Noname manuscript No.
(will be inserted by the editor)

A characterization of the time-rescaled gamma process as a model for spike trains

Takeaki Shimokawa · Shinsuke Koyama · Shigeru Shinomoto

October 14, 2009

Abstract The occurrence of neuronal spikes may be characterized by not only the rate but also the irregularity of firing. We have recently developed a Bayes method for characterizing a sequence of spikes in terms of instantaneous rate and irregularity, assuming that interspike intervals (ISIs) are drawn from a distribution whose shape may vary in time. Though any parameterized family of ISI distribution can be installed in the Bayes method, the ability to detect firing characteristics may depend on the choice of a family of distribution. Here, we select a set of ISI metrics that may effectively characterize spike patterns and determine the distribution that may extract these characteristics. The set of the mean ISI and the mean log ISI are uniquely selected based on the statistical orthogonality, and accordingly the corresponding distribution is the gamma distribution. By applying the Bayes method equipped with the gamma distribution to spike sequences derived from different ISI distributions such as the log-normal and inverse-Gaussian distribution, we confirm that the gamma distribution effectively extracts the rate and the shape factor.

Keywords firing irregularity · firing rate · gamma distribution · point processes · Bayesian estimation

1 Introduction

In the analysis of neuronal spike trains, attention has been paid mainly to the rate of spike occurrences correlated to the stimulus or the behavioral context of the animal [Gerstein and Kiang 1960, Abeles 1982, DiMatteo et al. 2001, Dayan and Abbott 2001, Kass et al. 2005, Shimazaki and Shinomoto 2007]. From the same sequence, it may be

T. Shimokawa · S. Shinomoto
Department of Physics, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
E-mail: shimokawa@ton.scphys.kyoto-u.ac.jp
E-mail: shinomoto@scphys.kyoto-u.ac.jp

S. Koyama
Department of Statistics and Center for Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, USA
E-mail: koyama@stat.cmu.edu

possible to further extract information about irregularity of firing [Kara et al. 2000, Kostal and Lansky 2006, Mitchell et al. 2007]. The irregularity of firing has been measured by the ISI distribution, or the coefficient of variation C_V defined by the ratio of the standard deviation to the mean of ISIs. However, these metrics are easily affected by fluctuation in the mean rate of firing caused by extrinsic stimuli [Gabbiani and Koch 1998, Nawrot et al. 2008]. A sequence of spikes that looks irregular can be interpreted either as being randomly generated at a fixed rate, or as being derived regularly with a fluctuating rate. We have proposed the Bayes method to select an interpretation from a continuous spectrum between these two extremes in light of the statistical plausibility for each given spike train [Koyama and Shinomoto 2005]. The nature of intrinsic randomness of neuronal firing has also been discussed using metrics such as the local variation L_V measuring the instantaneous irregularity [Shinomoto et al. 2003, Shinomoto et al. 2005, Shinomoto et al. 2009, Davies et al. 2006, Holt et al. 1996, Miura et al. 2006]. The use of information on the firing regularity was proposed to improve the firing rate estimation [Cunningham et al. 2008].

In these studies, the degree of irregularity or regularity is assumed to be constant throughout a single train of neuronal spikes. However, it was also reported that the firing irregularity may vary with behavioral context [Davies et al. 2006, Churchland et al. 2006, Mitchell et al. 2007]. To examine how the firing irregularity varies in time or with context, we have recently extended the Bayes method so that it can estimate both the rate and the regularity moment by moment, by assuming that the shape of the ISI distribution varies in time [Shimokawa and Shinomoto 2009].

In measuring the firing irregularity, any parameterized family of ISI distributions cannot perfectly account for the underlying distribution that has actually generated a sequence of spikes. Therefore, the accuracy of the inference would depend on the choice of a family of distributions that is installed in the Bayes method. Here, we first select a set of ISI metrics that may characterize any spike sequence effectively and a family of distributions suitable for extracting them. By imposing the information geometrical orthogonality, we found that a suitable set of metrics is the mean ISI and the mean log ISI, and the corresponding distribution is the gamma distribution. Next, we examine the Bayes method equipped with the gamma distribution in its ability in estimating the instantaneous rate and regularity of spike sequences. It is found that the Bayes method equipped with the gamma distribution can capture the instantaneous firing rate and regularity reasonably well, even for spike sequences derived from the log-normal and inverse-Gaussian ISI distributions.

2 Orthogonal ISI metrics

For the purpose of characterizing spike trains, we seek a set of ISI metrics that satisfy information geometrical orthogonality, under which the estimators can be independent, thereby the estimation of each metric is not affected by the estimation error of another metric. Here we assume a renewal process, in which ISIs are derived from an identical distribution $p(T)$. Under the principle of maximum entropy [Jaynes 1957, Schneidman et al. 2006], the ISI distribution $p(T)$ may be uniquely determined, given information on the spiking characteristics.

The most basic information about a spike sequence is the firing rate $\lambda(= 1/E[T] = 1/\int_0^\infty Tp(T)\,dT)$. We define a rate-rescaled distribution assuming scale invariance [Reich et al. 1998] by

$$p(T)dT = \lambda g(\lambda T)dT, \tag{1}$$

thus satisfying

$$\int_0^\infty g(\Lambda)d\Lambda = 1, \tag{2}$$

$$\int_0^\infty \Lambda g(\Lambda)d\Lambda = 1, \tag{3}$$

where $\Lambda \equiv \lambda T$. With the rate-rescaled distribution $g(\Lambda)$, the differential entropy can be separated into rate and shape terms,

$$h = -\int_0^\infty p(T)\log p(T)dT \tag{4}$$

$$= -\log\lambda - \int_0^\infty g(\Lambda)\log g(\Lambda)d\Lambda. \tag{5}$$

Given the firing rate λ, the distribution that maximizes the differential entropy is an exponential function, $\exp(-\Lambda)$. The process of generating spikes with this exponential ISI distribution is the Poisson process.

