospars et al., 1994) there is a bursting character in this activity. The bursting activity of the neuron is usually described by a mixture of two distributions, one for interburst ISIs and the other for intraburst ISIs (Mandl, 1992). For example Bhumbra et al. (2004) combines two lognormal distributions given by formula (20), which results in a very flexible model with five unknown parameters. An alternative, and more common, description could be a probability density function f of the mixture of two exponential distributions

$$f(t) = pae^{-ax} + (1-p)be^{-bx} \tag{29}$$

with $p \in (0,1)$ and $a > 0$, $b > 0$, $a \neq b$. We can ask, whether model (29) can be as far from the single exponential model as are the data in Fig. 3. To check this idea we first compute $E(T)$ and CV of the distribution (29)

$$E(T) = \frac{pb + (1-p)a}{ab}, \quad CV = \sqrt{\frac{2pb^2 + 2(1-p)a^2}{(pb + (1-p)a)^2} - 1}. \tag{30}$$

Expressing p and b from equations (30), we re-parameterize the original formula (29) using parameters $E(T), CV$ and a. Though the expressions grow in size quickly and get difficult to handle analytically, we can evaluate the KL distance $\mathrm{KL}(E(T), CV, a)$ from formula (8) numerically. The results for $CV = 1.2$ show that for any $E(T)$ we can find a value of parameter a such that the KL distance of the double exponential is greater than the KL distance of gamma distribution with the same mean and CV. Though it is not possible to fit the theoretical distributions to the data based solely on the KL number, we deduce, that the double exponential distribution has a chance to describe bursting behavior better than gamma distribution.

We chose two pairs of data sets with CV close to unity to show the situation in more detail. The results are summarized in Tab. 1. The parameters of the gamma, inverse Gaussian, lognormal and shifted exponential were estimated from the data. Then the Kolmogorov-Smirnov test was performed to test the goodness of fit at a 5 %

Filename	Exp.	n	$\nu\,[\mathrm{s}^{-1}]$	CV	KL_{data}	KL_{γ}	KL_{shift}
RN20_01	−	1906	6.862	1.005	0.074	–	–
RT48_01	+	355	1.988	1.032	0.027	0.001	–
RT58_01	−	788	5.246	1.073	0.236	–	–
RN30_01	+	549	1.963	0.930	0.097	0.006	0.073

Tab. 1: Comparison of two pairs of spike trains (encircled in Fig. 3 with $CV \approx 1$). The KL distances were estimated both from data (non-parametrically) and from theoretical models (parametrically). The "−" sign in the "Exp." column indicates that the hypothesis of exponentiality was rejected by the Kolmogorov-Smirnov test, while "+" states that it was not rejected (within a 5 % significance level). n is the number of ISIs in the record, ν the firing frequency, CV is the computed coefficient of variation, KL_{data} is estimated directly from data (non-parametrically), KL_{γ} is the corresponding distance of the gamma distribution and KL_{shift} of the shifted exponential. We see, that the KL distances computed parametrically and non-parametrically are rather different. This can be attributed both to wide confidence intervals of estimated parameters and to the behavior of Vasicek's estimator.

significance level. In the case of not rejecting the null hypothesis, the KL distance was estimated parametrically using the previous theoretical results. (Note that for all of these four data sets the inverse Gaussian and lognormal model were rejected.) Though all four data sets have $CV \approx 1$ there are differences in the KL distance from the exponential distribution. One may note that $\mathrm{KL}_{RN20\_01} > \mathrm{KL}_{RT48\_01}$ even though $|1 - CV_{RN20\_01}| < |1 - CV_{RT48\_01}|$. The spike trains and corresponding ISI histograms of the four above mentioned data sets (together with two other records also encircled in Fig. 3) are shown for comparison in Fig. 4, each plotted together with the exponential distribution with the corresponding mean ISI.

