LETTER **Communicated by Jakob Macke**

Information Transmission Using Non-Poisson Regular Firing

Shinsuke Koyama
skoyama@ism.ac.jp
Department of Statistical Modeling, Institute of Statistical Mathematics, Tokyo 190-8562, Japan

Takahiro Omi
omitakahiro@gmail.com
FIRST, Aihara Innovative Mathematical Modelling Project, Japan Science and Technology Agency, Tokyo 153-8505, Japan, and Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan

Robert E. Kass
kass@stat.cmu.edu
Department of Statistics and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Shigeru Shinomoto
shinomoto@scphys.kyoto-u.ac.jp
Department of Physics, Kyoto University, Kyoto 606-8502, Japan

In many cortical areas, neural spike trains do not follow a Poisson process. In this study, we investigate a possible benefit of non-Poisson spiking for information transmission by studying the minimal rate fluctuation that can be detected by a Bayesian estimator. The idea is that an inhomogeneous Poisson process may make it difficult for downstream decoders to resolve subtle changes in rate fluctuation, but by using a more regular non-Poisson process, the nervous system can make rate fluctuations easier to detect. We evaluate the degree to which regular firing reduces the rate fluctuation detection threshold. We find that the threshold for detection is reduced in proportion to the coefficient of variation of interspike intervals.

1 Introduction

Neural coding has been central to computational neuroscience for more than 40 years (see, e.g., Perkel & Bullock, 1968). The earliest work on the physiological significance of spike trains made clear the importance of rate coding, according to which information about stimuli or actions is contained

in fluctuating firing rates. Recently alternative coding schemes beyond the rate code have been examined by decoding sensory or behavior signals from recorded spike trains (Barbieri, Quirk, Frank, Wilson, & Brown, 2001; Jacobs et al., 2009; Pillow, Paninski, Uzzell, Simoncelli, & Chichilnisky, 2005). These studies showed that if spike counts or fluctuating firing rates were used as inputs, a Bayesian decoder could not sufficiently reproduce the signals, but when more precise firing timing information was included, it could. In these analyses, timing information was incorporated by non-Poisson spike train models. These investigations raise the question, What are the neural coding advantages of non-Poisson spiking? It could be that additional information is carried in non-Poisson characteristics of a spike train (Burns & Pritchard, 1964; Davies, Gerstein, & Baker, 2006; Lundstrom & Fairhall, 2006; Olypher, Lansky, & Fenton, 2002; Ratliff, Hartline, & Lange, 1968). Alternatively, temporal variation of the rate might be transmitted more accurately by using non-Poisson regular firing (Cunningham, Yu, Shenoy, & Sahani, 2008; Cunningham, Gilja, Ryu, & Shenoy, 2009; Shimokawa & Shinomoto, 2009; Grun, 2009; Omi & Shinomoto, 2011). In this study, we examine the latter possibility by testing the detectability of the rate fluctuation in inhomogeneous non-Poisson spike trains.

In the problem of estimating the rate of an inhomogeneous Poisson spike train, there is a threshold of rate fluctuation below which rate estimators, such as an optimized histogram with respect to minimizing the mean squared error or a Bayesian rate estimator, cannot discern rate fluctuations (Koyama & Shinomoto, 2004; Koyama, Shimokawa, & Shinomoto, 2007). The Bayesian rate estimator in previous studies was based on an inhomogeneous Poisson process likelihood with a roughness-penalizing prior on the rate process, where the hyperparameter was estimated by maximizing the marginal likelihood or the evidence. (See MacKay, 1992, and Kass & Raftery, 1995, for use of the evidence in model selection.) Based on this model, when the estimated hyperparameter suggests a constant-rate process for a spike train derived from a temporally fluctuating rate, it would be concluded that the rate fluctuation is undetectable. In this letter, we reformulate those results, replacing the Poisson spike trains with non-Poisson regular processes, and we evaluate the degree to which the threshold of undetectable fluctuations is thereby reduced so that information transmission is improved. In order to obtain analytical results, we assume that rate fluctuations have a long timescale so that the firing rate in successive spikes does not change drastically. The analytical results are then compared with numerical simulations.

2 Encoding an Inhomogeneous Rate

We first define the stochastic processes of generating spikes, starting from the homogeneous Poisson process, then adding inhomogeneity, and finally adding non-Poisson response statistics. Note that throughout this letter,

inhomogeneity stands for spike trains with a time-dependent firing rate $\lambda(t)$, whereas *non-Poissonian* means a departure from a Poisson process characterized by the shape parameter κ of the gamma interspike interval (ISI) distribution introduced below.