Next, let us consider the situation in which we obtain additional information with another firing metric $A(\Lambda)$,

$$\int_0^\infty A(\Lambda)g(\Lambda)d\Lambda = \eta. \tag{6}$$

The distribution that maximizes the entropy under the given λ and η is obtained by the method of Lagrange multipliers [Kapur 1989],

$$\int_0^\infty \delta g(\Lambda)\left\{\log g(\Lambda) + 1 + a + b\Lambda + cA(\Lambda)\right\}d\Lambda = 0, \tag{7}$$

giving the rate-rescaled distribution,

$$g(\Lambda) = \exp\left[-\left\{1 + a + b\Lambda + cA(\Lambda)\right\}\right]. \tag{8}$$

To make the ISI metric $A(\Lambda)$ effective in characterizing spike sequences, we require it to be orthogonal to the mean ISI. The information geometrical orthogonality [Amari and Nagaoka 2000, Ikeda 2005], or diagonalizing the Fisher information matrix, is given by

$$E\left[\frac{\partial^2}{\partial\lambda\partial\eta}\log p(T)\right] = 0. \tag{9}$$

Using Eqs.(1), (8), and (9), we obtain

$$\frac{1}{\lambda}\frac{\partial b}{\partial\eta} + \frac{\partial c}{\partial\eta}E[TA'(\lambda T)] = 0. \tag{10}$$

Here,

$$\frac{1}{\lambda}\frac{\partial b}{\partial\eta} + \frac{\partial c}{\partial\eta}E[TA'(\lambda T)] + c\frac{\partial}{\partial\eta}E[TA'(\lambda T)] = 0 \tag{11}$$

is derived by differentiating the following equation with respect to η,

$$E\left[\frac{\partial}{\partial\lambda}\log p(T)\right] = \frac{\partial}{\partial\lambda}\int_0^\infty dT p(T) = 0. \tag{12}$$

Combining Eqs.(10) and (11), we obtain

$$\frac{\partial}{\partial\eta}E[TA'(\lambda T)] = 0. \tag{13}$$

In this way, the condition for the metric $E[A(\lambda T)]$ being always orthogonal to $E[T]$ is that $TA'(\lambda T) = \text{const.}$, or equivalently,

$$A(\lambda T) \propto \log\lambda T. \tag{14}$$

Thus, the orthogonal set of metrics is found to be the mean ISI and the mean log ISI, which represent the scale and the shape of the ISI distribution, respectively. The suitability of the log ISI in measuring the firing pattern was also discussed by Dorval, who noted that with logarithmic partitioning of ISIs, firing rate changes become independent of firing pattern entropy [Dorval 2008]. Furthermore, for a rate-rescaled distribution (1) whose shape is parameterized by κ, the expectation of the metric, $-E[\log\lambda T] = \log E[T] - E[\log T]$ is computed as

$$\begin{aligned}-E[\log\lambda T] &= -\int_0^\infty p(T)\log\lambda T dT \\ &= -\int_0^\infty g_\kappa(\Lambda)\log\Lambda d\Lambda \\ &\equiv f(\kappa).\end{aligned} \tag{15}$$

(Here, we expressed the dependency of g_κ on κ explicitly.) Thus, $-E[\log\lambda T]$ is a function of only κ, suggesting that the shape of a rate-rescaled distribution may be parameterized by $-E[\log\lambda T]$. Hence, $-E[\log\lambda T]$ is suitable for characterizing the firing regularity independently from the firing rate.

$-E[\log\lambda T]$ may be expanded as

$$-E[\log\lambda T] = -E\left[\log\left(1+\frac{T-E[T]}{E[T]}\right)\right] \tag{16}$$

$$= \frac{E[(T-E[T])^2]}{2E[T]^2} - \frac{E[(T-E[T])^3]}{3E[T]^3} + \frac{E[(T-E[T])^4]}{4E[T]^4} - \cdots, \tag{17}$$

where the first term of the rhs in above equation is squared C_V. $-E[\log\lambda T]$ is, thus, interpreted as a generalization of C_V incorporating higher-order statistics.

Taking $f(\kappa)$ in Eq.(15) to be

$$f(\kappa) = \log\kappa - \psi(\kappa), \tag{18}$$

where $\psi(\kappa)$ is the digamma function, and combining it with Eq.(14), the distribution that maximizes the differential entropy given the mean ISI and the mean log ISI is found to be the gamma distribution [Kapur 1989],

$$p(T) = \frac{\lambda^\kappa\kappa^\kappa}{\Gamma(\kappa)}T^{\kappa-1}e^{-\lambda\kappa T}, \tag{19}$$

where $\Gamma(\kappa) \equiv \int_0^\infty x^{\kappa-1}e^{-\kappa x}dx$ is the gamma function. We will hereafter characterize any sequence of ISIs drawn from any distribution in terms of the firing rate λ and the firing regularity κ.

3 Bayesian inference of the rate and the regularity of firing

3.1 Spike generation

In deriving a Bayesian algorithm for characterizing spike sequences, we first consider a process of drawing ISIs independently from an identical ISI distribution $p(T \mid \theta)$. Here, we specify a set of parameters as $\theta = \{\lambda, \kappa\}$ that characterize the scale and shape of the ISI distribution (Fig.1).

