

Article

Nonparametric Estimation of Information-Based Measures of Statistical Dispersion

Lubomir Kostal and Ondrej Pokora *

Institute of Physiology, Academy of Sciences of the Czech Republic, Videnska 1083, 142 20 Prague, Czech Republic; E-Mail: kostal@biomed.cas.cz

* Author to whom correspondence should be addressed; E-Mail: pokora@math.muni.cz.

Received: 29 March 2012; in revised form: 20 June 2012 / Accepted: 4 July 2012 /

Published: 10 July 2012

Abstract: We address the problem of non-parametric estimation of the recently proposed measures of statistical dispersion of positive continuous random variables. The measures are based on the concepts of differential entropy and Fisher information and describe the “spread” or “variability” of the random variable from a different point of view than the ubiquitously used concept of standard deviation. The maximum penalized likelihood estimation of the probability density function proposed by Good and Gaskins is applied and a complete methodology of how to estimate the dispersion measures with a single algorithm is presented. We illustrate the approach on three standard statistical models describing neuronal activity.

Keywords: statistical dispersion; entropy; Fisher information; nonparametric density estimation

1. Introduction

Frequently, the dispersion (variability) of measured data needs to be described. Although standard deviation is used ubiquitously for quantification of variability, such approach has limitations. The dispersion of the probability distribution can be understood in different points of view: as “spread” with respect to the expected value, “evenness” (“randomness”) or “smoothness”. For example highly variable data might not be random at all if it consists only of “extremely small” and “extremely large”

measurements. Although the probability density function or its estimate provides a complete view, quantitative methods are needed in order to compare different models or experimental results.

In a series of recent studies [1,2] we proposed and justified alternative measures of dispersion. The effort was inspired by various information-based measures of signal regularity or randomness and their interpretations that have gained significant popularity in various branches of science [3–9]. For convenience, in what follows we discuss only the *relative* dispersion coefficients (*i.e.*, the data or the probability density function is first normalized to unit mean). Besides the coefficient of variation, c_v , which is the relative dispersion measure based on standard deviation, we employ the entropy-based dispersion coefficient c_h , and the Fisher information-based coefficient c_J . The difference between these coefficients lies in the fact that the Fisher information-based coefficient, c_J , describes how “smooth” is the distribution and it is sensitive to the modes of the probability density, while the entropy-based coefficient, c_h , describes how “even” it is, hence being sensitive to the overall spread of the probability density over the entire support. Since multimodal densities can be more evenly spread than unimodal ones, the behavior of c_h cannot be generally deduced from c_J (and vice versa).

If a complete description of the data is available, *i.e.*, the probability density function is known, the values of the above mentioned dispersion coefficients can be calculated analytically or numerically. However, the estimation of these coefficients from data is more problematic, and so far we employed either the parametric approach [2] or non-parametric estimation of c_h based on the popular Vasicek’s estimator of differential entropy [10,11]. The goal of this paper is to provide a self-contained method of non-parametric estimation. We describe a method that can be used to estimate both c_h and c_J as a result of a single procedure.

2. Methods

2.1. Dispersion Coefficients

We briefly review the proposed dispersion measures, for more details see [2]. Let T be a continuous positive random variable with probability density function $f(t)$ defined on $[0, \infty)$. By far, the most common measure of dispersion of T is the standard deviation, σ , defined as the square root of the second central moment of the distribution. The corresponding *relative* dispersion measure is known as the coefficient of variation, c_v ,

$$c_v = \frac{\sigma}{E(T)} \quad (1)$$

where $E(T)$ is the mean value of T .

The entropy-based dispersion coefficient is defined as

$$c_h = \frac{\exp[h(T) - 1]}{E(T)} \quad (2)$$

where $h(T)$ is the differential entropy [12],

$$h(T) = - \int_0^{\infty} f(t) \ln f(t) dt \quad (3)$$

The numerator in (2), $\sigma_h = \exp[h(T) - 1]$, is the entropy-based dispersion, “analogous” to the notion of standard deviation σ . The interpretation of σ_h relies on the asymptotic equipartition property theorem [12]. Informally, the theorem states that almost any sequence of n realizations of the random variable T comes from a rather small subset (the typical set) in the n -dimensional space of all possible values. The volume of this subset is approximately $\sigma_h^n = \exp[nh(T)]$, and the volume is bigger for those random variables, which generate more diverse (or unpredictable) realizations. The values of σ_h and c_h quantify how “evenly” is the probability distributed over the entire support. From this point of view, σ_h is more appropriate than σ to describe the randomness of outcomes generated by the random variable T .

The Fisher information-based dispersion coefficient, c_J , is defined as

$$c_J = \frac{1}{E(T) \sqrt{J(T)}} \tag{4}$$

where

$$J(T) = \int_0^\infty \left[\frac{\partial \ln f(t)}{\partial t} \right]^2 f(t) dt \tag{5}$$

The Fisher information is traditionally interpreted by means of the Cramer–Rao bound, *i.e.*, the value of $1/J(T)$ is the lower bound on the error of any unbiased estimator of a location parameter of the distribution. Due to the derivative in (5), certain regularity conditions are required on $f(t)$ in order for $J(T)$ to be interpreted according to the Cramer–Rao bound, namely continuous differentiability for all $t > 0$ and $f(0) = f'(0) = 0$, [13]. However, the integral in (5) exists and is finite also for, e.g., an exponential distribution. Any locally steep slope or the presence of modes in the shape of $f(t)$ increases $J(T)$, [4].