- Homogeneous Poisson Process. The process of generating spikes randomly in time at a constant rate λ can be realized by repeating Bernoulli trials with a small probability $\lambda\delta t (\ll 1)$ in every small interval δt. The probability of having no spikes for the first n intervals and finally having a spike at the $(n+1)$st interval is given by $(1-\lambda\delta t)^n \lambda\delta t$. By taking the infinitesimal limit of δt, fixing $t = n\delta t$ finite, the probability is given by

$$p(t|\lambda)\,\delta t = \lim_{n\to\infty} (1-\lambda t/n)^n\,\lambda\,\delta t = \exp(-\lambda t)\lambda\,\delta t. \tag{2.1}$$

  Thus, this process is identical to a process of drawing ISIs independently from an exponential distribution $\lambda \exp(-\lambda t)$. If the ISIs are drawn from a distribution function different from the exponential function, the process is called the *homogeneous non-Poisson renewal process*.
- Inhomogeneous Poisson Process. By allowing the occurrence rate to fluctuate in time as $\lambda(t)$, the probability of having no spike for the first n intervals and finally having a spike at the $(n+1)$st interval is given by $\prod_{j=1}^{n}(1-\lambda(j\delta t)\delta t)\lambda(n\delta t)\delta t$. When the limit of the infinitesimal interval δt is taken, the probability of having an interspike interval t can be obtained as

$$\begin{aligned} p(t_i|\{\lambda(t)\})\delta t &= \lim_{n\to\infty}\prod_{j=1}^{n}\big(1-\lambda(jt/n)t/n\big)\,\lambda(t)\,\delta t \\ &= \exp\left(-\int_0^t \lambda(t')dt'\right)\lambda(t)\,\delta t. \end{aligned} \tag{2.2}$$

  It follows from equation 2.2 that the inhomogeneous Poisson process can be mapped to the homogeneous Poisson process if the time axis is suitably scaled (Ogata, 1988; Daley & Vere-Jones, 1988; Reich, Victor, & Knight, 1998; Oram, Wiener, Lestienne, & Richmond, 1999; Barbieri et al., 2001; Brown, Barbieri, Ventura, Kass, & Frank, 2001; Smith & Brown, 2003; Koyama & Shinomoto, 2005) as

$$x = \Lambda(t) \equiv \int_0^t \lambda(t')\,dt', \tag{2.3}$$

  or,

$$dx = \lambda(t)\,dt. \tag{2.4}$$

- Inhomogeneous Non-Poisson Process. Using the time-scaling transformation, equation 2.4, the homogeneous non-Poisson renewal process given by a nonexponential ISI distribution can be transformed into an inhomogeneous non-Poisson process. Given a time-dependent rate process $\lambda(t)$ with spike aftereffects specified by a nonexponential ISI distribution function $f(x)$, the probability of having an event in a small time interval $[x, x + dx]$ given a preceding event at 0 can be represented as

$$f(x)\,dx = f(\Lambda(t) - \Lambda(t_0))\lambda(t)\,dt. \tag{2.5}$$

  Here we have assumed that the shape of the ISI distribution $f(x)$, which defines the firing irregularity, is unchanged in time. This is in agreement with the empirical fact that the degree of irregularity of neuronal firing is generally maintained in vivo (Shinomoto, Shima, & Tanji, 2003; Shinomoto, Miura, & Koyama, 2005; Shinomoto et al., 2009), while the firing rate $\lambda(t)$ is changing in time.

On the basis of the time-rescaling theory, an inhomogeneous non-Poisson spike train can be generated by the following process:

1. Generate a uniform non-Poisson (renewal) sequence. Draw ISIs $\{x_1, x_2, \ldots, x_n\}$ independently from a given distribution function $f(x)$ and arrange them sequentially to form a spike train. Here the ith spike time is given by summation of the previous ISIs as $y_i = \sum_{j=1}^{i} x_j$.
2. Rescale the time axis to conform to given rate $\lambda(t)$. Transform the original sequence $\{y_1, y_2, \ldots, y_n\}$ into another sequence $\{t_1, t_2, \ldots, t_n\}$ according to $t_i = \Lambda^{-1}(y_i)$, where $\Lambda^{-1}(y)$ is the inverse of the function $\Lambda(t)$ defined by equation 2.3.

Given time-dependent rate $\lambda(t)$ with aftereffects specified by nonexponential ISI distribution function $f(x)$, the probability density for the occurrence of spikes at $\{t_i\} = \{t_1, t_2, \ldots, t_n\}$ can be represented as

$$p(\{t_i\}|\{\lambda(t)\}) = \prod_{i=1}^{n-1} \lambda(t_{i+1}) f(\Lambda(t_{i+1}) - \Lambda(t_i)), \tag{2.6}$$

where t_i is the time of ith spike. Note that this manner of representing the non-Poisson spiking is similar to that of inhomogeneous Markov interval processes (Kass & Ventura, 2001; Koyama & Kass, 2008), in that both models confine the non-Poisson characteristics within an interspike correlation between two consecutive spikes.

