MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **11**, Number **1**, February **2014**

doi:10.3934/mbe.2014.11.63

pp. **63–80**

THE EFFECT OF INTERSPIKE INTERVAL STATISTICS ON THE INFORMATION GAIN UNDER THE RATE CODING HYPOTHESIS

Shinsuke Koyama

The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

Lubomir Kostal

Institute of Physiology, Academy of Sciences of the Czech Republic
Videnska 1083, 14220 Prague, Czech Republic

Abstract. The question, how much information can be theoretically gained from variable neuronal firing rate with respect to constant average firing rate is investigated. We employ the statistical concept of information based on the Kullback-Leibler divergence, and assume rate-modulated renewal processes as a model of spike trains. We show that if the firing rate variation is sufficiently small and slow (with respect to the mean interspike interval), the information gain can be expressed by the Fisher information. Furthermore, under certain assumptions, the smallest possible information gain is provided by gamma-distributed interspike intervals. The methodology is illustrated and discussed on several different statistical models of neuronal activity.

1. **Introduction.** Since the classical works of [2, 53, 43], the challenges of understanding the principles of neuronal coding have attracted an increasing number of scientists from different fields. It is generally accepted that neurons communicate using series of action potentials (spike trains) via chemical and electrical synapses. Currently, there are two main hypotheses that describe the representation of information in neuronal signal. In the first, denoted as the *rate* (or frequency) coding hypothesis, information is represented by the neuronal firing rate. In the second hypothesis, denoted as the *temporal* coding, features of the spiking activity beyond the firing rate are employed [54]. In this paper we are concerned only with the rate coding point of view.

A fundamental mathematical framework for the theoretical approach to the problem of information processing in neuronal systems is provided by information theory [52], e.g., in the works by [10, 9, 53] and by the statistical estimation theory [32, 40], e.g., in [24, 22, 23]. Along the latter approach, Koyama (2013) recently introduced the Kullback-Leibler (KL) divergence for rate-modulated renewal processes to investigate how much information on time-varying firing rates is carried by spike trains. It was shown that the KL divergence determines a lower bound for detectability of rate fluctuations with a Bayesian rate estimator [37]. It was also found that the information, as well as the lower bound, could significantly depend on the dispersion

2010 *Mathematics Subject Classification.* Primary: 60G55, 62P10; Secondary: 94A17.

Key words and phrases. Fisher information, Kullback-Leibler divergence, rate-modulated renewal processes, neural spike trains, neural coding.

The first author is supported by JSPS KAKENHI grant 24700287.

properties of neuronal firing, the effect of which, however, has not been investigated systematically.

In this paper, we employ the approximation of KL divergence in terms of Fisher information, which essentially relates the *information gain* and the *dispersion* of neuronal firing in a single formula. Consequently, we investigate the effect of the interspike interval (ISI) distribution dispersion on the overall encoding performance. In particular, we show that among all scale-family ISI distributions that share the same coefficient of variation, the gamma distribution attains the minimum information gain. We also illustrate typical behavior of the Fisher information by using generalized inverse Gaussian and lognormal distributions.

2. Methods.

In this manuscript we employ the general concept of *information* arising in statistics and introduced by [39], following the works of [52], [20] and [50]. In particular, assume that H_1 is the hypothesis that the random variable X follows probability distribution $f_1(x)$, and let H_2 denote the hypothesis $X \sim f_2(x)$. The *information* contained (on average) in the observation $X = x$ for *discrimination* in favor of H_1 against H_2 is defined as the Kullback-Leibler (KL) divergence $D(f_1||f_2)$ (see [39] for more details)

$$D(f_1||f_2) = \int_X f_1(x) \ln \frac{f_1(x)}{f_2(x)}\, dx. \tag{1}$$

The integral defining $D(f_1||f_2)$ always exists (although it may be infinite), and "$0 = 0 \ln 0$" as follows by taking the limits. The units of information are either "bits" (for the base-2 logarithm) or "nats" for the natural logarithm. The definition of information in Eq. (1) is apparently distinct from the Shannon's measures of *self-information* and *mutual information*, although there is no contradiction as can be demonstrated by a particular choice of H_1, H_2; the mutual information, for example, is obtained by taking H_1 and H_2 to be joint density and the product of the marginal densities, respectively [39, pp. 7–8]. See also [18] for further details.

Another measure of "information", employed especially in the theory of statistical estimation of continuously varying parameters [45], is the Fisher information. Let $f(x;\theta)$ be a parametric family of probability densities, then

$$J(\theta|X) = \int_X \left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right]^2 f(x;\theta)\, dx, \tag{2}$$

is denoted as the Fisher information about parameter θ contained in a single observation of r.v. X, since $J(\theta|X)$ determines how well the value of θ can be estimated from observing X (roughly stated, see e.g., [56] for details). In particular, for any unbiased estimator $\hat{\theta}(X)$ of parameter θ holds $Var(\hat{\theta}(X)) \geq 1/J(\theta|X)$ provided that $f(x;\theta)$ satisfies certain technical conditions [45]. The Fisher information has been used for measuring the encoding accuracy in the context of neural coding [8, 51, 57].

Eqns. (1) and (2) differ obviously, nonetheless, many fundamental relationships between these information measures have been found over the years, thus showing the depth of correspondence between information theory, theory of point estimation and statistics in general, see e.g., [5, 10, 33, 39, 45, 49] among others. An asymptotic relation between Fisher information and KL divergence is also employed in this paper.

2.1. **KL divergence for rate-modulated spike trains.** We consider a rate-modulated renewal process as a model of spike trains. Let $N(t)$ be the number of spikes that have already occurred at time t. A point process is generally defined by the conditional intensity function:

$$r(t; H(t)) = \lim_{\Delta t \to 0} \frac{P(N(t+\Delta t) - N(t) = 1; H(t))}{\Delta t}, \tag{3}$$

where $H(t)$ represent the history of spikes up to time t. Using this, the probability density of a sequence of spikes $\{t_i\} := \{t_1, \dots, t_n\}$ in an interval $[0, T]$ is expressed as [16, 31]

$$p(\{t_i\}) = \exp\left[-\int_0^T r(t; H(t)) dt \right] \prod_{i=1}^{n} r(t_i; H(t_i)). \tag{4}$$

Here, the exponential factor describes the probability of no spikes between the spike times. (Intuitively, this arises from the product of $1 - r(t; H(t))\Delta t$ as Δt goes to 0 and the number of terms in the product goes to infinity.)

Consider a renewal process whose ISI density is $f(x)$ with unit mean. The conditional intensity, or hazard function, of this process is given by

$$r(s; s_*) = \frac{f(s - s_*)}{1 - \int_{s_*}^{s} f(u - s_*) du}, \tag{5}$$

where s_* is the spike time preceding s.

A renewal process with arbitrary mean firing rate can be generated from Eq. (5) by rescaling the time axis. First, consider that the mean firing rate is given by a constant μ. The conditional intensity function is, then, obtained by simply rescaling the time $t = s/\mu$ as

$$r(t; t_*, \mu) = \frac{\mu f(\mu(t - t_*))}{1 - \int_{t_*}^{t} \mu f(\mu(u - t_*)) du}. \tag{6}$$

Substituting Eq. (6) into Eq. (4), the probability density of a spike train $\{t_i\}$ in the interval $[0, T]$ is obtained as

$$\begin{aligned} p(\{t_i\}; \mu) & \\ &= \exp\left[-\int_0^T r(t; t_{N(t)}, \mu) dt \right] \prod_{i=1}^{n} r(t_i; t_{i-1}, \mu) \\ &= p(\{t_i\}; \mu) = p_1(t_1; \mu) \cdot \prod_{i=2}^{n} \mu f(\mu(t_i - t_{i-1})) \cdot P((t_n, T]; \mu), \end{aligned} \tag{7}$$

where $p_1(t_1; \mu)$ is the probability density of the first spike occurring at t_1, and $P((t_n, T]; \mu)$ is the probability of no spikes being observed on $(t_n, T]$. The second line of Eq. (7) is derived in Appendix A.