2.2. Methods of Non-Parametric Estimation

In the following we assume that $\{t_1, t_2, \dots, t_N\}$ are N independent realizations of the random variable $T \sim f(t)$, defined on $[0, \infty)$.

The c_v is most often estimated by the ratio of sample standard deviation to sample mean, however, the estimate may be considerably biased, [14].

Estimation of c_h relies on the estimate \hat{h} of the differential entropy $h(T)$, as follows from (2). The problem of differential entropy estimation is well exploited in literature [11,15,16]. It is preferable to avoid estimations based on data binning (histograms), because discretization affects the results. The most popular approaches are represented by the class of Kozachenko–Leonenko estimators [17,18] and the Vasicek’s estimator [10]. Our experience shows that the simple Vasicek’s estimator gives good results on a wide range of data [19–21], thus in this paper we employ it for the sake of comparison with the estimation method described further below. Given the ranked observations $t_{[1]} < t_{[2]} < \dots < t_{[N]}$, the Vasicek’s estimator \hat{h} is defined as

$$\hat{h} = \frac{1}{N} \sum_{i=1}^N \ln (t_{[i+m]} - t_{[i-m]}) + \ln \frac{N}{2m} + \varphi_{\text{bias}} \tag{6}$$

where $t_{[i+m]} = t_{[m]}$ for $i + m > N$ and $t_{[i-m]} = t_{[1]}$ for $i - m < 1$. The integer parameter $m < N/2$ is set prior to the calculation, roughly one may set m to be the integer part of \sqrt{N} . The bias-correcting factor is

$$\varphi_{\text{bias}} = \ln \frac{2m}{N} + \left(1 - \frac{2m}{N}\right) \Psi(2m) + \Psi(N + 1) - \frac{2}{N} \sum_{i=1}^m \Psi(i + m - 1) \tag{7}$$

where $\Psi(z) = \frac{d}{dz} \ln \Gamma(z)$ denotes the digamma function, [22].

The estimation of c_J requires estimation of $J(T)$, and it is more problematic and as far as we know no “standard” algorithms have been proposed.

Huber [23] showed that there exists a unique interpolation of the empirical cumulative distribution function such that the resulting Fisher information is maximized. In theory, one may use this method to estimate $J(T)$. Unfortunately, it is very complicated and probably not feasible even for moderate values of N .

Kernel density estimators are widely used for estimation of the density. They can be successfully used for the calculation of the differential entropy too [24,25]. But, according to our experience, they are unsuitable for estimation of the Fisher information, mainly due to the inability to control the “smoothness” of the ratio f'/f , which is the principal term in the integral (5). Inappropriate choice of the kernel and the bandwidth leads to many local extremes of the kernel density estimator which substantially increases the Fisher information. Theoretically, the kernel should be “optimal” in the sense of minimizing the mean square error of the ratio f'/f , which is difficult to derive in a closed form, generally. One may also derive Vasicek-like estimator of $J(T)$, based on the empirical cumulative distribution function and considering differences of higher order, however, we discovered that this approach is numerically unstable.

We found that the maximum penalized likelihood (MPL) method of Good and Gaskins [26] for probability density estimation offers a possible solution. The idea is to represent the whole probability density function using a suitable orthogonal base of functions (Hermite functions), and the “best” estimate of $f(t)$ is obtained by solving a set of equations for the base coefficients given the sample $\{t_1, \dots, t_N\}$.

In order to proceed, we first log-transform and normalize the sample,

$$x_i = \frac{\ln t_i - \mu}{\sqrt{2\sigma^2}} \tag{8}$$

where

$$\mu = \frac{1}{N} \sum_{i=1}^N \ln t_i \quad \text{and} \quad \sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (\ln t_i - \mu)^2 \tag{9}$$

The idea here is to obtain a new random variable X , $X = (\ln T - \mu)/\sqrt{2\sigma^2}$, with a more Gaussian-like density shape unlike the shape of the distribution of T , which is usually highly skewed. Furthermore, the limitation of the support of T to the real half-line may cause numerical difficulties. The probability density function of X is expressed by the real probability amplitude $c(x)$, so that $X \sim c^2(x)$. For the purpose of numerical implementation, $c(x)$ is represented by the first r Hermite functions, [22], as

$$c(x) = \sum_{m=0}^{r-1} c_m \psi_m(x) \tag{10}$$

where

$$\psi_m(x) = [\sqrt{\pi} 2^m m!]^{-1/2} \exp(-x^2/2) H_m(x) \tag{11}$$

and $H_m(x)$ denotes the Hermite polynomial $H_m(x) = (-1)^m \exp(x^2) \frac{d^m}{dx^m} \exp(-x^2)$.