The ISI distribution used in this study is a parametric family of gamma distribution functions,

$$f(x) = f_\kappa(x) = \kappa(\kappa x)^{\kappa-1} \exp(-\kappa x)/\Gamma(\kappa), \tag{2.7}$$

where $\Gamma(\kappa) = \int_0^\infty x^{\kappa-1} \exp(-x)\, dx$ is the gamma function. Accordingly, the shape of the ISI distribution can be controlled by κ, keeping the scale factor or the mean ISI to be unity, $\int_0^\infty x f_\kappa(x) = 1$; the shape factors $\kappa > 1, = 1$, and < 1 represent regular, (Poisson) random, and bursty firings, respectively.

3 Decoding the Rate from a Spike Train

Next, we construct an algorithm to decode the rate from a spike train. The inverse probability of time-dependent rate $\lambda(t)$ given the spike train $\{t_i\}$ can be obtained according to Bayes' theorem:

$$p(\{\lambda(t)\}|\{t_i\}) = \frac{p(\{t_i\}|\{\lambda(t)\})p(\{\lambda(t)\})}{p(\{t_i\})}. \tag{3.1}$$

We employ the maximum a posteriori (MAP) estimate for inference of the firing rate.[1] The prior distribution of the latent rate process is chosen such that the large gradients of $\lambda(t)$ are penalized with

$$p(\{\lambda(t)\}) = p_\gamma(\{\lambda(t)\}) = \frac{1}{Z(\gamma)} \exp\left(-\frac{1}{2\gamma^2}\int_0^T (d\lambda/dt)^2 dt\right). \tag{3.2}$$

Here, $Z(\gamma)$ is the normalization constant given by

$$\begin{aligned} Z(\gamma) &= \int D\{\lambda(t)\} \exp\left(-\frac{1}{2\gamma^2}\int_0^T (d\lambda/dt)^2 dt\right) \\ &= \frac{1}{\sqrt{2\pi\gamma^2 T}} \exp\left[-\frac{\{\lambda(T)-\lambda(0)\}^2}{2\gamma^2 T}\right], \end{aligned} \tag{3.3}$$

where $\int D\{\lambda(t)\}$ represents integration over all possible latent rate processes or the Wiener integral over all paths of $\lambda(t)$ (Karatzas & Shreve, 1997). Equation 3.2, which defines a gaussian process whose covariance function is $\mathrm{cov}[\lambda(t), \lambda(s)] = \gamma^2 \min(t, s)$ (i.e., a Brownian motion), can be called the roughness-penalizing prior, the log of which is the same as the smoothing-spline penalization with the first-order derivative (Ramsay & Silverman, 2010). The MAP estimate, then, corresponds to the penalized maximum likelihood estimate. Here, the roughness hyperparameter γ controls the

[1] The MAP estimate does not involve a loss function with respect to which a posterior estimate is optimized, but instead is chosen so that the posterior probability is maximized. However, when the posterior distribution is unimodal and has a sharp peak, as in the case of our analysis below, the MAP estimate provides a good approximation to the posterior expectation (Koyama, Perez-Bolde, Shalizi, & Kass, 2010), which minimizes the mean squared loss between the true and estimated firing rates.

smoothness of the time-dependent rate $\lambda(t)$; with small values of γ, the model requires the smooth rate process and vice versa. Note that γ has units of $s^{-3/2}$ if $\lambda(t)$ has units of s^{-1}.

We may construct a Bayesian rate decoder based on the non-Poisson spike aftereffects by adopting the non-Poisson spike generation process (see equation 2.6) as

$$p(\{t_i\}|\{\lambda(t)\}) = p_\kappa(\{t_i\}|\{\lambda(t)\}) = \prod_{i=1}^{n-1} \lambda(t_{i+1}) f_\kappa(\Lambda(t_{i+1}) - \Lambda(t_i)) \tag{3.4}$$

into the Bayesian formula, equation 3.1, while the denominator of the Bayesian formula is given by marginalizing the joint probability over all possible paths of $\lambda(t)$,

$$p(\{t_i\}) = p_{\kappa,\gamma}(\{t_i\}) = \int D\{\lambda(t)\} p_\kappa(\{t_i\}|\{\lambda(t)\}) p_\gamma(\{\lambda(t)\}). \tag{3.5}$$

Upon applying the empirical Bayes method (MacKay, 1992; Bruce & Saad, 1994; Carlin & Louis, 2000; Rasmussen & Williams, 2006), the hyperparameters γ and κ can be determined so that this "marginal likelihood" or the "evidence" is maximized according to

$$(\hat{\gamma}, \hat{\kappa}) = \arg\max_{\gamma,\kappa} p_{\kappa,\gamma}(\{t_i\}). \tag{3.6}$$

In Bayesian analysis, in principle the hyperparameters may be determined by maximizing the posterior probability of them, which is computed by combining the prior probability over the hyperparameters with the evidence via Bayes' theorem. Having no basis on which to formulate a prior distribution, we instead maximize the marginal likelihood. As MacKay (1992) and Kass and Raftery (1995), pointed out, the marginal likelihood, or evidence, naturally embodies Occam's razor.