This transformation can be generalized to time-dependent firing rate [3, 4, 15, 38, 44, 47]. Consider next that the firing rate $\lambda(t)$ is given as a function of t. By defining the transformation $\Lambda(t) := \int_0^t \lambda(u) du$ and rescaling the time $t = \Lambda^{-1}(s)$ (Figure 1), the conditional intensity function is obtained from Eq. (5) as

$$r(t; t_*, \{\lambda(t)\}) = \frac{\lambda(t) f(\Lambda(t) - \Lambda(t_*))}{1 - \int_{t_*}^{t} \lambda(u) f(\Lambda(u) - \Lambda(t_*)) du}. \tag{8}$$

In the same way as Eq. (7), the probability density of $\{t_i\}$ is obtained as

$$\begin{aligned} &p(\{t_i\};\{\lambda(t)\}) \\ &= p_1(t_1;\{\lambda(t)\}) \cdot \prod_{i=2}^{n} \lambda(t_i) f(\Lambda(t_i) - \Lambda(t_{i-1})) \cdot P((t_n, T];\{\lambda(t)\}). \end{aligned} \tag{9}$$

This rate-modulated renewal process is a generalization of both the inhomogeneous Poisson process (if $f(x)$ is the exponential density) and the renewal process (if $\lambda(t)$ is constant).

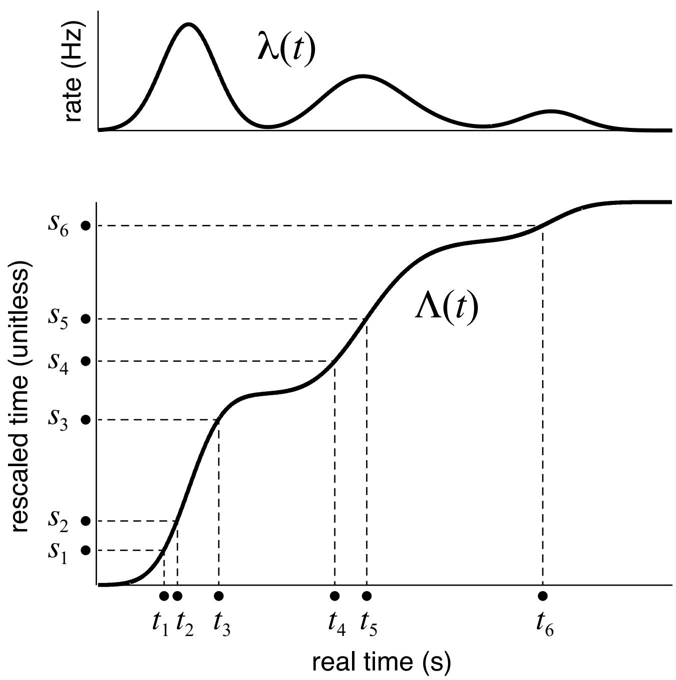

FIGURE 1. The time-rescaling transformation. A renewal spike train $\{s_i\}$, whose ISI distribution is given by $f(s_i - s_{i-1})$, is transformed to $\{t_i\}$ via $t_i = \Lambda^{-1}(s_i)$. Accordingly, the transformed spike train $\{t_i\}$ has the instantaneous firing rate $\lambda(t)$.

Figure 2 depicts four probability densities $f(x)$ with $C_V = 1$ (left) and sample spike trains derived from the rate-modulated renewal processes (9) with the sinusoidal rate process (17) (right). These models will be used for examples in the following section.

To measure information gained by the fluctuating rate $\lambda(t)$, Koyama (2013) introduced the KL divergence between the two probability densities of spike trains (7) and (9) [37]:

$$\begin{aligned} D[p(\cdot;\{\lambda(t)\})||p(\cdot;\mu)] &= \frac{1}{\mu T}\sum_{n=0}^{\infty}\int_0^T\int_{t_1}^T\cdots\int_{t_{n-1}}^T p(\{t_i\};\{\lambda(t)\}) \\ &\quad\times \ln\frac{p(\{t_i\};\{\lambda(t)\})}{p(\{t_i\};\mu)}dt_1 dt_2\cdots dt_n. \end{aligned} \tag{10}$$

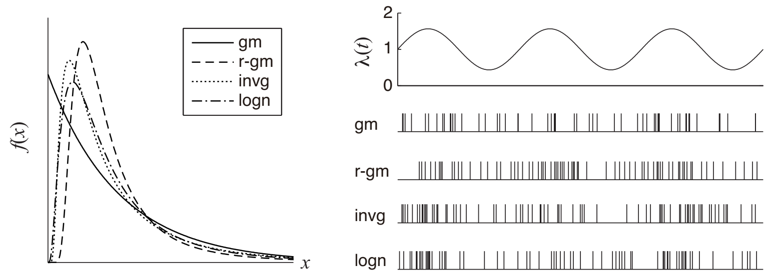

FIGURE 2. Left: the probability density functions of gamma ('gm'), reciprocal gamma ('r-gm'), inverse Gaussian ('invg') and lognormal ('logn') distributions for $C_V = 1$. Right: sample spike trains derived from the rate-modulated renewal processes with sinusoidal rate variation (17), whose parameters are $\mu = 1$, $\sigma = 0.4$ and $\tau = 20$.

Note that the unit is "nats/spike". Eq. (10) is generally difficult to be analyzed because it is a functional of $\lambda(t)$ and $f(x)$, the effects of which are not separated from each other in the formula, and contains a high-dimensional integration. As shown in the section 3.1, it can be approximated to more tractable form if temporal variation of the rate is sufficiently slow and small.

For the case of N identically and independent trials, or population of N neurons, which may be more practical situation in estimating the firing rate, the KL divergence is multiplied by the factor N due to the additivity of information gain. Hence, it is enough to consider the case of single spike trains.

2.2. **Fisher information for scale family of probability densities.** We consider the Fisher information for a scale family of probability densities $p(t;\lambda)$ that is generated from a probability density $f(x)$ as

$$p(t;\lambda) = \lambda f(\lambda t), \tag{11}$$

where $\lambda > 0$ is a scale constant. $p(t;\lambda)$ has the same shape as $f(x)$ but a different scale. Here, $f(x)$ is a distribution with unit mean, and λ is interpreted as the mean firing rate.

Inserting Eq. (11) into Eq. (2), it is found that the Fisher information about the scale parameter has the scaling property:

$$J(\lambda|T) = \frac{1}{\lambda^2} J(1|T), \tag{12}$$

where $J(1|T)$ is dimensionless and is uniquely determined by the shape of the density $f(x)$. Therefore, $J(1|X)$ can be interpreted as a kind of "dispersion" measure of random variable $X \sim f(x)$. In the following, we use a notation $I[f] := J(1|T)$ to indicate that it is a functional of $f(x)$. $I[f]$ is expressed as

$$\begin{aligned} I[f] &= \int_0^\infty \left[1 + x\frac{\partial \ln f(x)}{\partial x}\right]^2 f(x)dx \\ &= 1 - \int_0^\infty x^2 \frac{\partial^2 \ln f(x)}{\partial x^2} f(x)dx, \end{aligned} \tag{13}$$

where the second line is obtained by integration by parts.