The goal is to find such $c = c(x)$, for which the score, ω , is maximized,

$$\omega(c; x_1, \dots, x_N) = l(c; x_1, \dots, x_N) - \Phi(c) \tag{12}$$

subject to constraint (ensuring that $c^2(x)$ is a probability density function)

$$\sum_{m=0}^{r-1} c_m^2 = 1 \tag{13}$$

The term $l(c; x_1, \dots, x_N)$ in (12) is the log-likelihood function,

$$l(c; x_1, \dots, x_N) = 2 \sum_{i=1}^N \ln c(x_i) \tag{14}$$

and Φ is the *roughness penalty* proposed by [26], controlling the smoothness of the density of X (and hence the smoothness of $f(t)$),

$$\Phi(c; \alpha, \beta) = 4\alpha \int_{-\infty}^{\infty} c'(x)^2 dx + \beta \int_{-\infty}^{\infty} c''(x)^2 dx \tag{15}$$

The two nonnegative parameters α and β ($\alpha + \beta > 0$) should be set prior to the calculation. Note that the first term in (15) is equal to α -times the Fisher information $J(X)$.

The system of equations for c_i 's which maximize (12) can be written as [26]

$$\begin{aligned} & \frac{1}{2}\beta\sqrt{k(k-1)(k-2)(k-3)} c_{k-4} - \sqrt{k(k-1)} [4\alpha + (2k-1)\beta] c_{k-2} + \\ & + [2\lambda + 8\alpha k + \beta k(k+1)] c_k - \sqrt{(k+1)(k+2)} [4\alpha + (2k+3)\beta] c_{k+2} + \\ & + \frac{1}{2}\beta\sqrt{(k+1)(k+2)(k+3)(k+4)} c_{k+4} = 2 \sum_{i=1}^N \frac{\psi_k(x_i)}{\sum_{m=0}^{r-1} c_m \psi_m(x_i)} \end{aligned} \tag{16}$$

for $k = 0, 1, \dots, r - 1$. The system can be solved iteratively as follows. Initially $c_0 = 1$, $c_m = 0$ for $m = 1, 2, \dots$ and $\lambda = N$, which gives the approximation of X by a Gaussian random variable. Then, in each iteration step, current values of c_0, \dots, c_{r-1} are substituted into the right hand side of (16), and the system is solved as a linear system with unknown variables c_0, \dots, c_{r-1} appearing on the left hand side of (16). Once the linear system is solved, the coefficients are normalized to satisfy (13). The corresponding Lagrange multiplier is calculated as

$$\begin{aligned} \lambda = N + \alpha & \left\{ 2 - 4 \sum_{m=0}^{r-1} \left[(m + \frac{1}{2}) c_m^2 - \sqrt{(m+1)(m+2)} c_m c_{m+2} \right] \right\} + \\ & + \beta \left\{ \frac{3}{4} - \sum_{m=0}^{r-1} \left[\frac{3}{4} (2m^2 + 2m + 1) c_m^2 - (2m + 3) \sqrt{(m+1)(m+2)} c_m c_{m+2} + \right. \right. \\ & \left. \left. + \frac{1}{2} \sqrt{(m+1)(m+2)(m+3)(m+4)} c_m c_{m+4} \right] \right\} \end{aligned} \tag{17}$$

The $r + 1$ variables (r coefficients and the Lagrange multiplier λ) are computed iteratively. In accordance with [26], the algorithm is stopped when λ does not change within desired precision in subsequent iterations.

After the score-maximizing c_i 's are obtained, and thus the density of X is estimated, the estimates of dispersion coefficients in (2) and (4) are calculated from the following entropy and Fisher information estimators. The change of variables gives

$$\widehat{h(T)} = \widehat{h(X)} + \mu + \ln(\sqrt{2}\sigma) + \sqrt{2}\sigma\widehat{E(X)} \tag{18}$$

where $\widehat{h(X)} = -2 \int_{-\infty}^{\infty} c(x)^2 \ln|c(x)| dx$ is the estimator of the differential entropy of X , μ and σ are given by (9) and $\widehat{E(X)} = \int_{-\infty}^{\infty} xc(x)^2 dx$ is the estimated expected value of X . (Numerically, $\widehat{E(X)}$ is usually small and can be neglected. Its value is influenced by the number, r , of considered Hermite functions.)

The Fisher information estimator then follows, by an analogous change of variables, as

$$\widehat{J(T)} = \frac{\exp(-2\mu)}{\sigma^2} \int_{-\infty}^{\infty} \left[\sqrt{2}c'(x) - \sigma c(x) \right]^2 \exp(-2\sqrt{2}\sigma x) dx \tag{19}$$

The calculation requires the first derivative of $c(x)$ [22],

$$c'(x) = \frac{1}{\sqrt{2}} \sum_{m=1}^{r-1} c_m \sqrt{m} \psi_{m-1}(x) - \frac{1}{\sqrt{2}} \sum_{m=0}^{r-1} c_m \sqrt{m+1} \psi_{m+1}(x) \tag{20}$$