The posterior distribution, equation 3.1, can be continuously extended to $\gamma = 0$, in which case the posterior probability for the rate process, $p_{\kappa,\gamma}(\{\lambda(t)\}|\{t_i\})$, consequently indicates a constant rate, $d\lambda(t)/dt = 0$. When a train of an infinite number of spikes is analyzed, the distribution of the constant rate $\lambda(t) = \lambda$ becomes a delta function located at a value of the mean firing rate. Thus, if the optimized roughness hyperparameter vanishes, $\hat{\gamma} = 0$, it should be interpreted that the Bayesian rate decoder has uniquely selected a constant rate process.

4 Performing the Marginalizing Integration

For the inhomogeneous Poisson process, the marginalizing integration over all possible latent rate paths $\int D\{\lambda(t)\}$ in equation 3.5 was solved for some

given rate processes using the path integral method (Koyama et al., 2007). Here we reformulate the analytical method and extend the marginalizing integration method to make it applicable to non-Poisson processes.

4.1 Evaluation of the Log-Likelihood Function. From equations 2.7 and 3.4, the log likelihood is explicitly given by

$$\begin{aligned}\log p_\kappa(\{t_i\}|\{\lambda(t)\}) &= \sum_i [\log\lambda(t_i) + \kappa\log\kappa + (\kappa-1)\log(\Lambda(t_i) - \Lambda(t_{i-1})) \\ &\quad - \kappa(\Lambda(t_i) - \Lambda(t_{i-1})) - \log\Gamma(\kappa)] \\ &= \sum_i [\log\lambda(t_i) + (\kappa-1)\log(\Lambda(t_i) - \Lambda(t_{i-1}))] \\ &\quad + n\kappa\log\kappa - \kappa\int_0^T \lambda(t)\,dt - n\log\Gamma(\kappa). \end{aligned} \tag{4.1}$$

By introducing the "averaged" firing rate in the ith ISI,

$$\bar{\lambda}_i \equiv \frac{\Lambda(t_i) - \Lambda(t_{i-1})}{t_i - t_{i-1}}, \tag{4.2}$$

the rate fluctuation and the ISI in the term $\log(\Lambda(t_i) - \Lambda(t_{i-1}))$ are separated as

$$\log(\Lambda(t_i) - \Lambda(t_{i-1})) = \log\bar{\lambda}_i + \log(t_i - t_{i-1}). \tag{4.3}$$

Assuming that the timescale of the rate fluctuation is longer than the mean ISI so that $\bar{\lambda}_i$ is approximated to $\lambda(t_i)$, the log likelihood, equation 4.1, is rewritten as[2]

$$\begin{aligned}\log p_\kappa(\{t_i\}|\{\lambda(t)\}) &= \kappa\sum_i \log\lambda(t_i) + n\kappa\log\kappa - n\log\Gamma(\kappa) \\ &\quad - \kappa\int_0^T \lambda(t)\,dt + (\kappa-1)\sum_i \log(t_i - t_{i-1}). \end{aligned} \tag{4.4}$$

[2]Suppose that the timescale of rate fluctuation $\lambda(t)$ is given by τ. Then,

$$\lambda(t_i) = \lambda(t_{i-1}) + \frac{d\lambda(t_{i-1})}{d(t_{i-1}/\tau)}\frac{t_i - t_{i-1}}{\tau} + O\left(\left(\frac{t_i - t_{i-1}}{\tau}\right)^2\right),$$

from which the error is estimated as $\lambda(t_i) - \bar{\lambda}_i \sim (t_i - t_{i-1})/\tau$. Our analysis is hence valid under the condition in which τ is large enough compared with the mean ISI so that this error is negligible.

We decompose the rate $\lambda(t)$ into the mean μ and fluctuation $x(t)$ as

$$\lambda(t) = \mu + x(t). \tag{4.5}$$

Accordingly, the log-likelihood function is decomposed into two parts, as

$$\log p_\kappa(\{t_i\}|\{\lambda(t)\}) = \mathcal{L} + \mathcal{I}, \tag{4.6}$$

where $\mathcal{L}$ is the log-likelihood function of the gamma interval distribution,

$$\begin{aligned}\mathcal{L} = &-T\kappa\mu + n\kappa\log\mu + n\kappa\log\kappa - n\log\Gamma(\kappa)\\ &+ (\kappa-1)\sum_i \log(t_i - t_{i-1}),\end{aligned} \tag{4.7}$$

whereas $\mathcal{I}$ is given by

$$\mathcal{I} = -\int_0^T \left[\kappa x(t) - \kappa\sum_i \delta(t-t_i)\log\left(1+\frac{x(t)}{\mu}\right)\right]dt. \tag{4.8}$$

$\mathcal{L}$ and $\mathcal{I}$ represent the components of mean rate and fluctuation around the mean rate, respectively.