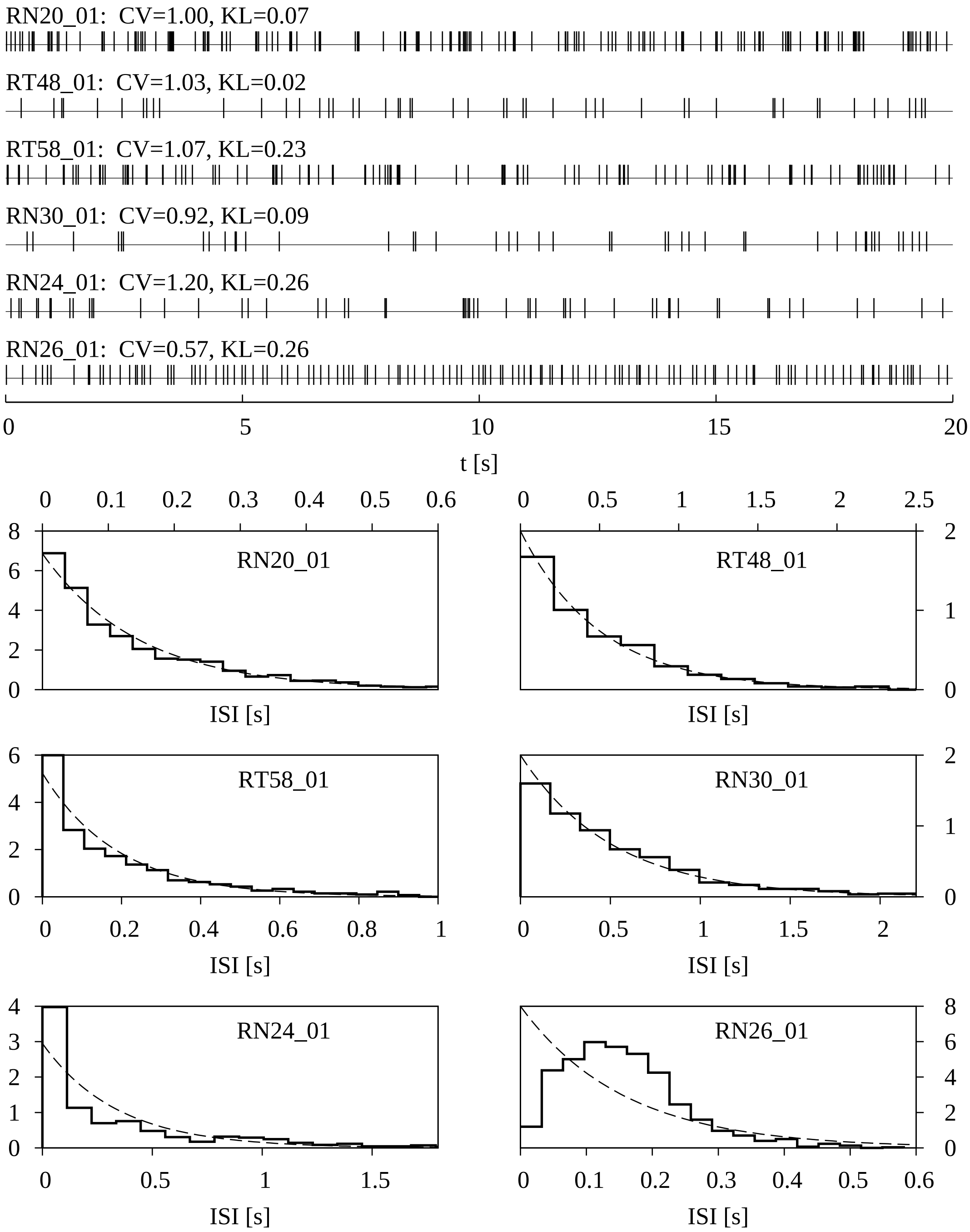

Fig. 4.: (*Caption on the following page.*)

4 Conclusions

The Kullback-Leibler (KL) distance was proposed as a measure of similarity between two interspike interval (ISI) distributions. Choosing the exponential one as the