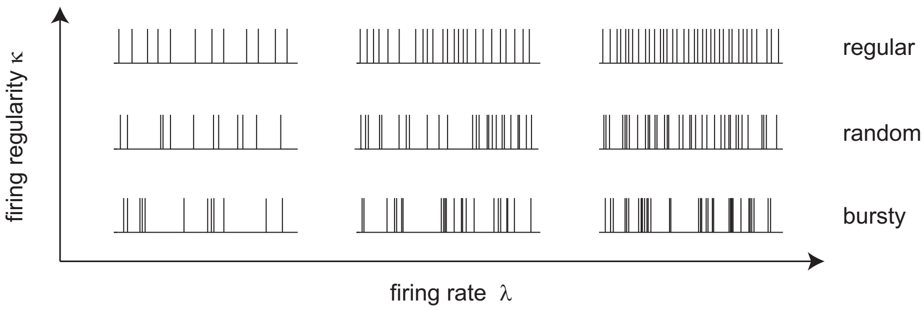

Fig. 1 A variety of ISI sequences derived from the gamma distributions in the parameter space $\theta = \{\lambda, \kappa\}$ representing the scale and shape of the distribution.

Let us consider the situation in which the parameters or the state variables θ vary in time. The probability for spikes to occur at times $\{t_i\}_{i=0}^{n}$ is approximated by the product of the probabilities with the instantaneous state variables at the occurrence of spikes,

$$p(\{t_i\} \mid \{\theta(t)\}) = \prod_{i=0}^{n-1} p(T_i \mid \theta(t_i)), \tag{20}$$

where $T_i \equiv t_{i+1} - t_i$ is the ith interspike interval.

In the preceding study [Shimokawa and Shinomoto 2009], we employed an accurate formula that reproduces the probability for the inhomogeneous Poisson process, $\exp\left(-\int r(t)dt\right)$, in the case of $\kappa = 1$. However, the approximated formula (20) has an advantage that the state space model can be applied straightforwardly. Note that approximation (20) is secured if the state variables $\theta(t)$ are slowly modulated. This is inevitably required for the inference of rate and regularity, since a certain number of ISIs are needed for the estimation of parameters, and it is essentially impossible to infer the variation of state variables during an ISI in which there is no spike.

3.2 Parameter inference

Next, we estimate the set of parameters, $\theta(t)$, from a given sequence of spikes, $\{t_i\}_{i=0}^{n}$, by inverting the conditional distribution Eq.(20) with the Bayes formula,

$$p(\{\theta(t)\} \mid \{t_i\}; \gamma) = \frac{p(\{t_i\} \mid \{\theta(t)\})\, p(\{\theta(t)\}; \gamma)}{p(\{t_i\}; \gamma)}. \tag{21}$$

The prior distributions for the parameters that represent the scale and shape factors are introduced here so that their large gradients are penalized; we introduce the random-walk-type priors [Bialek et al. 1996] or equivalently the Gaussian process priors [Rasmussen and Williams 2006, Cunningham et al. 2008] for the rate and the shape parameter,

$$p(\{\theta(t)\};\gamma) = p(\{\lambda(t)\};\gamma_\lambda)\, p(\{\kappa(t)\};\gamma_\kappa), \tag{22}$$

$$p(\{\lambda(t)\};\gamma_\lambda) = \frac{1}{Z(\gamma_\lambda)} \exp\left[-\frac{1}{2\gamma_\lambda^2}\int_0^T \left(\frac{d\log\lambda(t)}{dt}\right)^2 dt\right], \tag{23}$$

$$p(\{\kappa(t)\};\gamma_\kappa) = \frac{1}{Z(\gamma_\kappa)} \exp\left[-\frac{1}{2\gamma_\kappa^2}\int_0^T \left(\frac{d\log\kappa(t)}{dt}\right)^2 dt\right], \tag{24}$$

where $\gamma = (\gamma_\lambda, \gamma_\kappa)^T$ is a set of hyperparameters representing the degrees of flatness for the modulation of the logarithms of rate $\lambda(t)$ and the shape parameter $\kappa(t)$. $Z(\gamma_\lambda)$ and $Z(\gamma_\kappa)$ are the normalization constants. The result of the inference would be robust to the order of differentiation in the prior distribution, as has been studied for the rate estimation [Nemenman and Bialek, 2002].

In the empirical Bayes method, hyperparameters γ is determined so that it maximizes the marginalized likelihood [MacKay 1992, Carlin and Louis 2000],

$$p(\{t_i\};\gamma) \equiv \int d\{\theta(t)\}\, p(\{t_i\} \mid \{\theta(t)\})\, p(\{\theta(t)\};\gamma). \tag{25}$$

Under the optimized hyperparameter $\gamma = \gamma^*$, we can obtain the maximum *a posteriori* (MAP) estimates of the rate and the shape parameter $\theta(t)$ for which the posterior distribution,

$$p(\{\theta(t)\} \mid \{t_i\};\gamma^*) \propto p(\{t_i\} \mid \{\theta(t)\})\, p(\{\theta(t)\};\gamma^*), \tag{26}$$

is maximized.

The maximization of marginal likelihood (Eq.(25)) can be performed by the expectation and maximization (EM) method and Kalman filtering. Both the marginal likelihood maximization and the exact evaluation of the MAP solution (Eq.(26)) require computation of the linear order of data size, $O(n)$. Algorithms that allow practical estimation are summarized in the Appendix.