4.2 The Path Integral Method. Substituting equations 3.2 and 4.6 into equation 3.5 and taking into account the change of variable from $\lambda(t)$ to $x(t)$ in equation 4.5, the marginal likelihood function is written as

$$\begin{aligned}p_{\gamma,\kappa}(\{t_i\}) &= \int D\{x(t)\}\exp(\mathcal{L}+\mathcal{I})\frac{1}{Z(\gamma)}\exp\left(-\frac{1}{2\gamma^2}\int_0^T \dot{x}^2 dt\right)\\ &= \frac{1}{Z(\gamma)}e^{\mathcal{L}}\mathcal{F}.\end{aligned} \tag{4.9}$$

Here, the contribution of rate fluctuation can be represented in the form of a path integral (Feynman & Hibbs, 1965; Kleinert, 2009),

$$\mathcal{F} = \int\exp\left[-\int_0^T L(\dot{x},x)\,dt\right]D\{x(t)\}, \tag{4.10}$$

where $L(\dot{x},x)$ is a Lagrangian of the form,

$$L(\dot{x},x) = \frac{1}{2\gamma^2}\dot{x}^2 + \kappa x(t) - \kappa\sum_i \delta(t-t_i)\log\left(1+\frac{x(t)}{\mu}\right). \tag{4.11}$$

The MAP estimate of the rate fluctuation $\hat{x}(t)$ is obtained by taking the minimum of the action integral $S[x(t)] \equiv \int_0^T L(\dot{x}, x)\, dt$ in equation 4.10. The extremum condition for S is expressed by the variational equation $\delta S = 0$, and an integration by part in δS with fixed boundary values leads to the Euler-Lagrange equation:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0. \tag{4.12}$$

Note that it can also be shown that the quadratic form of the action integral $\delta^2 S$ is nonnegative (Schulman, 2005), and thus the solution of equation 4.12 is guaranteed to be the MAP estimate $\hat{x}(t)$. In the following analysis, we consider a long spike train $T \gg 1$, so that the boundary effect, which may be caused by a particular choice of the boundary values in deriving the Euler-Lagrange equation, 4.12, is negligible.

By considering the deviation from the MAP path as $x(t) = \hat{x}(t) + \eta(t)$, $(\eta(0) = \eta(T) = 0)$, and approximating the action integral to a range of the second-order terms with respect to the deviation $\eta(t)$,

$$\begin{aligned}\int_0^T & L(\dot{x}, x)\, dt \\ &\approx \int_0^T L(\dot{\hat{x}}, \hat{x})\, dt + \frac{1}{2}\int_0^T \left(\frac{\partial^2 L}{\partial \dot{x}^2}\dot{\eta}^2 + 2\frac{\partial^2 L}{\partial \dot{x}\partial x}\dot{\eta}\eta + \frac{\partial^2 L}{\partial x^2}\eta^2\right)dt,\end{aligned} \tag{4.13}$$

where we ignored the terms in $O(\eta^3)$, the path integral can be performed analytically as

$$\mathcal{F} = Re^{-S[\hat{x}(t)]}, \tag{4.14}$$

where $\exp\{-S[\hat{x}(t)]\}$ represents the contribution of the mode to the path integral, whereas R represents that of quadratic derivation:

$$R = \int \exp\left[-\frac{1}{2}\int_0^T \left(\frac{\partial^2 L}{\partial \dot{x}^2}\dot{\eta}^2 + 2\frac{\partial^2 L}{\partial \dot{x}\partial x}\dot{\eta}\eta + \frac{\partial^2 L}{\partial x^2}\eta^2\right)dt\right] D\{\eta(t)\}. \tag{4.15}$$

The path integral method presented here can be regarded as a functional version of the Laplace approximation used in the field of statistics and machine learning (Kass, Tierney, & Kadane, 1991; Rasmussen & Williams, 2006).

4.3 Evaluation of the Marginal Likelihood. We formulate a method for computing a marginal likelihood function averaged over the ensemble of event sequences derived from a given underlying rate,

$$\lambda(t) = \mu + \sigma g(t), \tag{4.16}$$

where μ is the mean rate and $\sigma g(t)$ represents a rate fluctuation characterized by the amplitude σ, $\langle g(t)\rangle = 0$ and $\langle g(t)g(t')\rangle = \phi(t - t')$. We suppose that event sequences are generated from a time-rescaled gamma interval process with the shape parameter κ^*.