Fig. 4: First 20 seconds of the spike trains and ISI histograms (of the whole record) for the selected data encircled in Fig. 3 compared with exponential probability density with the corresponding mean ISI. The values of CV and KL distance estimated from data are also indicated. The first four records with CV closest to unity are explored in Tab. 1. The record RN24_01 was chosen to represent the group of data with $CV > 1$ and relatively large KL distance. The sharp increase in frequency of very small ISI values together with relatively flat tail of the histogram and large KL distance value favors the double exponential model as the more probable ISI descriptor compared to the gamma model. On the other hand, the record RN26_01 was chosen for comparison only due to its small value of CV. Except RT48_01 all records have $n > 500$.

template we analyzed the KL distance from both data and models. We selected four common two-parametric distributions: gamma, lognormal, inverse Gaussian and shifted exponential.

Fixing the mean values of exponential and model distribution (or data) to be equal, the KL distance corresponds to information gain coming from a state described by the exponential distribution of ISI to another state described by the model distribution. Thus, it reveals a different mechanism from rate coding of information transmission. Furthermore the KL distance is interpreted in terms of specific information and its usefulness when determining the efficiency of the stimulus encoding is shown. It is natural to express the information gain as a function of a spike train variability, commonly reflected by coefficient of variation (CV). For exponential distribution $CV = 1$, however, the reverse implication does not hold. The KL distance offers an alternative tool to CV to judge exponentiality of the model distribution (or data), because the exponentiality is guaranteed if, and only if, the distance is zero. Furthermore, while there are tools to measure the variability of spike trains, the KL distance measures a different characteristics – the randomness of the underlying process.

The following inference can be made on the basis of the KL distance of ISIs distributions:

1. Even if neither spike frequency nor coefficient of variation changes, the KL distances to the exponential distribution can be different for different models (data) and thus there is still a gain of information coming from one state to another.

2. It is well known that the lognormal and inverse Gaussian distributions never become exponential, but surprisingly their minimal KL distances to this distribution are not located at $CV = 1$.

3. For CV increasing from zero (regular spiking) the KL distances of lognormal, inverse Gaussian and gamma distributions are initially the same. Then gamma branches off at $CV \approx 0.25$ and the lognormal and inverse Gaussian depart at $CV \approx 1$. For low values of CV the differences among the distributions are hardly distinguishable for usual sample sizes available in neural spiking data studies.

4. For lognormal and inverse Gaussian distributions, the KL distance grows very slowly for $CV > 1$, compared to the gamma distribution, and their distances to the exponential distribution are practically the same for $CV = 1$ as for $CV < 2$.

5. The KL distance of shifted exponential ($CV < 1$) is the steepest from all of the investigated alternatives.

6. As shown in experimental data, even if $CV \approx 1$, the ISI distribution may not be exponential, this is confirmed by the Kolmogorov-Smirnov test. Although the data follow the general features indicated by theoretical results, the "course" of their KL distance for $CV > 1$ is steeper even than that of gamma distribution. This suggests/confirms the bursting character of this data.

7. The occurrence of data where exponentiality cannot be rejected, is asymmetric around $CV = 1$ and closely follows the theoretical curve for the shifted exponential distribution.

Acknowledgements

Helpful conversations with Priscilla E. Greenwood and Jean-Pierre Rospars are acknowledged. The authors wish to thank Patricia Duchamp-Viret and Michel Chaput for making their experimental data available. The authors are grateful to Nick Dorrel for linguistic correction and to an anonymous referee for comments which improved the quality of the manuscript. This work was supported by the Research project AV0Z 5011922, Center for Neuroscience LC554 and by Academy of Sciences of the Czech Republic Grant (Information Society, 1ET400110401).