3. Results

Neurons communicate via the process of synaptic transmission, which is triggered by an electrical discharge called the *action potential* or *spike*. Since the time intervals between individual spikes are relatively large when compared to the spike duration, and since for any particular neuron the “shape” or character of a spike remains constant, the spikes are usually treated as point events in time. Spike train consists of times of spike occurrences $\tau_0, \tau_1, \dots, \tau_n$, equivalently described by a set of n interspike intervals (ISIs) $t_i = \tau_i - \tau_{i-1}$, $i = 1 \dots n$, and these ISIs are treated as independent realizations of the random variable $T \sim f(t)$. The probabilistic description of the spiking results from the fact that the positions of spikes cannot be predicted deterministically due to presence of intrinsic noise, only the probability that a spike occurs can be given [27–29]. In real neuronal data, however, the non-renewal property of the spike trains is often observed [30,31]. Taking the serial correlation of the ISIs as well as any other statistical dependence into account would result in the decrease of the entropy and hence of the value of c_h , see [12].

We compare exact and estimated values of the dispersion coefficients on three widely used statistical models of ISIs: gamma, inverse Gaussian and lognormal. Since only the relative coefficients are discussed, c_v, c_h, c_J , we parameterize each distributions by its c_v while keeping $E(T) = 1$.

The gamma distribution is one of the most frequent statistical descriptors of ISIs used in analysis of experimental data [32]. Probability density function of gamma distribution can be written as

$$f_{\Gamma}(t) = (c_v^2)^{-1/c_v^2} [\Gamma(1/c_v^2)]^{-1} t^{1/c_v^2-1} \exp(-t/c_v^2) \tag{21}$$

where $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$ is the gamma function, [22]. The differential entropy is equal to, [2],

$$h_\Gamma = \frac{1}{c_v^2} + \ln c_v^2 + \ln \Gamma\left(\frac{1}{c_v^2}\right) + \left(1 - \frac{1}{c_v^2}\right) \Psi\left(\frac{1}{c_v^2}\right) \tag{22}$$

where $\Psi(z) = \frac{d}{dz} \ln \Gamma(z)$ denotes the digamma function, [22]. The Fisher information about the location parameter is

$$J_\Gamma = \frac{1}{c_v^2(1 - 2c_v^2)} \text{ for } 0 < c_v < \frac{1}{\sqrt{2}} \tag{23}$$

The Fisher information diverges for $c_v \geq 1/\sqrt{2}$, with the exception of $c_v = 1$ (corresponds to exponential distribution) where $J_\Gamma = 1$, but the Cramer–Rao based interpretation of J_Γ does not hold in this case since $f_\Gamma(0) \neq 0$, see [13] for details.

The inverse Gaussian distribution is often used for description of ISIs and fitted to experimental data. It arises as result of spiking activity of a stochastic variant of the perfect integrate-and-fire neuronal model [33]. The density of this distribution is

$$f_{IG}(t) = \sqrt{\frac{1}{2\pi c_v^2 t^3}} \exp\left[-\frac{(t-1)^2}{2c_v^2 t}\right] \tag{24}$$

The differential entropy of this distribution is equal to

$$h_{IG} = \frac{\ln(2\pi e c_v^2)}{2} - \frac{3}{\sqrt{2\pi c_v^2}} \exp(c_v^{-2}) K^{(1,0)}\left(\frac{1}{2}, c_v^{-2}\right) \tag{25}$$

where $K^{(1,0)}(\nu, z)$ denotes the first derivative of the modified Bessel function of the second kind [22], $K^{(1,0)}(\nu, z) = \frac{d}{d\nu} K(\nu, z)$. The Fisher information of the inverse Gaussian distribution results in

$$J_{IG} = \frac{21c_v^6 + 21c_v^4 + 9c_v^2 + 2}{2c_v^2} \tag{26}$$

The lognormal distribution is rarely presented as model distribution of ISIs. However, it represents a common descriptor in analysis of experimental data [33], with density

$$f_{LN}(t) = \sqrt{\frac{1}{2\pi t \ln(1 + c_v^2)}} \exp\left\{-\frac{[\ln(1 + c_v^2) + \ln t^2]^2}{8 \ln(1 + c_v^2)}\right\} \tag{27}$$