The fluctuation in the apparent spike count is given by the variance to mean ratio as represented by the Fano factor (Fano, 1947). For the renewal process in which ISIs are drawn from a given distribution function, it is proven that the Fano factor is related to the ISI variability with $F \approx C_V^2$, where C_V is the coefficient of variation defined as the standard deviation of the ISIs divided by the mean (Cox, 1962). Though this approximation is proven to hold in the long interval limit, it was found that the relation practically holds in a short interval that contains only a few spikes (Figure 4 in Omi & Shinomoto, 2011). The ISI variability of the gamma distribution is given by $C_V = 1/\sqrt{\kappa^*}$. Thus, in each realization of a spike generation, the occurrence of events fluctuates around the underlying rate, and $\sum_i \delta(t - t_i)$ in equation 4.11 can be represented as a stochastic process,

$$\sum_i \delta(t - t_i) \approx \lambda(t) + \sqrt{\frac{\lambda(t)}{\kappa^*}}\xi(t), \tag{4.17}$$

where $\xi(t)$ is a white noise characterized by the ensemble averages $\langle \xi(t)\rangle = 0$ and $\langle \xi(t)\xi(t')\rangle = \delta(t - t')$. Note that this approximation, equation 4.17, holds only for a long timescale fluctuation in the rate in which a serial correlation of spikes is negligible. Because the rate is nonnegative, the path integral defined by equation 4.10 should be carried out in the range of $\lambda \geq 0$. Under the condition that the rate fluctuation is small, $\sigma/\mu \ll 1$; however, the orbits passing through $\lambda < 0$ practically do not contribute to the integral, and we can ignore the case of $\lambda < 0$.

Under the same condition $\sigma/\mu \ll 1$ and a large time interval $T \gg 1$, the "free energy" or the negative log marginal likelihood function is derived by evaluating each factor in equation 4.9 as

$$\begin{aligned} F(\gamma, \kappa) &\equiv -\frac{1}{T}\log p_{\gamma,\kappa}(\{t_i\}) \\ &= -\frac{1}{T}\left(\log R - \log Z(\gamma) - \int_0^T L(\dot{\hat{x}}, \hat{x})\, dt + \mathcal{L}\right) \\ &= \frac{\gamma}{4}\sqrt{\frac{\kappa}{\mu}}\left\{2 - \frac{\kappa}{\kappa^*}\left(1 + 2\beta\int_0^\infty \phi(u)e^{-\gamma\sqrt{\frac{\kappa}{\mu}}u}du\right)\right\} \end{aligned}$$

$$- \mu \Bigg\{ \log \mu - \kappa + \kappa \log \kappa - \log \Gamma(\kappa) + (\kappa - 1) \left[\psi(\kappa^*) - \log \kappa^* - \frac{\sigma^2 \phi(0)}{2\mu^2} \right] \Bigg\}, \tag{4.18}$$

where $\beta = \kappa^* \sigma^2 / \mu$ represents the effective degree of fluctuation and $\psi(\kappa) \equiv d \log \Gamma(\kappa)/d\kappa$ is the digamma function. The derivation is given in the appendix. The hyperparameters (γ, κ) are selected so that the marginal likelihood is maximized (see equation 3.6), or the free energy is minimized.

5 Applying the Decoding Methods to Non-Poisson Data

We apply this method of analysis to the time-rescaled gamma interval processes, in which the rate fluctuation is given by specific processes. Note that the free energy, equation 4.18, depends on the rate process via its correlation function $\phi(u)$. We use two rate processes fluctuating according to the sinusoidal and Ornstein-Uhlenbeck processes; these are the simplest processes whose correlation function is periodic or decaying with a given correlation timescale, respectively.

5.1 Sinusoidally Modulated Gamma Interval Process. First we consider detecting rate fluctuation from the gamma interval process in which the underlying rate is modulated sinusoidally in time:

$$\lambda(t) = \mu + \sigma \sin t/\tau. \tag{5.1}$$

Inserting the correlation function of this process, $\phi(u) = \frac{1}{2} \cos \frac{u}{\tau}$, into equation 4.18, the free energy for this case is obtained as

$$F(\gamma, \kappa) = \frac{\gamma}{4} \sqrt{\frac{\kappa}{\mu}} \left(2 - \frac{\kappa}{\kappa^*} \right) - \frac{\beta \kappa}{4\kappa^*} \frac{\tau^2 \gamma^2 \kappa}{\mu + \tau^2 \gamma^2 \kappa} - \mu \Bigg\{ \log \mu - \kappa + \kappa \log \kappa - \log \Gamma(\kappa) + (\kappa - 1) \left[\psi(\kappa^*) - \log \kappa^* - \frac{\sigma^2}{4\mu^2} \right] \Bigg\}. \tag{5.2}$$