4 The firing rate and regularity for other distributions

We consider here how the firing rate λ and firing regularity κ are translated into the parameters of other distributions. For mathematical convenience, we take up here the log-normal and inverse-Gaussian distributions. The log-normal distribution is given by

$$p(T \mid \mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma T} \exp\left[-\frac{(\log T - \mu)^2}{2\sigma^2}\right]. \tag{27}$$

The firing rate λ and firing regularity κ of a log-normal function are given in terms of the parameter of the log-normal distribution μ and σ by

$$\lambda = \frac{1}{E[T]} = \exp\left\{-(\mu + \frac{\sigma^2}{2})\right\}, \tag{28}$$

$$\log\kappa - \psi(\kappa) = -E[\log\lambda T] = \frac{\sigma^2}{2}. \tag{29}$$

These relationships may be inverted to determine the parameter of the log-normal distribution μ and σ, given the firing rate λ and firing regularity κ,

$$p(T \mid \lambda, \kappa) = \frac{1}{2T\sqrt{\pi(\log\kappa - \psi(\kappa))}} \exp\left[-\frac{(\log\lambda T + \log\kappa - \psi(\kappa))^2}{4(\log\kappa - \psi(\kappa))}\right]. \quad (30)$$

In the same way, the inverse-Gaussian distribution is represented in terms of λ and κ as

$$p(T \mid \lambda, \kappa) = \frac{1}{\sqrt{2\pi\lambda\sigma^2(\kappa)T^3}} \exp\left[-\frac{(\lambda T - 1)^2}{2\lambda\sigma^2(\kappa)T}\right], \quad (31)$$

where $\sigma^2(\kappa)$ is obtained by solving

$$\log\kappa - \psi(\kappa) = -E[\log\lambda T] = e^{\frac{2}{\sigma^2}} E_1\left(\frac{2}{\sigma^2}\right), \quad (32)$$

$E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt$ being the exponential integral. Figure 2 compares the log-normal, inverse-Gaussian and gamma distributions that are equivalent in terms of λ and κ.

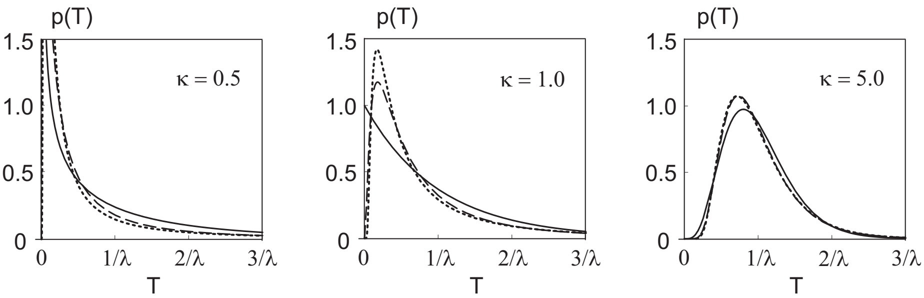

Fig. 2 The log-normal distributions (dashed lines), inverse-Gaussian distributions (dotted lines), and gamma distributions (solid lines) that yield the regularity $\kappa = 0.5, 1$, and 5, while the rate λ is fixed at 1.

5 Efficiency comparison of different Bayes methods

Here we examine how well the Bayes method equipped with the gamma distribution estimates the underlying rate and the regularity. For this purpose, we first generate a variety of spike sequences with the rate $\lambda(t)$ and regularity $\kappa(t)$ that are modulated in time, according to Eq.(20) on the basis of the gamma distribution, Eq.(19), the log-normal distribution, Eq.(30), or the inverse-Gaussian distribution, Eq.(31). Next, we apply the Bayes method to each sequence.

Figure 3 depicts the integrated squared errors (ISEs) of the estimates from the designated rate $\lambda(t)$ and regularity $\kappa(t)$, respectively, defined by $1/T \int_0^T (\hat{\lambda}(t) - \lambda(t))^2 dt$ and $1/T \int_0^T (\hat{\kappa}(t) - \kappa(t))^2 dt$. While the ISEs of estimated rate and regularity are the smallest when the underlying model is the gamma distribution, the Bayes method equipped with the gamma distribution is able to capture the rate and the regularity even when it does not match the underlying ISI distributions (i.e. the log-normal or inverse Gaussian distributions). The identification of the rate and regularity seems to hold (ISE converges to zero) in the adiabatic limit, in which the speed of the modulation is infinitesimal. At the opposite extreme of the rapid modulation of the firing rate, the optimal estimate of the hyperparameter, γ_λ, takes its value to be zero, resulting that the ISE of the estimated rate becomes practically independent of time scale above a certain $1/\tau_1$. This implies that rate estimation is unrealizable for processes whose underlying rate is modulated too rapidly, as has been reported for Bayesian rate estimation [Koyama et al. 2007]. The ISE of estimated regularity $\hat{\kappa}$ also exhibits similar dependence, implying that the estimation of regularity is also unrealizable for processes whose underlying regularity is modulated too rapidly. It is also observed that the time scales above which the estimated rate or that of regularity becomes independent of time scale of modulation are approximately the same among the distributions. These results support that the gamma distribution works reasonably well for estimating the rate and regularity even when it does not match the underlying distribution.