3. **Results.** We first show that the KL divergence (10) can be approximated by the Fisher information (13) for slow and small rate variation. From this point of view, all the properties of the Fisher information play important role. In particular, we show that the minimum of the Fisher information $I[f]$, given the coefficient of variation, is achieved by the gamma distribution. Typical behavior of the Fisher information is illustrated by using generalized inverse Gaussian and lognormal distributions.

3.1. **Relation between the KL and Fisher information.** Let $\mu = \frac{1}{T}\int_0^T \lambda(t)dt$ and $\sigma^2 = \frac{1}{T}\int_0^T [\lambda(t)-\mu]^2 dt$ be the mean of and variance of $\lambda(t)$. We also let τ be the characteristic time scale of $\lambda(t)$[1]. In the limit of $T \to \infty$ and under the conditions of $\tau\mu \gg 1$ (slow variation) and $\sigma/\mu \ll 1$ (small variation), the KL divergence (10) is approximated as

$$D[p(\cdot;\{\lambda\})||p(\cdot;\mu)] = \frac{\sigma^2}{2\mu^2} I[f] + O(\frac{\sigma}{\tau\mu^2}). \tag{14}$$

The details are given in Appendix B.

It must be noticed that the rhs of (14) is analytically more tractable than Eq. (10), because i) $I[f]$ is defined for *single* ISIs while $D[p(\cdot;\{\lambda\})||p(\cdot;\mu)]$ is defined for whole spike trains, from which we can avoid performing the high-dimensional integration, and ii) $I[f]$ is separated from the effect of rate variation.

By using this approximation, we can study the effect of dispersion of firing, which is described by $I[f]$, on the information gain, separably from the effect of the rate variation. For this purpose, it is necessary to verify that the leading term in the rhs of Eq. (14) dominates the KL divergence under reasonable range of parameter values. This is done by computing the KL divergence (10) numerically, and comparing it with the approximation (14). The KL divergence is computed by simulating a large number of spikes $\{t_i\}$, and averaging the log likelihood ratio, $\ln p(\{t_i\};\{\lambda(t)\}) - \ln p(\{t_i\};\mu)$, over the sample spikes:

$$\begin{aligned} D[p(\cdot;\{\lambda\})||p(\cdot;\mu)] &= \frac{1}{n-1}\sum_{i=2}^{n}\big\{\ln\lambda(t_i) f(\Lambda(t_i)-\Lambda(t_{i-1})) \\ &\qquad - \ln \mu f(\mu(t_i - t_{i-1}))\big\}. \end{aligned} \tag{15}$$

The spike trains are simulated by generating spikes $\{s_i\}$ from the renewal process with ISI density $f(x)$, and then by applying the time-rescaling transformation, $t_i = \Lambda^{-1}(s_i)$.

In the simulation, we use the gamma density function with unit mean for $f(x)$:

$$f(x) = \frac{\kappa^\kappa x^{\kappa-1} e^{-\kappa x}}{\Gamma(\kappa)}, \tag{16}$$

where $\Gamma(\kappa) = \int_0^\infty x^{\kappa-1}e^{-x}dx$ is the gamma function, and a sinusoidally varying rate:

$$\lambda(t) = \mu + \sqrt{2}\sigma \sin\frac{2\pi}{\tau}t. \tag{17}$$

Figure 3 depicts the KL divergence, and its approximation (14) as a function of σ/μ. Note that the difference between these two quantities corresponds to the error term $O(\frac{\sigma}{\tau\mu^2})$. For instance, this error term is relatively small even if the rate fluctuation

[1] For a stochastic process $\lambda(t)$ whose correlation function is $\phi(u) = \lim_{T\to\infty}\frac{1}{T}\int_0^T[\lambda(t) - \mu][\lambda(t+u)-\mu]dt$, it is given by $\tau = \frac{1}{\phi(0)}\int_0^\infty \phi(u)du$. For a periodic process such as a sinusoidal function (17) used as an example below, τ is given by the period.

is relatively large ($\sigma/\mu \doteq 1$) for $\tau\mu = 3$, with which 3 spikes on average appear in a cycle of the sinusoidal firing rate. it is therefore confirmed that the expression of the KL divergence (14) with $I[f]$ provides a good approximation.

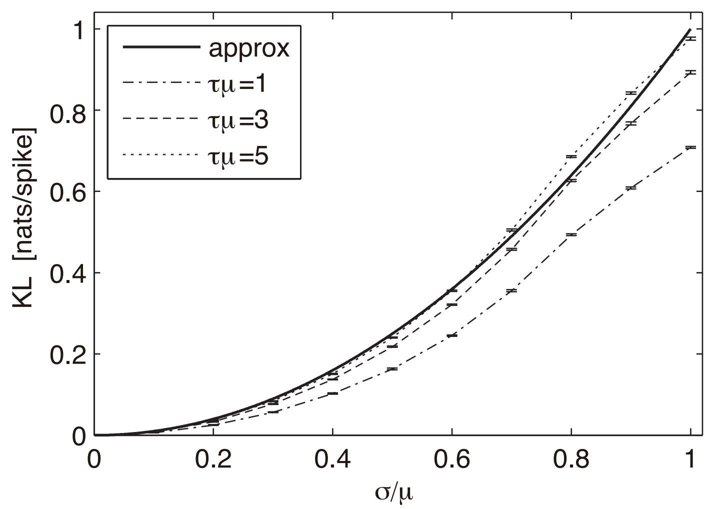

FIGURE 3. The information gain due to sinusoidally varying firing rate and gamma distribution of interspike intervals with shape parameter $\kappa = 2$. The value of σ/μ is the ratio of the amplitude to the mean of the driving sinusoid (the "fluctuation rate"). The number $\tau\mu$ determines the average number of spikes per sinusoidal period, here shown for $\tau\mu = 1$ (dash-dotted line), $\tau\mu = 3$ (dashed line) and $\tau\mu = 5$ (dotted line). The solid line is the approximation using the Fisher information. Note that the approximation holds well even for relatively high fluctuation rates ($\sigma/\mu \doteq 1$).

3.2. **The density function that achieves the minimum Fisher information.** Given the mean and variance, or the coefficient of variation C_V, the minimum of the Fisher information is achieved by the gamma distribution. This has been known in the literature [30]. In this paper, we provide another proof. The proof is done by showing that the Fisher information $I[f]$ is bounded as

$$I[f] \geq C_V^{-2}, \tag{18}$$

and then, showing that only the gamma distribution achieves the minimum of the Fisher information, C_V^{-2}. The details are given in Appendix C.

3.3. **Examples.** We use two specific models, generalized inverse Gaussian (GIG) family and lognormal distribution to illustrate behavior of the Fisher information (13).