The differential entropy of this distribution is equal to

$$h_{LN} = \ln \sqrt{2\pi e \frac{\ln(1 + c_v^2)}{1 + c_v^2}} \tag{28}$$

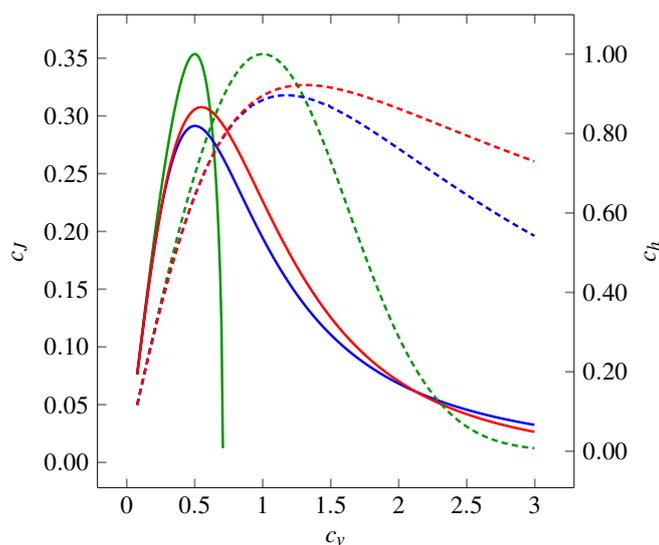
and the Fisher information is given by

$$J_{LN} = \frac{(1 + c_v^2)^3 [1 + \ln(1 + c_v^2)]}{\ln(1 + c_v^2)} \tag{29}$$

The theoretical values of the coefficients of c_h and c_J as functions of c_v are shown in Figure 1 for all the three distributions mentioned above. We see that the functions form hill-shaped curves with local

maxima achieved for different values of c_v . Asymptotically, c_h as well as c_J tend to zero for $c_v \rightarrow 0$ or $c_v \rightarrow \infty$ for the three models.

Figure 1. Variability represented by the coefficient of variation c_v , the entropy-based coefficient c_h (dashed curves, right-hand-side axis) and the Fisher information-based coefficient c_J (solid curves, left-hand-side axis) of three probability distributions: gamma (green curves), inverse Gaussian (blue curves) and lognormal (red curves). The entropy-based coefficient, c_h , expresses the evenness of the distribution. In dependency on c_v , it shows a maximum at $c_v = 1$ for gamma distribution (corresponds to exponential distribution) and between $c_v = 1$ and $c_v = 1.5$ for inverse Gaussian and lognormal distributions. For all the distributions holds $c_h \rightarrow 0$ as $c_v \rightarrow 0$ or $c_v \rightarrow \infty$. The Fisher information-based coefficient, c_J , grows as the distributions become “smoother”. The overall dependence on c_v shows a maximum around $c_v = 0.5$. Similarly to $c_h(c_v)$ dependencies, $c_h \rightarrow 0$ as $c_v \rightarrow 0$ or $c_v \rightarrow \infty$ (does not hold for gamma distribution, where c_J can be calculated only for $c_v < 1/\sqrt{2}$).



To explore the accuracy of the estimators of the coefficients c_h and c_J when the MPL estimations of the densities are employed, we did three separate simulation studies. All the simulations and calculations were performed in the free software package R [34]. For each model with the probability distribution (21), (24) and (27), respectively, the coefficient of variation, c_v , varied from 0.1 to 3.0 in steps of 0.1. One thousand samples, each consisting of 1000 random numbers (a common number of events in experimental records of neuronal firing), were taken for each value of c_v from the three distributions.

The MPL method was employed on each generated sample to estimate the density. We chose the number of the base functions equal to $r = 20$. The larger bases, $r = 50$ or $r = 100$, were examined too, with negligible differences in the estimation for the selected models. The values of the parameters for (15) were chosen as $\alpha = 3$ and $\beta = 4$, in accordance with the suggestion ([26], Appendix A).

The outcome of the simulation study for the gamma density (21) is presented in Figure 2, where the theoretical and estimated values of c_h and c_J for given c_v are plotted. In addition, the values of coefficient c_h calculated by Vasicek’s estimator are shown. We see that the MPL method results in precise estimation

of c_h for low values of c_v and in slightly overestimated values of c_h for $c_v > 0.7$. The Vasicek’s estimator gives slightly overestimated values too. Overall, the performances of Vasicek’s and MPL estimators are comparable. The maximum of the MPL estimator of c_h is achieved at the theoretical value, $c_v = 1$. We can see that the standard deviation of the c_h estimate is higher as c_v grows from zero. As $c_v > 1$, the standard deviation begins to decrease slowly.

We conclude that the MPL estimator of c_J is accurate for low c_v . For $c_v > 0.5$ it results in underestimated c_J and it tends to zero as $c_v \rightarrow 1/\sqrt{2}$ as well as the theoretical values. The high bias of c_J for high c_v is caused by inappropriate choice of the parameters α and β , which were kept fixed in accordance with the suggestion of [26]. Nevertheless, the main shape of the dependency on c_v remains and the MPL estimates of c_J achieves local maximum at $c_v \doteq 0.55$, which is slightly lower than the theoretical value.

Figure 2. The entropy-based variability coefficient c_h (panel a) and the Fisher information-based variability coefficient c_J (panel b), calculated nonparametrically from (18) and (19), respectively, for gamma distribution (24). The mean values (indicated by red discs) accompanied by the standard error (red error bars) are plotted in dependency on the coefficient of variation, c_v . The dashed lines are the theoretical curves. In panel a, the results obtained by estimation (6) are added (blue triangles indicate mean values and blue error bars stand for standard error). The results are based on 1000 trials of samples of size 1000 for each value of c_v .