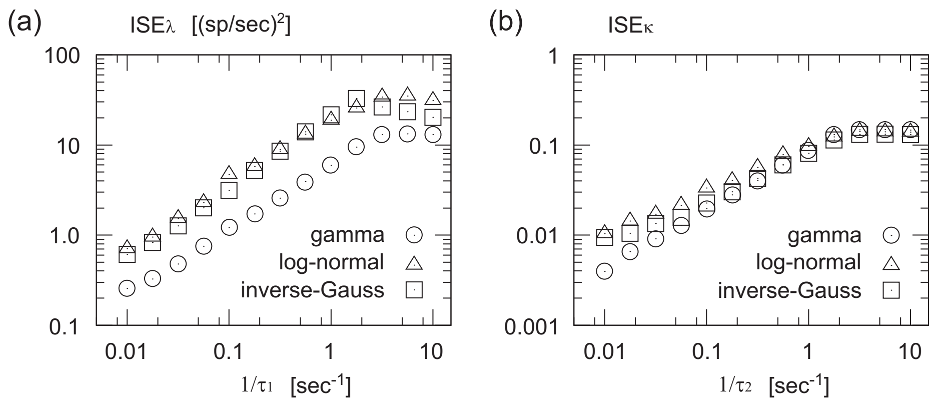

Fig. 3 Integrated squared errors (ISEs) of the estimates of (a) the rate $1/T \int_0^T (\hat{\lambda}(t) - \lambda(t))^2 dt$ and (b) regularity $1/T \int_0^T (\hat{\kappa}(t) - \kappa(t))^2 dt$ plotted against the inverse timescales $1/\tau$ for the modulation of rate and regularity (in each case, another timescale was fixed, $\tau_2 = 10$ sec, or $\tau_1 = 10$ sec). The designated rate and regularity were given by $\lambda(t) = 10 + 5.0 \sin(t/\tau_1)$ and $\kappa(t) = 1 + 0.5 \sin(t/\tau_2 + \pi/2)$. "◯", "△" and "□" represent the results of the cases in which the underlying model is given by the gamma, log-normal and inverse-Gaussian distribution, respectively.

6 Discussion

In this paper, we selected a set of ISI metrics that may characterize spike patterns effectively on the basis of information geometrical orthogonality. The orthogonality was imposed in deriving the metrics so that the estimates of the rate and the shape factor have no correlation asymptotically [Amari and Nagaoka 2000, Ikeda 2005]. The most fundamental quantity used for characterizing a spike sequence is the firing rate, or its inverse, the mean ISI. Another metric that is orthogonal to the mean ISI is found to be the mean log ISI. It was also shown that the shape of a rate-rescaled distribution may be parameterized by the expectation of these metrics, $\log E[\mathrm{ISI}] - E[\log \mathrm{ISI}]$. It is noteworthy that the log ISI has heuristically been used for characterizing the firing patterns [Sigworth and Sine 1987, Selinger et al. 2007], and in particular, Dorval recently pointed out that with logarithmic partitioning of ISIs, firing rate changes become independent of firing pattern entropy [Dorval 2008]. The distribution that parametrically represents the mean ISI and the mean log ISI is the gamma distribution, which has frequently been adopted in neuroscience [Stein 1965, Teich et al., 1997, Baker and Lemon 2000]. The gamma distribution is also derived as the optimal ISI distribution under the energy constraint [Berger and Levy 2009]. From statistical point of view, the mean ISI and the mean log ISI are sufficient statistics for the gamma distribution, which means that the estimator for the gamma distribution only depends on these metrics. Because the Bayes estimator has a good statistical properties in terms of sufficiency and consistency, the variance of it asymptotically achieves the Cramér-Rao lower bound and it converges to the underlying parameters of the gamma distribution in the limit of large sample size [Lehmann and and Casella, 1998]. Therefore, the gamma distribution effectively captures firing patterns that can be characterized by those metrics.

We also examined the adaptability of gamma distribution in characterizing spike trains. In particular, we installed the gamma distribution into the Bayesian framework, and evaluated the accuracy of estimation of the rate and regularity of a spike sequence. The Bayesian method equipped with the gamma distribution works reasonably well in estimating time-dependent firing rate and regularity even when spike trains are derived from the log-normal or inverse-Gaussian distributions, implying the generality of the gamma distribution.

Note, however, that the simulation result does not guarantee that the gamma distribution always works well in any cases; if we have a prior knowledge about the type of the underlying distribution, it should be incorporated into estimations to gain statistical efficiency. To summarize, we proved that the gamma distribution is derived from the orthogonality of ISI metrics between the rate and regularity, which supports using the gamma distribution when we do not have a prior knowledge of the underlying ISI distribution.

Acknowledgements This study is supported in part by Grants-in-Aid for Scientific Research to SS from the MEXT Japan (20300083, 20020012). TS is supported by the Research Fellowship of the JSPS for Young Scientists. SK is supported by grants RO1 MH064537, RO1 EB005847 and RO1 NS050256.

Appendix

Numerical Algorithm

We construct a fast algorithm that may put the empirical Bayesian inference into practice. First, the marginalized likelihood function is maximized with the expectation maximization (EM) method. Under the hyperparameters determined with the EM method, the maximum *a posteriori* (MAP) estimates of time-dependent rate and regularity are obtained with the Newton-Raphson method. The following algorithm can be applied to any kind of distribution.

Maximization of marginal likelihood

We select the set of hyperparameters $\gamma = (\gamma_\lambda, \gamma_\kappa)^T$ so that the marginal likelihood Eq.(25) is maximized.

State space model

For the numerical marginalization path integral of Eq.(25), we discretize the time coordinate at every time a spike occurred. Then the spike occurrence can be treated as the state space model or the hidden Markov model, in which $\{\theta_i \equiv (\log \lambda(t_i), \log \kappa(t_i))^T\}_{i=0}^{n-1}$ are states or hidden variables, and $\{T_i\}_{i=0}^{n-1}$ are observations. The system model is given with the prior distribution Eq.(22),

$$p(\theta_{j+1}|\theta_j;\gamma) = \frac{1}{\sqrt{2\pi|R_j|}} \exp\left[-\frac{1}{2}(\theta_{j+1}-\theta_j)^T R_j^{-1}(\theta_{j+1}-\theta_j)\right], \tag{33}$$

where

$$R_j = \begin{pmatrix} \gamma_\lambda^2 T_j & 0 \\ 0 & \gamma_\kappa^2 T_j \end{pmatrix}. \tag{34}$$