References

Abramowitz M, Stegun IA (1972) *Handbook of mathematical functions*, Dover, New York

Adrian ED (1928) *The basis of sensation*, W. W. Norton, New York

Beirlant J, Dudewicz EJ, Gyorfi L, van der Meulen EC (1997) Nonparametric entropy estimation: an overview, *Int. J. Math. Stat. Sci.*, 6, 17–39

Berger D, Pribram K, Wild H, Bridges C (1990) An analysis of neural spike train distributions: determinants of the response of visual cortex neurons to changes in orientation and spatial frequency, *Exp. Brain. Res.*, 80, 129–134

Berger D, Pribram K (1992) The relationship between the gabor elementary function and a stochastic model of interspike interval distribution in the responses of visual cortex neurons, *Biol. Cybern.*, 67, 191–194

Bershadskii A, Dremencov E, Fukayama D, Yadid G (2001) Probabilistic properties of neuron spiking time series obtained in vivo, *Eur. Phys. J. B*, 24, 409–413

Berry II MJ, Meister M (1998) Refractoriness and neural precision, *J. Neurosci*, 6, 2200–2211

Bhumbra GS, Inyushkin AN, Dyball REJ (2004) Assessment of spike activity in the supraoptic nucleus, *J. Neuroendocrin.*, 16, 390–397

Borst A, Theunissen FE (1999), Information theory and neural coding, *Nature neuroscience*, 2, 947–957

Buracas GT, Albright TD (1999) Gauging sensory representations in the brain, *Trends Neurosci.*, 22, 303–309

Chacron MJ, Longtin A, Maler L (2001) Negative interspike interval correlations increase the neuronal capacity for encoding time-dependent stimuli, *J. of Neurosci*, 21, 5328–5343

Chhikhara RS, Folks JL (1989) *The inverse Gaussian distribution: theory, methodology and applications*, Marcel Dekker, Inc, New York

Choi B, Kim K, Song SH (2004) Goodness-of-fit test for exponentiality based on Kullback-Leibler information, *Commun. Stat-Simul. C. (33)*, 2, 525–536

Christodoulou C, Bugmann G (2001) Coefficient of variation vs. mean interspike interval curves: what do they tell us about the brain?, *Neurocomputing*, 38–40, 1141–1149

Cover TM, Thomas JA (1991) *Elements of information theory*, John Wiley & sons

Inc., New York

Cox DR, Lewis PAW (1966) The Statistical Analysis of Series of Events, John Wiley & Sons, Inc., New York

DeWeese MR, Meister M (1999) How to measure the information gained from one symbol, *Network: Comput. Neural. Syst.*, 10, 325–340

Duchamp-Viret P, Duchamp A, Chaput MA (2003) Single olfactory sensory neurons simultaneously integrate the components of an odor mixture, *Eur. J. Neurosci.*, 10, 2690–2696

Duchamp-Viret P, Chaput MA, Kostal L, Lansky P, Rospars J-P (2005) Patterns of spontaneous activity in single rat olfactory receptor neurons are different in normally breathing and tracheotomized animals, *Journal of Neurobiology*, 65, 97–114

Ebrahimi N, Habibullah M, Soofi ES (1992) Testing exponentiality based on Kullback-Leibler Information, *J. Statist. Soc. B*, 3, 739–748

Gerstein G, Mandelbrot B (1964) Random walk models for the spike activity of a single neuron, *Biophys. J.*, 4, 41–68

Gerstner W, Kistler W (2002) *Spiking neuron models*, Cambridge, Cambridge University Press

Gibbons JD (1971) *Nonparametric Statistical Inference*, McGraw-Hill, Inc, New York

Han YMY, Chan YS, Lo KS, Wong TM (1998) Spontaneous activity and barosensitivity of the barosensitive neurons in the rostral ventrolateral medulla of hypertensive rats induced by transection of aortic depressor nerves, *Brain. Res.*, 813, 262–267

Hentall ID (2000) Interactions between brainstem and trigeminal neurons detected by cross-spectral analysis, *Neurosci.*, 96, 601–610

Holt GR, Softky WR, Koch C, Douglas RJ, (1996) Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons, *J. Neurophysiology*, 75, 1806–1814

Iynegar S, Liaom Q (1997) Modeling neural activity using the generalized inverse Gaussian distribution, *Biol. Cybern.*, 77, 289–295

Jaynes ET (1968) Prior probabilities, *IEEE Transactions on systems science and cybernetics*, 3, 227–241

Johnson DH, Glantz RM (2004) When does interval coding occur?, *Neurocomputing*, 58–60, 13–18