The GIG family may be used as a statistical descriptor of ISIs for several reasons [28]. First, for certain values of the parameters of this family, its members are the first passage time distributions of certain diffusion processes to a constant boundary. It also contains gamma and inverse Gaussian distributions as sub-families, and thus unifies commonly used distributions. Gamma distribution is one of the most frequent statistical descriptors of ISIs [17, 41, 46]. Note that for $C_V = 1$ gamma distribution becomes exponential, resulting in the neuronal firing described by the Poisson process [55]. The inverse Gaussian distribution [11] is often used to describe neural activity [19, 55] and fitted to experimentally observed ISIs [41, 46]. This distribution results from the Wiener process with positive drift (the depolarization has a linear trend to the threshold) as the stochastic perfect integrate-and-fire neuronal model [55].

The lognormal distribution of ISIs, with some exceptions [6], is rarely presented as a result of a neuronal model. However, it represents quite a common descriptor in ISI data, see e.g. [46, 55] and references therein. Furthermore, a mixture of two lognormal distributions has been used recently [7] as an statistical ISI descriptor.

3.3.1. *Generalized inverse Gaussian distribution.* The GIG distribution has a density function:

$$f(x;w,\eta,a)=\frac{1}{2\eta^a K_a(w)}x^{a-1}\exp\left[-\frac{1}{2}\left(\frac{w}{\eta}x+\frac{w\eta}{x}\right)\right], \tag{19}$$

where $K_a(w)$ is the modified Bessel function of the second kind with index $a \in (-\infty,\infty)$ [1], and $\eta \geq 0$ and $w \geq 0$ represent a scale and concentration parameters, respectively. This becomes the inverse Gaussian distribution (for $a=-1/2$) and gamma distribution (16) (for $a=\kappa>0$, $w/(2\eta)=\kappa$ and $w\to 0$) as special cases. See [29] for statistical properties of the GIG distribution.

The mean and variance of X are, respectively, given by

$$E(X)=\eta\frac{K_{a+1}(w)}{K_a(w)}, \tag{20}$$

and

$$Var(X)=\eta^2\left[\frac{K_{a+2}(w)}{K_a(w)}-\left(\frac{K_{a+1}(w)}{K_a(w)}\right)^2\right], \tag{21}$$

from which the square of the coefficient of variation is obtained as

$$C_V^2=\frac{K_{a+2}(w)K_a(w)}{K_{a+1}(w)^2}-1. \tag{22}$$

$I[f]$ of the GIG is calculated as

$$\begin{aligned} I[f] &= 1-E\left[x^2\frac{\partial^2 \ln f(x;w,\eta,a)}{\partial x^2}\right] \\ &= a+w\eta E\left(\frac{1}{X}\right). \end{aligned} \tag{23}$$

Using Eq. (70), $I[f]$ is obtained as

$$I[f]=w\frac{K_{a+1}(w)+K_{a-1}(w)}{2K_a(w)}. \tag{24}$$

In the following, we analyze the behavior of Eqs. (22) and (24) in asymptotic cases.

1. Limit of $w\to 0$ for $a>0$. The gamma distribution is obtained in this limit. Using the asymptotic formula:

$$K_a(w)\sim\frac{\Gamma(a)2^{a-1}}{w^a}, \tag{25}$$

C_V^2 and $I[f]$ are evaluated as

$$C_V^2 \quad \sim \quad \frac{\Gamma(a+2)\Gamma(a)}{\Gamma(a+1)^2} - 1 = \frac{1}{a}, \tag{26}$$

and

$$I[f] \quad \sim \quad \frac{\Gamma(a+1)}{\Gamma(a)} + \frac{\Gamma(a-1)}{4\Gamma(a)} w^2 \to a = \frac{1}{C_V^2}, \tag{27}$$

which is consistent with the result in the section 3.2, that is, $I[f]$ of the gamma distribution corresponds to the minimum value $1/C_V^2$.

2. Limit of $w \to 0$ for $a < 0$. Using the asymptotic formula for this limit:

$$K_a(w) \sim 2^{-a-1}\Gamma(-a)w^a, \quad \text{for } a < 0 \tag{28}$$

and

$$K_0(w) \sim -\ln\frac{w}{2} - \gamma, \tag{29}$$

where γ is the Euler constant, C_V^2 and $I[f]$ are obtained as

$$C_V^2 \sim \begin{cases} \frac{\Gamma(-a-2)\Gamma(-a)}{\Gamma(-a-1)^2} - 1 & = & -\frac{1}{a+2}, & a < -2 \\ -2(\ln\frac{w}{2} + \gamma) & \to & \infty, & a = -2 \\ \frac{\Gamma(a+2)\Gamma(-a)}{\Gamma(-a-1)^2} \frac{2^{2a+4}}{w^{2a+4}} - 1 & \to & \infty, & -2 < a < -1 \\ \frac{1}{w^2(\ln\frac{w}{2}+\gamma)^2} - 1 & \to & \infty, & a = -1 \\ \frac{\Gamma(a+2)\Gamma(-a)}{\Gamma(a+1)^2} \frac{w^{2a}}{2^{2a}} - 1 & \to & \infty, & -1 < a < 0 \end{cases} \tag{30}$$

and

$$I[f] \sim \begin{cases} \frac{\Gamma(-a+1)+2^{-2}\Gamma(-a-1)w^2}{\Gamma(-a)} & \to & -a, & a < -1 \\ 1 - w^2\ln\frac{w}{2} - w^2\gamma & \to & 1 & a = -1 \\ \frac{\Gamma(-a+1)+2^{2a}\Gamma(a+1)w^{-2a}}{\Gamma(-a)} & \to & -a, & -1 < a < 0 \end{cases} \tag{31}$$

Particularly, $I[f]$ for $a < -2$ is expressed as

$$I[f] = \frac{1}{C_V^2} + 2. \tag{32}$$

It is worth noting that the GIG for $a < 0$, $w\eta = 2\beta$ and $w \to 0$ becomes the reciprocal gamma distribution:

$$f(x;\beta,a) = \frac{\beta^{-a}}{\Gamma(-a)} x^{a-1} \exp\left(-\frac{\beta}{x}\right). \tag{33}$$

3. Limit of $w \to \infty$. Using the asymptotic formula in this limit:

$$K_a(w) \sim \sqrt{\frac{\pi}{2}} e^{-w} w^{-\frac{1}{2}} \left(1 + \frac{4a^2-1}{8w}\right), \tag{34}$$

we obtain

$$C_V^2 \sim \frac{[1+O(w^{-1})][1+O(w^{-1})]}{[1+O(w^{-1})]^2} - 1 \to 0, \tag{35}$$

and

$$I[f] \sim w \cdot \frac{1+O(w^{-1})}{1+O(w^{-1})} \to \infty. \tag{36}$$

Thus, the Fisher information diverges as $C_V^2 \to 0$.