The state equation for θ_i is represented as

$$\theta_{j+1} = \theta_j + w_j, \tag{35}$$

where w_j is the Gaussian noise with the mean 0, and variance-covariance matrix R_j. The observation probability is given by the distribution, in which parameters θ are varying in time,

$$p(T_j\,|\,\theta_j). \tag{36}$$

EM algorithm

The principle of the marginal likelihood maximization for selecting hyperparameters may be regarded as the maximum likelihood principle for selecting the parameters of the statistical model accompanied by the hidden variables. For the purpose of maximizing the likelihood, one may apply the expectation maximization (EM) algorithm [Dempster et al. 1977], in which the parameters of (marginal) likelihood Eq.(25) is obtained by iterating the Q function maximization,

$$Q(\gamma\,|\,\gamma^{(p)}) = E\left[\log p(\{T_j\}_{j=0}^{n-1}, \{\theta_j\}_{j=0}^{n-1};\gamma)\,|\,\{T_i\}_{i=0}^{n-1};\gamma^{(p)}\right], \tag{37}$$

where $\gamma^{(p)} = (\gamma_\lambda^{(p)}, \gamma_\kappa^{(p)})^T$ is the (hyper)parameters of pth iteration. The $p+1$st γ is obtained by maximizing this Q function, or equivalently the conditions for $dQ/d\gamma = 0$,

$$\gamma_\lambda^{(p+1)} = \sqrt{\frac{1}{n-1}\sum_{j=0}^{n-2}\frac{1}{T_j}E\left[(\log\lambda_{j+1}-\log\lambda_j)^2\,|\,\{T_i\}_{i=0}^{n-1};\gamma^{(p)}\right]}, \tag{38}$$

$$\gamma_\kappa^{(p+1)} = \sqrt{\frac{1}{n-1}\sum_{j=0}^{n-2}\frac{1}{T_j}E\left[(\log\kappa_{j+1}-\log\kappa_j)^2\,|\,\{T_i\}_{i=0}^{n-1};\gamma^{(p)}\right]}. \tag{39}$$

Kalman filter and smoothing algorithm

In carrying out the EM algorithms, Eqs.(38) and (39), we need the conditional probability distribution given ISIs $\{T_i\}_{i=0}^{n-1}$ and hyperparameters $\gamma^{(p)}$. Under the Gaussian assumption, the conditional distribution can be obtained with the Kalman filter and smoothing [Smith and Brown 2003, Truccolo et al. 2005] applied to the mean, variance and covariance of the distribution, respectively defined by

$$\theta_{j|l} \equiv E\left[\theta_j \mid \{T_i\}_{i=0}^{l}; \gamma^{(p)}\right], \tag{40}$$

$$V_{j|l} \equiv E\left[\{\theta_j - \theta_{j|l}\}\{\theta_j - \theta_{j|l}\}^T \mid \{T_i\}_{i=0}^{l}; \gamma^{(p)}\right], \tag{41}$$

$$V_{j,k|l} \equiv E\left[\{\theta_j - \theta_{j|l}\}\{\theta_k - \theta_{k|l}\}^T \mid \{T_i\}_{i=0}^{l}; \gamma^{(p)}\right]. \tag{42}$$

[prediction]: The mean θ and variance V are iterated in the prediction process as

$$\theta_{j|j-1} = \theta_{j-1|j-1}, \tag{43}$$

$$V_{j|j-1} = V_{j-1|j-1} + R_{j-1}. \tag{44}$$

[filtering]: The algorithm recursively computes the posterior probability density [Mendel 1995, Kitagawa and Gersh 1996],

$$p(\theta_j \mid \{T_i\}_{i=0}^{j}) = \frac{p(\theta_j \mid \{T_i\}_{i=0}^{j-1})\, p(T_j \mid \theta_j, \{T_i\}_{i=0}^{j-1})}{p(T_j \mid \{T_i\}_{i=0}^{j-1})} \propto p(\theta_j \mid \{T_i\}_{i=0}^{j-1})\, p(T_j|\theta_j). \tag{45}$$

By assuming $p(\theta_j \mid \{T_i\}_{i=0}^{j-1})$ to be Gaussian, the log posterior probability density can be given by

$$\log p(\theta_j \mid \{T_i\}_{i=0}^{j}) = -\frac{1}{2}(\theta_j - \theta_{j|j-1})^T V_{j|j-1}^{-1}(\theta_j - \theta_{j|j-1}) + \log p(T_j|\theta_j) + \text{const.} \tag{46}$$

By assuming further the posterior probability density $p(\theta_j \mid \{T_i\}_{i=0}^{j})$ to be Gaussian, the mean $\theta_{j|j}$ and variance $V_{j|j}$ are given by

$$\frac{d}{d\theta_j} \log p(\theta_j \mid \{T_i\}_{i=0}^{j})\Big|_{\theta_j = \theta_{j|j}} = 0, \tag{47}$$

$$V_{j|j} = \left[\frac{d^2}{d\theta_j^2} \log p(\theta_j \mid \{T_i\}_{i=0}^{j})\Big|_{\theta_j = \theta_{j|j}}\right]^{-1}. \tag{48}$$

Note that the concavity of the log posterior probability density (46) is not guaranteed. A reasonable posterior mode $\theta_{j|j}$ can be found, however, by choosing $\theta_{j|j-1}$ as a starting value of an iterative algorithm in solving Eq. (47).