A vanishing hyperparameter $\gamma = 0$ indicates a constant rate. This free energy defined for $\gamma \geq 0$ always has a minimum at $(\gamma, \kappa) = (0, \hat{\kappa}_c)$, where $\hat{\kappa}_c = \kappa^* - \frac{\sigma^2 \phi(0)}{2\mu^2 I(\kappa^*)}$, $I(\kappa^*) = \dot{\psi}(\kappa^*) - 1/\kappa^*$ being the Fisher information, as $\partial F/\partial \gamma|_{(\gamma,\kappa)=(0,\hat{\kappa}_c)} > 0$ and $\partial F/\partial \kappa|_{(\gamma,\kappa)=(0,\hat{\kappa}_c)} = 0$. As the rate fluctuation is increased from zero to $\beta\tau \equiv \kappa^* \sigma^2 \tau/\mu > 2$, the minimum at a finite γ becomes lower than the minimum at $\gamma = 0$, implying that the event sequence should

be interpreted as being derived from a fluctuating rate. It should be noted that the critical amplitude of the rate fluctuation σ_c is scaled with $1/\sqrt{\kappa^*}$ (see Figure 1A, top panel), demonstrating that the smallest amplitude of detectable fluctuation reduces in proportion with the coefficient of variation of ISIs, C_V.

A Poissonian decoder is obtained by simply setting $\kappa = 1$ instead of optimizing κ. By putting $\kappa = 1$ in equation 4.18 and deriving the condition under which $F(\gamma, 1)$ takes its minimum at $\gamma = 0$, the detection limit for the Poisson decoder is given by $\beta\tau \equiv \kappa^*\sigma^2\tau/\mu = 2(2\kappa^* - 1)$. In the regular domain, $\kappa^* > 1$ or $C_V < 1$, the detection limit amplitude for the Poisson decoder is larger than that of the non-Poisson decoder, implying the information transmission efficiency is dependent not only on the regularity of the spike train but also on whether the decoder takes account of the non-Poisson firing.

The analytical results are compared with numerical simulations in Figure 1. Here we generated a number of spike trains from inhomogeneous non-Poisson processes, equation 2.6, and attempted to estimate the rate with the Bayes decoder, equation 3.1. The maximization of the marginal likelihood, equation 3.6, was carried out by the expectation-maximization (EM) method prescribed in Koyama and Shinomoto (2005). A sample spike train (b) in Figure 1 demonstrates the situation in which a rate fluctuation was detected using a non-Poisson decoder, while the rate estimator that does not take account of non-Poisson firing cannot capture rate fluctuation.

To compare the non-Poisson and Poisson decoders in their performances in estimating the underlying rate, we have plotted ISEs computed for data derived from detectable rate fluctuations (see Figure 2A). The result indicates that the non-Poisson decoder has improved the fitting performance from the Poisson decoder.

Note that the relationship is reversed in the bursty domain, $\kappa^* < 1$ or $C_V > 1$; the Poisson decoder yields the detection of the rate fluctuation even for an amplitude smaller than the lower-bound amplitude for the non-Poisson decoder (see Figure 1A). If we assess the integrated squared error (ISE) of the estimated rate from the true rate, the Poisson decoder gives a larger error than the non-Poisson decoder. This implies that the Poisson decoder (mis)interpreted the bursty firing as signal rather than noise. A sample spike train (a) in Figure 1 demonstrates the situation in which a rate fluctuation was not detected using a non-Poisson decoder, while the Poisson decoder detected spurious fluctuation.

5.2 OUP Modulated Gamma Interval Process. In the second example, we consider the case in which the rate fluctuation is fluctuating with the Ornstein-Uhlenbeck (OUP) process,

$$\frac{d\lambda}{dt} = -\frac{\lambda - \mu}{\tau} + \sigma\sqrt{\frac{2}{\tau}}\xi(t), \tag{5.3}$$