4. Upper and lower bounds of $I[f]$. As shown in section 3.2, the gamma distribution gives the lower bound of $I[f]$ (among all ISI densities), which is obtained in the limit $w \to 0$ for $a > 0$. On the other hand, the GIG becomes the reciprocal gamma distribution in $w \to 0$ for $a < 0$, whose C_V^2 and $I[f]$ are obtained as Eqs. (30) and (31). Taking into account that $I[f]$ and C_V^{-2} monotonically increase as the concentration parameter w is increased, and the mapping from (w, a) to $(C_V^{-2}, I[f])$ is one-to-one and smooth, it can be shown that the reciprocal gamma distribution (32) gives the upper bound of $I[f]$ among the GIG family.
5. Case of $a = -1/2$. The GIG becomes the inverse Gaussian distribution for this case, the density function of which is given by

$$f(x; w, \eta) = \sqrt{\frac{w\eta}{2\pi x^3}} \exp\left[-\frac{w(x-\eta)^2}{2\eta x} \right] \tag{37}$$

Using the formula:

$$K_{\frac{1}{2}}(w) = \sqrt{\frac{\pi}{2}} e^{-w} w^{-\frac{1}{2}}, \quad K_{\frac{3}{2}}(w) = \sqrt{\frac{\pi}{2}} e^{-w} w^{-\frac{1}{2}} \Big(1 + \frac{1}{w}\Big), \tag{38}$$

and $K_{-a}(w) = K_a(w)$, or calculating directly from Eq. (37), we obtain $C_V^2 = 1/w$ and

$$I[f] = w + \frac{1}{2} = \frac{1}{C_V^2} + \frac{1}{2}. \tag{39}$$

3.3.2. *Lognormal distribution.* Next, we examine the lognormal distribution. The density function has a form:

$$f(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right]. \tag{40}$$

The mean and variance are given by $E(X) = \exp(\mu + \frac{\sigma^2}{2})$ and $Var(X) = (e^{\sigma^2} - 1)\exp(2\mu + \sigma^2)$, respectively. Using Eq. (13), $I[f]$ is obtained as

$$I[f] = \frac{1}{\sigma^2} = \frac{1}{\ln(C_V^2 + 1)}. \tag{41}$$

$I[f]$ is a monotonically decreasing function of C_V. Particularly, $I[f] \to \infty$ as $C_V \to 0$, and $I[f] \to 0$ as $C_V \to \infty$.

3.3.3. *Summary of the Fisher information.* Figure 4 summarizes the Fisher information $I[f]$ of the GIG family (gray region) and lognormal distribution (dash-dotted line). The gamma distribution (solid line) and its reciprocal (dashed line), respectively, give the lower and upper bounds of $I[f]$ of the GIG family.

The Fisher information generally decreases as C_V is increased. It, however, does not necessarily converge to zero as $C_V \to \infty$: for the GIG, the Fisher information converges to the finite value $I[f] \to -a$ for $-2 \le a < 0$.

As seen in this figure, the Fisher information can significantly differ among the distributions even if they share the same value of C_V. For $C_V = 1$, $I[f]$ of gamma, reciprocal gamma, inverse Gaussian and lognormal distributions are 1, 3, 1.5, and 1.44, respectively.

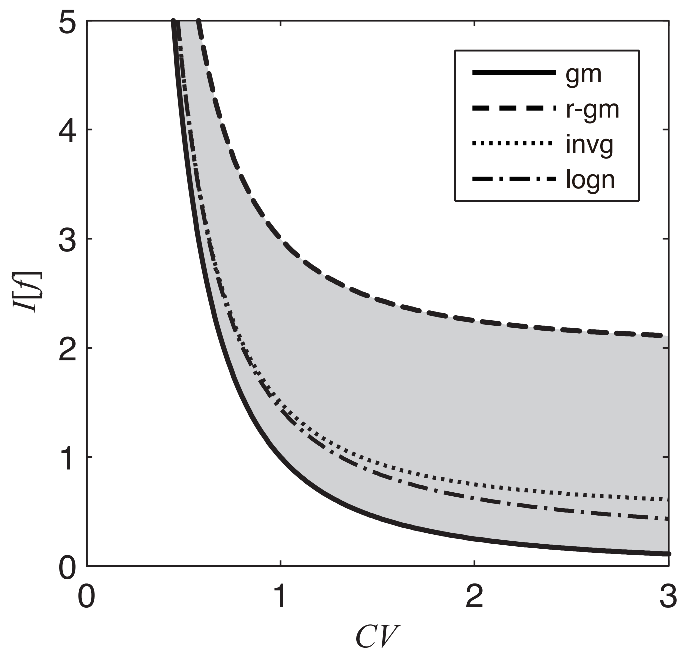

FIGURE 4. Fisher information-based dispersion of neuronal firing, $I[f]$, for the generalized inverse Gaussian family (gray region) and lognormal distribution ("logn") of interspike intervals, as a function of CV. The "gm", "r-gm" and "invg" represent the gamma, reciprocal gamma and inverse Gaussian distributions, respectively. The value of $I[f]$ plays a key role in the approximate expression for the information gain due to variable firing rate.

4. **Discussion.** In this paper, we studied how much information can be gained from variable neuronal firing rate with respect to constant firing rate. For this purpose, we employed the KL divergence and its approximation in terms of the Fisher information. It was shown that the KL divergence, defined for rate-modulated spike trains, can be reduced to the Fisher information for single ISIs in the limit of slow ($\tau\mu \gg 1$) and small ($\sigma/\mu \ll 1$) rate variation. The numerical study quantitatively verified that the Fisher information approximates the KL divergence reasonably well (Figure 3).

We stress at this point, that the overall methodology presented in this paper is not restricted to the scale-family probability distributions. The approximate relation between KL divergence and Fisher information holds for an arbitrary parameterization [39, p.26]. Here, however, we restrict ourselves to the scale parameterization due to its relationship to the rate coding scheme (see e.g., [27, 42]) and due to tractable properties of the Fisher information [26, 30]. In particular, we justify that among all scale-family ISI distributions that share the same coefficient of variation, the gamma distribution reaches the minimum information gain.

The KL divergence employed in this paper is different from the Shannon mutual information. The mutual information between the firing rate $\lambda(t)$ and spike train $\{t_i\}$ can be defined by introducing the probability distribution of $\lambda(t)$. Asymptotic relations between Fisher information and mutual information has also been well investigated, analogously to the relation of KL divergence with Fisher information [5, 10, 33, 34, 39, 45, 49]. In our case, it is easily speculated that under the same condition as the KL divergence (i.e., slow and small rate variation), the Fisher information $I[f]$ will appear in the approximate mutual information.

As the definition, $I[f]$ quantifies the change in the shape of the probability density under infinitesimally small changes in the scale. Therefore, "smooth" density shapes attain lower $I[f]$ than, e.g., multimodal ones. Similarly, [35] employed a dispersion measure based on the Fisher information for the location-family class, and analyzed its statistical properties. It may be interesting to investigate the relation among these measures.

As shown in Figure 4, $I[f]$ generally has a monotonic relationship with C_V^{-2}, but reflects additional statistical properties beyond the second moment of ISIs. $I[f]$ can take different values among various probability distributions that share the same value of C_V^{-2}, and vice versa. $I[f]$ is equal to C_V^{-2}, i.e., the minimum value, only for the gamma distribution. An interesting property of $I[f]$ is that it is directly connected with the information-theoretic measure (the KL divergence) on rate coding, while C_V is not. Therefore, characterizing the dispersion of ISI with $I[f]$ could give some new information if the rate reflected the information processing in neuronal systems (e.g., see [14, 25, 48]), and statistical properties of spike trains were significantly deviated from the gamma statistics. It would be interesting to examine if these were the case for biological spike trains.

In order to estimate the Fisher information $I[f]$ from experimentally observed spike trains, It would be preferable to perform nonparametric inference. Recently, [36] introduced a Fisher information estimator for the location family, based on the maximum penalized likelihood estimation of the probability density function [21]. Nonparametric inference of Fisher information is generally not straightforward because it contains the derivative of density, which is sensitive to estimates of the density. It remains for a future work to develop a reliable method for estimating $I[f]$ and examine how much information can be gained from real spike trains.