Kozachenko L F, Leonenko NN (1987) On statistical estimation of entropy of a random vector, *Probl. Inf. Trans.*, 23, 95–101

Lansky P, Rodriguez R, Sacerdote L (2004) Mean instantaneous firing frequency is always higher than the firing rate, *Neural Comp.*, 16, 477-489

Levine MW (1991) The distribution of the intervals between neural impulses in the maintained discharges of retinal ganglion cells, *Biol. Cybern.*, 65, 459–467

Lewis CD, Gebber GL, Larsen PD, Barman SM (2001) Long-term correlations in the spike trains of medullary sympathetic neurons, *J. Neurophysiol.*, 85, 1614–1622

Ling YB, He B (1993) Entropic Analysis of Biological Growth Models, *IEEE T. Bio-Med. Eng.*, 40, 48–54

van der Lubbe JCA (1997) *Information theory*, Cambridge University Press, Cambridge

Mandl G (1992) Coding for stimulus velocity by temporal patterning of spike discharges in visual cells of cat superior colliculus, *Vision Res.*, 33, 1451–1475

McKeegan DEF (2002) Spontaneous and odor evoked activity in single avian olfactory bulb neurons, *Brain Res.*, 929, 48–58

Miller EG, Fisher III JW (2003) ICA using space estimates of entropy, J of Machine

Learning Res, 4, 1271–1295

Nemenman I, Bialek W, de Ruyter van Steveninck RR (2004) Entropy and information in spike trains: progress on the sampling problem, *Phys. Rev. E*, 69, 056111-1–056111-6

Paninski L (2003) Estimation of entropy and mutual information, *Neural Comp.*, 15, 1191–1253

Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes, *Biophysical Journal*, 4, 391–418

Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) *Spikes: exploring the neural code*, Cambridge, MIT Press

Reeke NR, Coop AD (2004) Estimating the temporal entropy of neuronal discharge, *Neural Comp.*, 16, 941–970

Rospars J-P, Lansky P, Vaillant J, Duchamp-Viret P, Duchamp A (1994) Spontaneous activity of first- and second-order neurons in the frog olfactory system, *Brain Res.*, 662, 31–44

Rozell CJ, Johnson DH, Glantz RM (2004) Measuring information transfer in the spike generator of crayfish sustaining fibers, *Biol. Cybern.*, 90, 89–97

Ruskin DN, Bergstrom D A, Walters JR (2002) Nigrostriatal lesion and dopamine agonists affect firing patterns of rodent entopeduncular nucleus neurons, *J. Neurophysiol.*, 88, 487–496

Shannon C (1948) A mathematical theory of communication, *AT&T Tech. J.*, 27, 379–423

Shinomoto S, Yutaka S, Hiroshi O (2002) Recording site dependence of the neuronal spiking statistics, *Biosystems*, 67, 259–263

Shinomoto S, Shima K, Tanji J (2003) Differences in spiking patterns among cortical neurons, *Neural Computation*, 15, 2823–2842

Soofi ES, Ebrahimi N, Habibullah M (1995) Information Distinguishability with Application to Analysis of Failure Data, *Journal of the American Statistical Association*, 90, 57-668.

Stein RB (1967) The information capacity of nerve cells using a frequency code, *Biophys. J.*, 7, 797–826

Stevens CF, Zador A (1995) Information through a spiking neuron, *Adv. Neural Inf. Proc. Sys.*, 8, 75–81

Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and information in spike trains, *Phys. Rev. Let.*, 80, 197–200

Tarantola A (1994) The Inverse Problem Theory, Elsevier, Amsterdam

Tarantola A, Valette B (1982) Inverse Problems = Quest for Information, *J. Geophys.* 50, 159–170

Tsybakov AB, van der Meulen EC (1996) Root-n Consistent Estimators of entropy for densities with unbounded support, *Scand. J. Statist.*, 23, 75–83

Vasicek O (1976) A test for normality based on sample entropy, *J. Statist. Soc. B*, 38, 54–59

Zador A (1998) Impact of synaptic unreliability on the information transmitted by spiking neurons, *J. Neurophysiol.*, 79, 1219–1229