[fixed interval smoothing algorithm]: The smoothing can be carried out with the iteration [Ansley and Kohn 1982],

$$\theta_{j|n} = \theta_{j|j} + A_j(\theta_{j+1|n} - \theta_{j+1|j}), \tag{49}$$

$$V_{j|n} = V_{j|j} + A_j(V_{j+1|n} - V_{j+1|j})A_j^T, \tag{50}$$

where

$$A_j = V_{j|j} V_{j+1|j}^{-1}. \tag{51}$$

[covariance algorithm]: The covariance is given by the combination of variances [de Jong and Mackinnon 1988],

$$V_{j+1,j|n} = A_j V_{j+1|n}. \tag{52}$$

Derivation of the exact MAP solutions

With the hyperparameters determined with the EM algorithms, we evaluate the MAP estimates of time-dependent rate and regularity. The estimates of the states $\{\theta_i\}_{i=0}^{n-1}$ are obtained by maximizing the log *a posteriori* distribution,

$$\begin{aligned}\log p(\{\theta_i\}_{i=0}^{n-1} \mid \{t_i\};\gamma) &= \log p(\{t_i\} \mid \{\theta_i\}_{i=0}^{n-1}) + \log p(\{\theta_i\}_{i=0}^{n-1};\gamma) + \text{const.} \\ &= \sum_{j=0}^{n-1} \log p(T_j \mid \theta_j) + \sum_{j=0}^{n-2} \log p(\theta_{j+1} \mid \theta_j;\gamma) + \text{const.}\end{aligned} \tag{53}$$

The log *a posteriori* distribution can be rewritten in matrix expression with the states $\Theta \equiv (\theta_0^T, \theta_1^T, \cdots, \theta_{n-1}^T)^T$ as

$$S(\Theta) \equiv \log p(\{\theta_i\}_{i=0}^{n-1} \mid \{t_i\};\gamma) = l(\Theta) - \frac{1}{2}\Theta^T R^{-1}\Theta, \tag{54}$$

where

$$l(\Theta) \equiv \sum_{j=0}^{n-1} \log p(T_j \mid \theta_j), \tag{55}$$

$$R^{-1} \equiv \begin{bmatrix} R_0^{-1} & -R_0^{-1} & 0 & 0 & \cdots & 0 \\ -R_0^{-1} & R_0^{-1}+R_1^{-1} & -R_1^{-1} & 0 & \cdots & 0 \\ 0 & -R_1^{-1} & R_1^{-1}+R_2^{-1} & -R_2^{-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -R_{n-3} & R_{n-3}^{-1}+R_{n-2}^{-1} & -R_{n-2}^{-1} \\ 0 & \cdots & \cdots & 0 & -R_{n-2}^{-1} & R_{n-2}^{-1} \end{bmatrix}. \tag{56}$$

The estimates of the states $\hat{\Theta}$ satisfy

$$\nabla_\Theta S(\hat{\Theta}) = 0. \tag{57}$$

The Newton-Raphson method [Press et al. 1992] can be used to efficiently solve Eq. (57) by choosing the provisional solution $\tilde{\Theta}$ obtained under the Gaussian approximation with the Kalman filter and smoothing as a starting value. We repeat the iteration process,

$$\nabla\nabla_\Theta S(\tilde{\Theta})\delta\Theta = -\nabla_\Theta S(\tilde{\Theta}). \tag{58}$$

We recursively add the $\delta\Theta$ obtained from this equation to revise the provisional solutions $\tilde{\Theta}$ until the solutions converge. As $\nabla\nabla_\Theta S(\tilde{\Theta})$ is a block tridiagonal matrix, the $\delta\Theta$ of the recurrence equation is obtained by the $O(n)$ computational complexity.

Once the MAP solution $\hat{\Theta}$ is obtained, the posterior covariance is approximated by $[-\nabla\nabla_\Theta S(\hat{\Theta})]^{-1}$ which can be used to obtain confidence intervals.

References

[Abeles 1982] Abeles, M. (1982). Quantification, smoothing, and confidence-limits for single-units histograms. *Journal of Neuroscience Methods*, 5(4), 317-325.

[Amari and Nagaoka 2000] Amari, S., and Nagaoka, H. (2000). *Methods of information geometry.* Oxford: Oxford University Press.

[Ansley and Kohn 1982] Ansley, C. F., and Kohn, R. (1982). A geometrical derivation of the fixed interval smoothing algorithm. *Biometrika*, 69(2), 486-487.

[Baker and Lemon 2000] Baker, S. N., and Lemon, R. N. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *Journal of Neurophysiology*, 84(4), 1770-1780.

[Berger and Levy 2009] Berger, T., and Levy, W. B. (2009). Information transfer by energy-efficient neurons. *Proc. IEEE ISIT 2009*, 1584-1588.

[Bialek et al. 1996] Bialek, W., Callan, C. G., and Strong, S. P. (1996). Field Theories for Learning Probability Distributions. *Physical review letters*, 77, 4693-4697.

[Carlin and Louis 2000] Carlin, B. P., and Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis, 2nd ed.* New York: Chapman and Hall/CRC.

[Churchland et al. 2006] Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G., and Shenoy K. V. (2006). Neural variability in premotor cortex provides a signature of motor preparation. *Journal of Neuroscience*, 26(14), 3697-3712.

[Cunningham et al. 2008] Cunningham, J. P., Yu, B. M., Shenoy, K. V., and Sahani, M. (2008). Inferring neural firing rates from spike trains using Gaussian processes. *Advances in Neural Information Processing Systems*, 20, 329-336.

[Davies et al. 2006] Davies, R. M., Gerstein, G. L., and Baker, S. N. (2006). Measurement of time-dependent changes in the irregularity of neural spiking. *Journal of Neurophysiology*, 96(2), 906-918.

[Dayan and Abbott 2001] Dayan, P., and Abbott, L. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems.* Cambridge, MA: MIT Press.

[de Jong and Mackinnon 1988] de Jong, P., and Mackinnon, M. J. (1988). Covariances for smoothed estimates in state space models. *Biometrika*, 75(3), 601-602.

[Dempster et al. 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1), 1-38.

[DiMatteo et al. 2001] DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve-fitting with freeknot splines. *Biometrika*, 88(4), 1055-1071.

[Dorval 2008] Dorval, A. D. (2008). Probability distributions of the logarithm of inter-spike intervals yield accurate entropy estimates from small datasets. *Journal of Neuroscience Methods*, 173(1), 129-139.