Acknowledgments. S.K. was supported by JSPS KAKENHI Grant Number 24700287. L.K. was supported by the Institute of Physiology RVO: 67985823, the Centre for Neuroscience, GAP304/12/G069 and by the Grant Agency of the Czech Republic projects GAP103/11/0282 and GPP103/12/P558.

Appendix A. **Derivation of Eq. (7).** Eq. (7) is written as

$$\begin{aligned} & p(\{t_i\};\mu) \\ = \quad & p_1(t_1;\mu)\cdot\prod_{i=2}^{n} r(t_i;t_{N(t_i)},\mu)\exp\left[-\int_{t_{i-1}}^{t_i} r(t;t_{i-1},\mu)dt\right]\cdot P((t_n,T];\mu), \end{aligned} \tag{42}$$

where

$$p_1(t_1;\mu) = \exp\left[-\int_0^{t_1} r(t;t_0,\mu)dt\right] r(t_1;t_0,\mu) \tag{43}$$

is the probability density of the first spike occurring at t_1[2], and

$$P((t_n,T];\mu) = \exp\left[-\int_{t_n}^{T} r(t;t_n,\mu)dt\right] \tag{44}$$

is the probability of no spikes being observed on $(t_n,T]$.

Consider the following derivative:

$$\begin{aligned}\frac{d}{dt_i}\ln\left[1-\int_{t_{i-1}}^{t_i}\mu f(\mu(u-t_{i-1}))du\right] &= -\frac{\mu f(\mu(t_i-t_{i-1}))}{1-\int_{t_{i-1}}^{t_i}\mu f(\mu(u-t_{i-1}))du}\\ &= -r(t_i;t_{i-1},\mu),\end{aligned} \tag{45}$$

from which we obtain

$$1-\int_{t_{i-1}}^{t_i}\mu f(\mu(u-t_{i-1}))du = \exp\left[-\int_{t_{i-1}}^{t_i} r(t;t_{i-1},\mu)dt\right]. \tag{46}$$

Using this, the conditional intensity function (6) is written as

$$\begin{aligned}r(t_i;t_{i-1},\mu) &= \frac{\mu f(\mu(t_i-t_{i-1}))}{1-\int_{t_{i-1}}^{t_i}\mu f(\mu(u-t_{i-1}))du}\\ &= \mu f(\mu(t_i-t_{i-1}))\exp\left[\int_{t_{i-1}}^{t_i} r(t;t_{i-1},\mu)dt\right].\end{aligned} \tag{47}$$

Substituting this into Eq. (42), we obtain Eq. (7).

Appendix B. **Derivation of Eq. (14).** Let λ_i be

$$\lambda_i := \frac{\Lambda(t_i)-\Lambda(t_{i-1})}{t_i-t_{i-1}} = \frac{1}{t_i-t_{i-1}}\int_{t_{i-1}}^{t_i}\lambda(t)dt. \tag{48}$$

From the mean-value theorem, there exists $s\in(t_{i-1},t_i)$ such that $\lambda(s)=\lambda_i$. Expanding $\lambda(t_i)$ around s, we obtain

$$\lambda(t_i) = \lambda_i + \frac{d\lambda(s)}{d(s/\tau)}\frac{t_i-s}{\tau} + O((\frac{t_i-s}{\tau})^2), \tag{49}$$

where the order of each factor is $d\lambda(s)/d(s/\tau)\sim\sigma$, $t_i-s\sim 1/\mu$. Thus, the error in approximating $\lambda(t_i)$ to λ_i is evaluated as

$$\lambda(t_i) = \lambda_i + O(\frac{\sigma}{\tau\mu}). \tag{50}$$

Using Eqs. (48) and (50), Eq. (9) is expressed as

$$p(\{t_i\};\{\lambda(t)\}) = p_1(t_1;\{\lambda(t)\})\cdot\prod_{i=2}^{n}\tilde{p}(x_i;\lambda_i)\cdot P((t_n,T];\{\lambda(t)\}), \tag{51}$$

where $x_i = t_i - t_{i-1}$ and

$$\tilde{p}(x_i;\lambda_i) = \left[\lambda_i + O(\frac{\sigma}{\tau\mu})\right] f(\lambda_i x_i). \tag{52}$$

[2] Here, $t_0(<0)$ represents the spike time preceding t_1. If we have no information about t_0, it is be set to be $p_1(t_1;\mu)=\mu[1-\int_0^{\mu t} f(u)du]$ [13].

Inserting Eq. (51) into Eq. (10), and taking it into account that $p_1(t_1|\{\lambda(t)\})$ and $P((t_n,T]|\{\lambda(t)\})$ are negligible in the limit of $T\to\infty$ if the mean firing rate is nonzero $\mu>0$, the KL divergence is rewritten as

$$\begin{aligned}
&D[p(\cdot;\{\lambda\})||p(\cdot;\mu)]\\
&\quad= \lim_{n\to\infty}\frac{1}{n-1}\sum_{i=2}^{n}\int_0^\infty\cdots\int_0^\infty\prod_{i=2}^{n}\tilde{p}(x_i;\lambda_i)\ln\frac{\prod_{i=2}^{n}\tilde{p}(x_i;\lambda_i)}{\prod_{i=2}^{n}p(x_i;\mu)}dx_2\cdots dx_n\\
&\quad= \lim_{n\to\infty}\frac{1}{n-1}\sum_{i=2}^{n}\int_0^\infty\tilde{p}(x_i;\lambda_i)\ln\frac{\tilde{p}(x_i;\lambda_i)}{p(x_i;\mu)}dx_i\\
&\quad= \lim_{n\to\infty}\frac{1}{n-1}\sum_{i=2}^{n}D_s(\lambda_i||\mu)+O(\frac{\sigma}{\tau\mu^2}),
\end{aligned}\tag{53}$$

where

$$D_s(\lambda_i||\mu):=\int_0^\infty p(x_i;\lambda_i)\ln\frac{p(x_i;\lambda_i)}{p(x_i;\mu)}dx_i \tag{54}$$

is the KL divergence between the two ISI densities. Thus, the KL divergence of the spike trains is reduced to the KL divergence of the single ISIs.

Since the mean of $\lambda(t)$ is given by μ, λ_i may be expressed by $\lambda_i=\mu+\delta_i$ with $\delta_i\sim O(\sigma)$. Using the scaling property of $D_s(\lambda||\mu)$ (i.e., $D_s(\lambda||\mu)=D_s(c\lambda||c\mu)$, $c>0$), we obtain

$$\begin{aligned}
&D_s(\lambda_i||\mu)\\
&\quad= D_s(1||\frac{\mu}{\mu+\delta_i})\\
&\quad= D_s(1||1)+\frac{\partial D_s(1||z)}{\partial z}\Big|_{z=1}\cdot\xi_i+\frac{1}{2}\frac{\partial^2 D_s(1||z)}{\partial z^2}\Big|_{z=1}\cdot\xi_i^2+O(\xi_i^3).
\end{aligned}\tag{55}$$

where $\xi_i=-\delta_i/\mu+(\delta_i/\mu)^2-\cdots$. The first and second terms in the above equation vanish, and the coefficient of the third term is computed as

$$\frac{\partial^2 D_s(1||z)}{\partial z^2}\Big|_{z=1} = -\int_0^\infty\frac{\partial^2\ln p(x;z)}{\partial z^2}p(x;1)dx\Big|_{z=1}=I[f]. \tag{56}$$