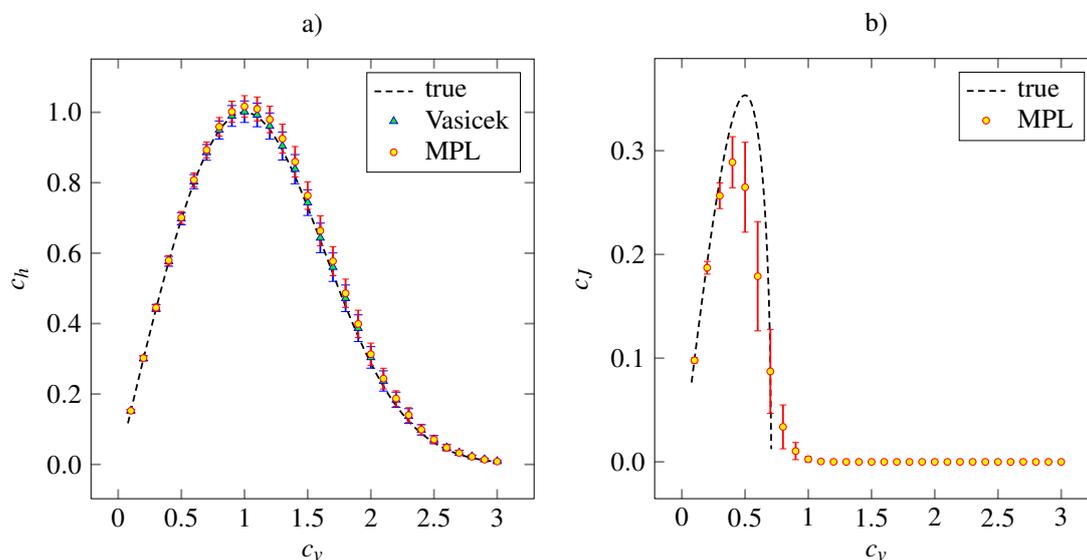
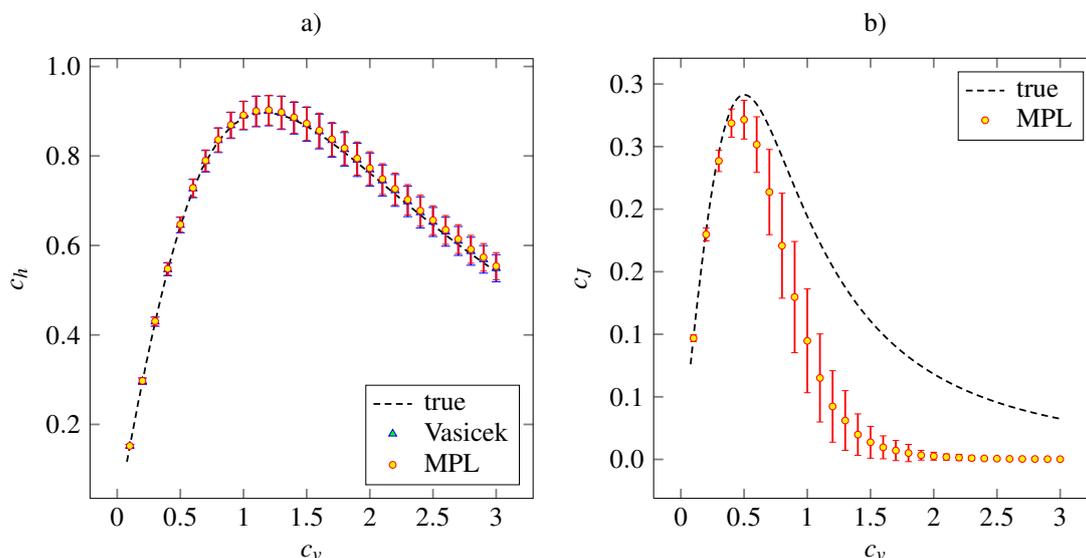


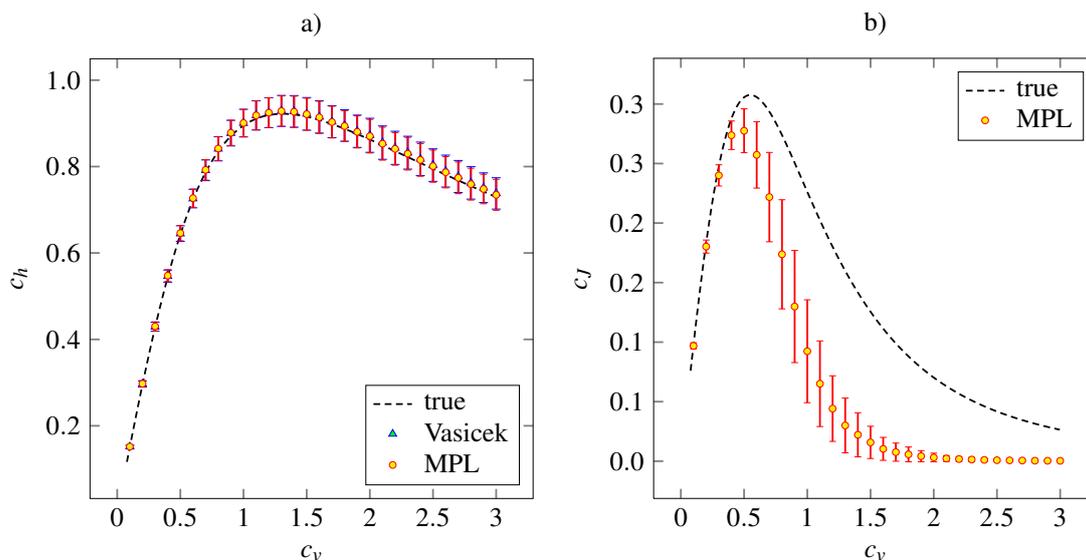
Figure 3 shows the corresponding results for the inverse Gaussian density. The Vasicek’s and MPL estimates of c_h are almost precise. Both the estimator of c_h based on MPL method and that based on the Vasicek’s entropy give the same mean values together with the same standard deviations. The MPL estimator of c_J gives accurate and precise results for $c_v < 0.5$. For higher c_v , the estimated value is lower than the true one. The maximum of estimated c_J is achieved at the same point, $c_v \doteq 0.5$. The asymptotical decrease of the estimated c_J to zero is faster than the true dependency is. By our experience, this can be improved by setting higher α and β in order to give higher impact to the roughness penalty (15).