[Gabbiani and Koch 1998] Gabbiani, F., and Koch, C. (1998). *Methods in neuronal modeling : from ions to networks,* 2nd ed, chapter 9. Cambridge, MA: MIT Press.

[Gerstein and Kiang 1960] Gerstein, G. L., and Kiang, N.Y.-S. (1960). An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophysical Journal*, 1(1), 15-28.

[Holt et al. 1996] Holt, G. R., Softky, W. R., Koch, C., and Douglas, R. J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysiology*, 75(5), 1806-1814.

[Ikeda 2005] Ikeda, K. (2005). Information geometry of interspike intervals in spiking neurons. *Neural Computation*, 17(12), 2719-2735.

[Jaynes 1957] Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106(4), 620-630.

[Kapur 1989] Kapur, J. N. (1989). *Maximum-entropy models in science and engineering.* New York: Wiley.

[Kara et al. 2000] Kara, P., Reinagel, P., and Reid, R. C. (2000). Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27(3), 635-646.

[Kass et al. 2005] Kass, R. E., Ventura V., and Brown, E. N. (2005). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*, 94, 8-25.

[Kitagawa and Gersh 1996] Kitagawa, G., and Gersh W. (1996). *Smoothness priors analysis of time series.* New York: Springer-Verlag.

[Kostal and Lansky 2006] Kostal, L., and Lansky, P. (2006). Classification of stationary neuronal activity according to its information rate. *Network: Computation in Neural Systems*, 17(2), 193-210.

[Koyama and Shinomoto 2005] Koyama, S., and Shinomoto, S. (2005). Empirical Bayes interpretations of random point events. *Journal of Physics A: Mathematical and Theoretical*, 38(29), L531-L537.

[Koyama et al. 2007] Koyama, S., Shimokawa, T., and Shinomoto, S (2007). Phase transitions in the estimation of event-rate: A path integral analysis. *Journal of Physics A: Mathematical and Theoretical*, 40(20), F383-F390.

[Lehmann and and Casella, 1998] Lehmann, E. L., and Casella, G. (1998). *Theory of point estimation.* New York: Springer.

[MacKay 1992] MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Computation*, 4(3), 415-447.

[Mendel 1995] Mendel, J. M. (1995). *Lessons in estimation theory for signal processing, communications, and control.* New Jersey: Prentice Hall.

[Mitchell et al. 2007] Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron*, 55, 131-141.

[Miura et al. 2006] Miura, K., Okada, M., and Amari, S. (2006). Estimating spiking irregularities under changing environments. *Neural Computation*, 18(10), 2359-2386.

[Nawrot et al. 2008] Nawrot, M. P., Boucsein, C., Rodriguez-Molina, V., Riehle, A., Aertsen, A., and Rotter, S. (2008). Measurement of variability dynamics in cortical spike trains. *Journal of Neuroscience Methods*, 169(2), 374-390.

[Nemenman and Bialek, 2002] Nemenman, I., and Bialek, W. (2002). Occam factors and model independent Bayesian learning of continuous distributions. *Physical Review E*, 65(2), 026137.

[Press et al. 1992] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical recipes in C: the art of scientific computing,* 2nd ed, chapter 17. Cambridge: Cambridge University Press.

[Rasmussen and Williams 2006] Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning.* Cambridge, MA: MIT Press.

[Reich et al. 1998] Reich, D. S., Victor, J. D., and Knight, B. W. (1998). The power ratio and the interval map: spiking models and extracellular recordings. *Journal of Neuroscience*, 18(23), 10090-10104.

[Schneidman et al. 2006] Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087), 1007-1012.

[Selinger et al. 2007] Selinger, J. V., Kulagina, N. V., O'Shaughnessy, T. J., Ma, W., Pancrazio, J. J. (2007). Methods for characterizing interspike intervals and identifying bursts in neuronal activity. *Journal of Neuroscience Methods*, 162(1-2), 64-71.

[Shimazaki and Shinomoto 2007] Shimazaki, H., and Shinomoto, S (2007). A method for selecting the bin size of a time histogram. *Neural Computation*, 19(6), 1503-1527.

[Shimokawa and Shinomoto 2009] Shimokawa, T., and Shinomoto, S (2009). Estimating instantaneous irregularity of neuronal firing. *Neural Computation*, 21(7), 1931-1951.

[Shinomoto et al. 2003] Shinomoto, S., Shima, K., and Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural Computation*, 15(12), 2823-2842.

[Shinomoto et al. 2005] Shinomoto, S., Miyazaki, Y., Tamura, H., and Fujita, I. (2005). Regional and laminar differences in in vivo firing patterns of primate cortical neurons. *Journal of Neurophysiology*, 94(1), 567-575.

[Shinomoto et al. 2009] Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., Fujita, I., Tamura, H., Doi, T., Kawano, K., Inaba, N., Fukushima, K., Kurkin, S., Kurata, K., Taira, M., Tsutsui, K., Komatsu, H., Ogawa, T., Koida, K., Tanji, J., and Toyama K. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Computational Biology*, 5, e1000433.

[Sigworth and Sine 1987] Sigworth, F. J., and Sine, S. M. (1987). Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophysical Journal*, 52(6), 1047-1054.

[Smith and Brown 2003] Smith, A. C., and Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Computation*, 15(5), 965-991.

[Stein 1965] Stein, R. B. (1965). A theoretical analysis of neuronal variability. *Biophysical Journal*, 5(2), 173-194.

[Teich et al., 1997] Teich, M. C., Heneghan, C., Lowen, S. B., Ozaki, T., and Kaplan, E. (1997). Fractal character of the neural spike train in the visual system of the cat. *Journal of the Optical Society of America*, 14(3), 529-546.

[Truccolo et al. 2005] Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2), 1074-1089.