Thus, the KL divergence of the single ISIs is expanded with respect to δ_i/μ as

$$D_s(\lambda_i||\mu)=\frac{I[f]}{2}(\frac{\delta_i}{\mu})^2+O((\frac{\delta_i}{\mu})^3). \tag{57}$$

Substituting Eq. (57) into Eq. (53), the KL divergence is obtained as

$$D[p(\cdot;\{\lambda\})||p(\cdot;\mu)] = \frac{I[f]}{2}\lim_{n\to\infty}\frac{1}{n-1}\sum_{i=2}^{n}(\frac{\delta_i}{\mu})^2+O(\frac{\sigma}{\tau\mu^2}). \tag{58}$$

If we further assume that the rate fluctuation is ergodic with a limiting density $p(\lambda)$, the summation can be replaced as

$$\lim_{n\to\infty}\frac{1}{n-1}\sum_{i=2}^{n}\to\int_0^\infty\frac{\lambda p(\lambda)}{\mu}d\lambda=\int_0^\infty(1+\frac{\delta}{\mu})p(\mu+\delta)d\delta, \tag{59}$$

where $\delta = \lambda - \mu$. Taking into account $\delta_i = \delta + O(\sigma/(\tau\mu))$ from Eq. (50), Eq. (58) becomes

$$\begin{aligned} & D[p(\cdot;\{\lambda\})||p(\cdot;\mu)] \\ = & \frac{I[f]}{2}\int_0^\infty \Big(\frac{\delta}{\mu} + O(\frac{\sigma}{\tau\mu^2})\Big)^2 (1+\frac{\delta}{\mu})p(\mu+\delta)d\delta + O(\frac{\sigma}{\tau\mu^2}) \\ = & \frac{\sigma^2}{2\mu^2}I[f] + O(\frac{\sigma}{\tau\mu^2}). \end{aligned} \tag{60}$$

Appendix C. **Proof of minimum Fisher information.** Consider the scale family of densities $p(t;\lambda)$ given by Eq. (11). The mean is given by $E(T) = 1/\lambda$ if the mean of $f(x)$ is unity. Then, applying the Cramér-Rao inequality [45] leads to

$$Var(T) \geq \frac{\big(\frac{dE(T)}{d\lambda}\big)^2}{J(\lambda)} = \frac{E(T)^2}{I[f]}, \tag{61}$$

from which we obtain

$$I[f] \geq \frac{E(T)^2}{Var(T)} = C_V^{-2}, \tag{62}$$

where $C_V = \sqrt{Var(T)}/E(T)$ is the coefficient of variation of $p(t;\lambda)$ as well as $f(x)$. Thus, given the coefficient variation $C_V = 1/\sqrt{\kappa}$, the density function $f(x)$ that achieves the minimum fisher information satisfies

$$I[f] = 1 - \int_0^\infty x^2 \frac{\partial^2 \ln f(x)}{\partial x^2} f(x)dx = \kappa, \tag{63}$$

which holds if $f(x)$ satisfies

$$x^2 \frac{\partial^2 \ln f(x)}{\partial x^2} = 1 - \kappa. \tag{64}$$

The solution of the differential equation (64) is found to be $f(x) = \exp[(\kappa-1)\log x - c_1 x + c_2]$, where $c_1 = \kappa$ and $c_2 = \kappa \log\kappa - \log\Gamma(\kappa)$ are the constants of integration that are determined from the normalization condition $\int_0^\infty f(x)dx = 1$ and the mean of $f(x)$. Therefore, it is shown that the gamma density function (16) attains the minimum of $I[f]$.

Let $f(x)$ and $g(x)$ be probability densities having fixed mean and the variance. Then, it is easily proven that, for $0 \leq \theta \leq 1$, $\theta f(x) + (1-\theta)g(x)$ also has the same mean and variance. Therefore, the set of distributions with fixed mean and variance is convex. Since $I[f]$ is a strictly convex functional [12], only the gamma density function achieves the minimum Fisher information.

Appendix D. $E(1/X)$ **of the GIG.** Applying $E[\frac{\partial}{\partial\theta}\lg f(X;\theta)] = 0$ to Eq. (19) leads to

$$E\Big[\frac{\partial \ln f(x;w,\eta,a)}{\partial w}\Big] = -\frac{\frac{\partial K_a(w)}{\partial w}}{K_a(w)} - \frac{1}{2}\Big[\frac{E(X)}{\eta} + \eta E\Big(\frac{1}{X}\Big)\Big] = 0. \tag{65}$$

Using Eq. (20) and

$$\frac{\partial K_a(w)}{\partial w} = -\frac{1}{2}[K_{a+1}(w) + K_{a-1}(w)], \tag{66}$$

$E(1/X)$ is obtained as

$$E\Big(\frac{1}{X}\Big) = \frac{1}{\eta}\frac{K_{a-1}(w)}{K_a(w)}. \tag{67}$$

Another expression of $E(1/X)$ is obtained by taking the derivative of $\log f$ with respect to η,

$$E\Big(\frac{\partial}{\partial\eta}\log f(x;w,\eta,a)\Big) \quad = \quad -\frac{a}{\eta}-\frac{1}{2}\Big[-\frac{wE(X)}{\eta^2}+wE\Big(\frac{1}{X}\Big)\Big]=0. \qquad (68)$$

Using Eq. (20), $E(1/X)$ is obtained as

$$E\Big(\frac{1}{X}\Big)=\frac{1}{\eta}\frac{K_{a+1}(w)}{K_a(w)}-\frac{2a}{w\eta}. \qquad (69)$$

From Eqs. (67) and (69), $E(1/X)$ is also expressed as

$$E\Big(\frac{1}{X}\Big)=\frac{1}{2\eta}\frac{K_{a+1}(w)+K_{a-1}(w)}{K_a(w)}-\frac{a}{w\eta}. \qquad (70)$$

REFERENCES

[1] M. Abramowitz and I. A. Stegun, eds., "Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables," Dover Publications, Inc., New York, 1966.

[2] E. D. Adrian, *The basis of sensation*, Br. Med. J., **1** (1954).

[3] R. Barbieri, M. C. Quirk, L. M. Frank, M. A. Wilson and E. N. Brown, *Construction and analysis of non-Poisson stimulus-response models of neural spiking activity*, Journal of Neuroscience Methods, **105** (2001), 25–37.

[4] M. Berman, *Inhomogeneous and modulated gamma processes*, Biometrika, **68** (1981), 143–152.

[5] J. M. Bernardo, *Reference posterior distributions for Bayesian inference. With discussion*, J. Roy. Stat. Soc. B, **41** (1979), 113–147.

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

[7] G. S. Bhumbra, A. N. Inyushkin and R. E. J. Dyball, *Assessment of spike activity in the supraoptic nucleus*, J. Neuroendocrinol., **16** (2004), 390–397.

[8] L. Bonnasse-Gahot and J.-P. Nadal, *Perception of categories: From coding efficiency to reaction times*, Brain Res., **1434** (2012), 47–61.

[9] A. Borst and F. E. Theunissen, *Information theory and neural coding*, Nature Neurosci., **2** (1999), 947–958.

[10] N. Brunel and J.-P. Nadal, *Mutual information, Fisher information, and population coding*, Neural Computation, **10** (1998), 1731–1757.

[11] R. S. Chhikara and J. L. Folks, "The Inverse Gaussian Distribution: Theory, Methodology, and Applications," Marcel Dekker, New York, 1989.