Figure 3. Estimations of the variability coefficients c_h (panel a) and c_J (panel b) for inverse Gaussian distribution (24). The notation and the layout is the same as in Figure 2.



The results of the simulation study on the lognormal distribution with density (27) is plotted in Figure 4, together with the theoretical dependencies and the results for the Vasicek’s estimator of the entropy. Both the MPL estimators of c_h and c_J have qualitatively same accuracy and precision features as the analogous estimators for the inverse Gaussian model.

Figure 4. Estimations of the variability coefficients c_h (panel a) and c_J (panel b) for lognormal distribution (27). The notation and the layout is the same as in Figure 2.



4. Discussion and Conclusions

In proposing the dispersion measures based on entropy and Fisher information we were motivated by the difference between frequently mixed up notions of ISI variability and randomness, which, however, represent two different concepts [1]. The proposed measures have been so far successfully applied

mainly to examine differences between various neuronal activity regimes, obtained either by simulation of neuronal models or from experimental measurements [32]. There, the comparison of neuronal spiking activity under different conditions plays a key role in resolving the question of neuronal coding. However, the methodology is not specific to the research of neuronal coding; it is generally applicable whenever one needs to quantify some additional properties of positive continuous random data.

In this paper, we used the MPL method of Good and Gaskins [26] to estimate the dispersion coefficients nonparametrically from data. We found that the method performs comparably with the classical Vasicek's estimator [10] in the case of entropy-based dispersion.

The estimation of Fisher information-based dispersion is more complicated, but we found that the MPL method gives reasonable results. In fact, so far the MPL method is the best option for c_J estimation among the possibilities we tested (modified kernel methods, spline interpolation and approximation methods). The key parameters of the MPL method which affect the estimated value of $J(T)$ (and consequently of c_J) are the values of α and β in (15). In this paper we employed the suggestion of [26], however, we found that different setting may sometimes lead to dramatic improvement in the estimation of $J(T)$. We tested the performance of the estimation for sample sizes less than 1,000 and we found out that significant and systematic improvement resulting in low bias can be reached if α and β are allowed to depend somehow on the sample size. In this sense, the parameters play a similar role to the parameter m in the Vasicek's estimator (6). We are currently working on a systematic approach to determine α, β optimally, but the fine tuning of α and β is a difficult numerical task. Nevertheless, even without the fine-tuning, the performance of the entropy estimation is essentially the same as in the case of Vasicek's estimator.

The length of the neuronal record and hence the sample size is another issue related to the choice of these parameters. As emphasized, e.g., by [35,36], particularly short record can considerably modify the empirical distribution. This can be adjusted by the parameter values, choosing whether the distribution should fit the data or it should be rather robust.

Acknowledgements

This work was supported by the Institute of Physiology RVO: 67985823, the Centre for Neuroscience P304/12/G069 and by the Grant Agency of the Czech Republic projects P103/11/0282 and P103/12/P558. We thank Petr Lansky for helpful discussion.

References

1. Kostal, L.; Lansky, P.; Rospars, J.P. Review: Neuronal coding and spiking randomness. *Eur. J. Neurosci.* **2007**, *26*, 2693–2701.
2. Kostal, L.; Lansky, P.; Pokora, O. Variability measures of positive random variables. *PLoS One* **2011**, *6*, e21998.
3. Bercher, J.F.; Vignat, C. On minimum Fisher information distributions with restricted support and fixed variance. *Inf. Sci.* **2009**, *179*, 3832–3842.
4. Frieden, B.R. *Physics from Fisher Information: A Unification*; Cambridge University Press: New York, NY, USA, 1998.