[12] M. Cohen, *The fisher information and convexity*, IEEE Transactions on Information Theory, **14** (1968), 591–592.

[13] D. R. Cox and P. A. W. Lewis, "The Statistical Analysis of Series of Events," Methuen & Co., Ltd., London; John Wiley & Sons, Inc., New York, 1966.

[14] J. P. Cunningham, V. Gilja, S. I. Ryu and K. V. Shenoy, *Methods for estimating neural firing rates, and their application to brain-machine interfaces*, Neural Networks, **22** (2009), 1235–1246.

[15] J. P. Cunningham, B. M. Yu, K. V. Shenoy and M. Sahani, *Inferring neural firing rates from spike trains using Gaussian processes*, in "Neural Information Processing Systems" (eds. J. C. Platt, D. Koller, Y. Singer and S. Roweis), Vol. 20, (2008), 329–336.

[16] D. J. Daley and D. Vere-Jones, "An Introduction to the Theory of Point Processes. Vol. I. Elementary Theory and Methods," Second edition, Probability and its Applications (New York), Springer-Verlag, New York, 2003.

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

[18] R. G. Gallager, "Information Theory and Reliable Communication," John Wiley & Sons, Inc., New York, 1968.

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

[20] I. J. Good, "Probability and the Weighing of Evidence," Charles Griffin & Co., Ltd., London; Hafner Publishing Co., New York, N. Y., 1950.

[21] I. J. Good and R. A. Gaskins, *Nonparametric roughness penalties for probability densities*, Biometrika, **58** (1971), 255–277.

[22] P. E. Greenwood and P. Lánský, *Optimal signal estimation in neuronal models*, Neural Comput., **17** (2005), 2240–2257.

[23] P. E. Greenwood and P. Lánský, *Optimum signal in a simple neuronal model with signal-dependent noise*, Biol. Cybern., **92** (2005), 199–205.

[24] P. E. Greenwood, L. M. Ward, D. F. Russell, A. Neiman and F. Moss, *Stochastic resonance enhances the electrosensory information available to paddlefish for prey capture*, Phys. Rev. Lett., **84** (2000), 4773–4776.

[25] A. Grémiaux, T. Nowotny, D. Martinez, P. Lucas and J.-P. Rospars, *Modelling the signal delivered by a population of first-order neurons in a moth olfactory system*, Brain Res., **1434** (2012), 123–135.

[26] P. J. Huber, "Robust Statistics," Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1981.

[27] S. Ikeda and J. H. Manton, *Capacity of a single spiking neuron channel*, Neural Comput., **21** (2009), 1714–1748.

[28] S. Iyengar and Q. Liao, *Modeling neural activity using the generalized inverse gaussian distribution*, Biological Cybernetics, **77** (1997), 289–295.

[29] B. Jørgensen, "Statistical Properties of the Generalized Inverse Gaussian Distribution," Lecture Notes in Statistics, **9**, Springer-Verlag, New York-Berlin, 1982.

[30] A. M. Kagan, I. V. Linnik and C. R. Rao, "Characterization Problems in Mathematical Statistics," John Wiley & Sons, New York, 1973.

[31] R. E. Kass and V. Ventura, *A spike-train probability model*, Neural Computation, **13** (2001), 1713–1720.

[32] S. M. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory," Prentice Hall, New Jersey, 1993.

[33] L. Kostal, *Information capacity in the weak-signal approximation*, Phys. Rev. E, **82** (2010), 026115.

[34] L. Kostal, *Approximate information capacity of the perfect integrate-and-fire neuron using the temporal code*, Brain Res., **1434** (2012), 136–141.

[35] L. Kostal, P. Lansky and O. Pokora, *Variability measures of positive random variables*, PLoS ONE, **6** (2011), e21998.

[36] L. Kostal and O. Pokora, *Nonparametric estimation of information-based measures of statistical dispersion*, Entropy, **14** (2012), 1221–1233.

[37] S. Koyama, *Coding efficiency and detectability of rate fluctuations with non-Poisson neuronal firing*, in "Neural Information Processing Systems," Vol. 25, The Institute of Statistical Mathematics, 2013.

[38] S. Koyama and R. E. Kass, *Spike train probability models for stimulus-driven leaky integrate-and-fire neurons*, Neural Computation, **20** (2008), 1776–1795.

[39] S. Kullback, "Information Theory and Statistics," Dover Publications, Inc., Mineola, New York, 1968.

[40] E. L. Lehmann and G. Casella, "Theory of Point Estimation," Second edition, Springer Texts in Statistics, Springer-Verlag, New York, 1998.

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

[42] Z. Pawlas, L. B. Klebanov, M. Prokop and P. Lansky, *Parameters of spike trains observed in a short time window*, Neural Comput., **20** (2008), 1325–1343.

[43] D. H. Perkel and T. H. Bullock, *Neural coding*, Neurosci. Res. Prog. Sum., **3** (1968), 405–527.

[44] J. W. Pillow, *Time-rescaling methods for the estimation and assessment of non-Poisson neural encoding models*, in "Neural Information Processing Systems" (eds. Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams and A. Culotta), Vol. 22, (2008), 1473–1481.

[45] E. J. G. Pitman, "Some Basic Theory for Statistical Inference," Monographs on Applied Probability and Statistics, Chapman and Hall, London; A Halsted Press Book, John Wiley & Sons, New York, 1979.

[46] C. Pouzat and A. Chaffiol, *Automatic spike train analysis and report generation. An implementation with R, R2HTML and STAR*, J. Neurosci. Methods, **181** (2009), 119–144.

[47] D. S. Reich, J. D. Victor and B. W. Knight, *The power ratio and the interval map: Spiking models and extracellular recordings*, Journal of Neuroscience, **18** (1998), 10090–10104.

[48] B. J. Richmond and L. M. Optican, *Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform*, Journal of Neurophysiology, **57** (1987), 147–161.

[49] J. J. Rissanen, *Fisher information and stochastic complexity*, IEEE Trans. Inf. Theory, **42** (1996), 40–47.

[50] L. J. Savage, "The Foundations of Statistics," John Wiley & Sons, Inc., New York; Chapman & Hill, Ltd., London, 1954.

[51] H. S. Seung and H. Sompolinsky, *Simple models for reading neuronal population codes*, Proceedings of the National Academy of Sciences of the United States of America, **90** (1993), 10749–10753.

[52] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," University of Illinois Press, Urbana, Illinois, 1949.

[53] R. B. Stein, *The information capacity of nerve cells using a frequency code*, Biophys. J., **7** (1967), 797–826.

[54] F. Theunissen and J. P. Miller, *Temporal encoding in nervous systems: A rigorous definition*, J. Comput. Neurosci., **2** (1995), 149–162.

[55] H. C. Tuckwell, "Introduction to Theoretical Neurobiology, Vol. 2. Nonlinear and Stochastic Theories," Cambridge Studies in Mathematical Biology, **8**, Cambridge University Press, Cambridge, 1988.

[56] A. W. van der Vaart, "Asymptotic Statistics," Cambridge Series in Statistical and Probabilistic Mathematics, **3**, Cambridge University Press, Cambridge, 1998.

[57] K. Zhang and T. Sejnowski, *Neural tuning: To sharpen or broaden?*, Neural Computation, **11** (1999), 75–84.

Received December 20, 2012; Accepted April 13, 2013.

E-mail address: skoyama@ism.ac.jp
E-mail address: kostal@biomed.cas.cz