5. Berger, A.L.; Della Pietra, V.J.; Della Pietra, S.A. A maximum entropy approach to natural language processing. *Comput. Linguist.* **1996**, *22*, 39–71.
6. Della Pietra, S.A.; Della Pietra, V.J.; Lafferty, J. Inducing features of random fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 380–393.
7. Di Crescenzo, A.; Longobardi, M. Entropy-based measure of uncertainty in past lifetime distributions. *J. Appl. Probab.* **2002**, *39*, 434–440.
8. Ramírez-Pacheco, J.; Torres-Román, D.; Rizo-Dominguez, L.; Trejo-Sanchez, J.; Manzano-Pinzón, F. Wavelet Fisher's information measure of $1/f^\alpha$ signals. *Entropy* **2011**, *13*, 1648–1663.
9. Pennini, F.; Ferri, G.; Plastino, A. Fisher information and semiclassical treatments. *Entropy* **2009**, *11*, 972–992.
10. Vasicek, O. A test for normality based on sample entropy. *J. Roy. Stat. Soc. B* **1976**, *38*, 54–59.
11. Tsybakov, A.B.; van der Meulen, E.C. Root-n consistent estimators of entropy for densities with unbounded support. *Scand. J. Statist.* **1994**, *23*, 75–83.
12. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley and Sons, Inc.: New York, NY, USA, 1991.
13. Pitman, E.J.G. *Some Basic Theory for Statistical Inference*; John Wiley and Sons, Inc.: New York, NY, USA, 1979.
14. Ditlevsen, S.; Lansky, P. Firing variability is higher than deduced from the empirical coefficient of variation. *Neural Comput.* **2011**, *23*, 1944–1966.
15. Beirlant, J.; Dudewicz, E.J.; Györfi, L.; van der Meulen, E.C. Nonparametric entropy estimation: An overview. *Int. J. Math. Stat. Sci.* **1997**, *6*, 17–39.
16. Gupta, M.; Srivastava, S. Parametric Bayesian estimation of differential entropy and relative entropy. *Entropy* **2010**, *12*, 818–843.
17. Kozachenko, L.F.; Leonenko, N.N. Sample estimate of the entropy of a random vector. *Prob. Inform. Trans.* **1987**, *23*, 95–101.
18. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 66138–16.
19. Esteban, M.D.; Castellanos, M.E.; Morales, D.; Vajda, I. Monte Carlo comparison of four normality tests using different entropy estimates. *Comm. Stat. Simulat. Comput.* **2001**, *30*, 761–785.
20. Miller, E.G.; Fisher, J.W. ICA using spacings estimates of entropy. *J. Mach. Learn. Res.* **2003**, *4*, 1271–1295.
21. Kostal, L.; Lansky, P. Similarity of interspike interval distributions and information gain in a stationary neuronal firing. *Biol. Cybern.* **2006**, *94*, 157–167.
22. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*; Dover: New York, NY, USA, 1965.
23. Huber, P.J. Fisher information and spline interpolation. *Ann. Stat.* **1974**, *2*, 1029–1033.
24. Hall, P.; Morton, S.C. On the estimation of the entropy. *Ann. Inst. Statist. Math.* **1993**, *45*, 69–88.
25. Eggermont, P.B.; LaRiccia, V.N. Best asymptotic normality of the kernel density entropy estimator for smooth densities. *IEEE Trans. Inform. Theor.* **1999**, *45*, 1321–1326.

26. Good, I.J.; Gaskins, R.A. Nonparametric roughness penalties for probability densities. *Biometrika* **1971**, *58*, 255–277.
27. Gerstner, W.; Kistler, W.M. *Spiking Neuron Models: Single Neurons, Populations, Plasticity*; Cambridge University Press: Cambridge, UK, 2002.
28. Stein, R.; Gossen, E.; Jones, K. Neuronal variability: Noise or part of the signal? *Nat. Rev. Neurosci.* **2005**, *6*, 389–397.
29. Shinomoto, S.; Shima, K.; Tanji, J. Differences in spiking patterns among cortical neurons. *Neural Comput.* **2003**, *15*, 2823–2842.
30. Avila-Akerberg, O.; Chacron, M.J. Nonrenewal spike train statistics: Cause and functional consequences on neural coding. *Exp. Brain Res.* **2011**, *210*, 353–371.
31. Farkhooi, F.; Strube, M.; Nawrot, M.P. Serial correlation in neural spike trains: Experimental evidence, stochastic modelling and single neuron variability. *Phys. Rev. E* **2009**, *79*, 021905.
32. Duchamp-Viret, P.; Kostal, L.; Chaput, M.; Lansky, P.; Rospars, J.P. Patterns of spontaneous activity in single rat olfactory receptor neurons are different in normally breathing and tracheotomized animals. *J. Neurobiol.* **2005**, *65*, 97–114.
33. Pouzat, C.; Chaffiol, A. Automatic spike train analysis and report generation. An implementation with R, R2HTML and STAR. *J. Neurosci. Meth.* **2009**, *181*, 119–144.
34. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2009.
35. Nawrot, M.P.; Boucsein, C.; Rodriguez-Molina, V.; Riehle, A.; Aertsen, A.; Rotter, S. Measurement of variability dynamics in cortical spike trains. *J. Neurosci. Meth.* **2008**, *169*, 374–390.
36. Pawlas, Z.; Lansky, P. Distribution of interspike intervals estimated from multiple spike trains observed in a short time window. *Phys. Rev. E* **2011**, *83*, 011910.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).