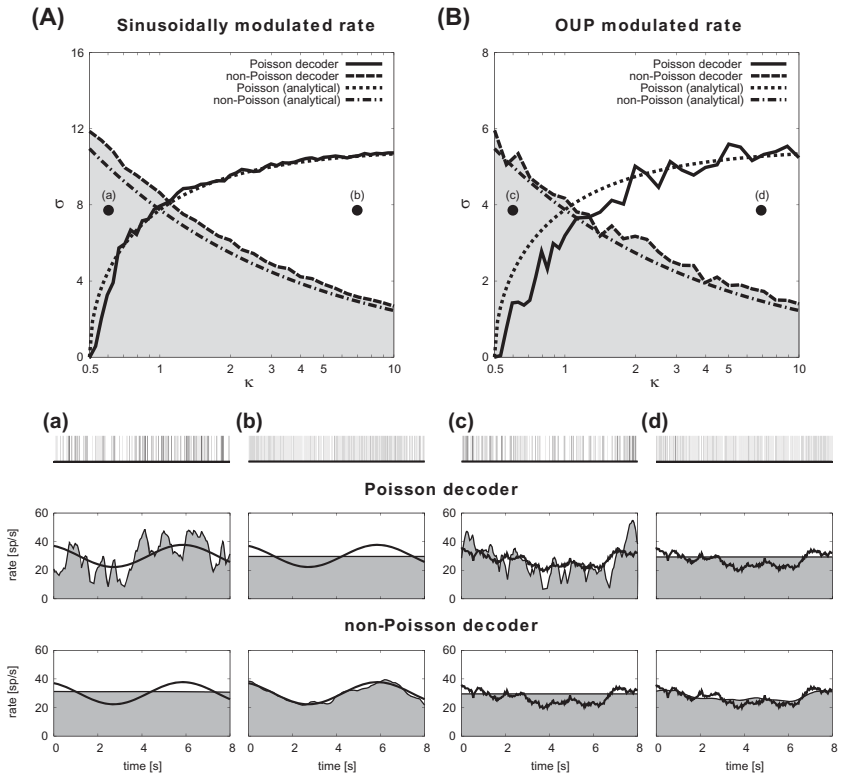

Figure 1: The smallest amplitudes of detectable rate fluctuations and estimation samples. (A) Sinusoidally modulated processes. (B) OUP modulated processes. Top: The ordinate represents the amplitude of the rate fluctuation σ, below which Bayesian decoders cannot detect fluctuations. The abscissa represents the degree of non-Poisson irregularity as represented by the shape factor of the gamma ISI distribution; $\kappa > 1$, $= 1$, and <1 represent regular, Poisson random, and bursty firing, respectively. Dashed and solid lines depict the numerical detection limit of the non-Poisson and Poisson empirical Bayes estimators, respectively. Spike sequences were derived from the inhomogeneous gamma interval process either sinusoidally or OUP modulated with the mean rate $\mu = 30$ Hz and the timescale of $\tau = 1$ s. Dot-dashed and dotted lines indicate theoretical bounds obtained from the path integral marginalization integral based on non-Poisson and Poisson assumptions, respectively. The analytical result for a non-Poisson decoder is given by $\sigma_c/\sqrt{\kappa}$. Bottom: Raster plots of sample spike trains and the estimated rates. The solid lines and the shaded areas represent the underlying rates and the rates estimated by the Poisson and non-Poisson Bayesian decoders. The parameters (κ, σ) of a, b, c, and d are $(0.6, \sqrt{60})$, $(7, \sqrt{60})$, $(0.6, \sqrt{15})$, and $(7, \sqrt{15})$, respectively.

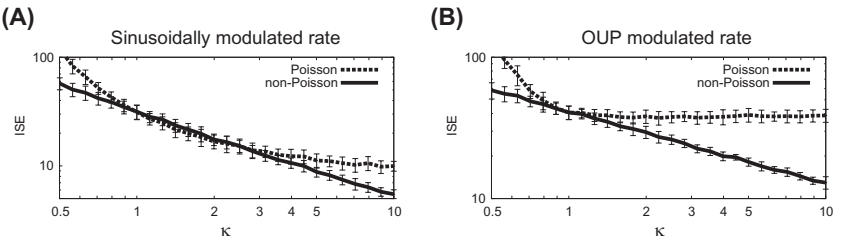

Figure 2: Estimation error of the non-Poisson and Poisson Bayesian decoders. (A) Sinusoidally modulated processes (parameters: $\mu = 30$ Hz, $\tau = 1$ s, $\sigma = 20$ Hz). (B) OUP modulated processes (parameters: $\mu = 30$ Hz, $\tau = 1$ s, $\sigma = 10$ Hz). The ordinate represents integrated squared error (ISE) between the underlying rate and the estimated rate, and the abscissa represents the shape factor of the gamma ISI distribution κ. The mean and standard deviation of ISEs are obtained from 50 samples computed for an interval of $T = 100$ s.

where $\xi(t)$ is a gaussian white noise. Substituting the correlation function of this process, $\phi(u) = e^{-|u|/\tau}$, into equation 4.18, the free energy for this case is derived as

$$\begin{aligned} F(\gamma, \kappa) = {} & \frac{\gamma}{4}\sqrt{\frac{\kappa}{\mu}}\left(2 - \frac{\kappa}{\kappa^*}\right) - \frac{\beta\kappa}{2\kappa^*}\frac{\tau\gamma\sqrt{\kappa}}{\sqrt{\mu} + \tau\gamma\sqrt{\kappa}} \\ & - \mu\left\{\log\mu - \kappa + \kappa\log\kappa - \log\Gamma(\kappa)\right. \\ & \left. + (\kappa - 1)\left[\psi(\kappa^*) - \log\kappa^* - \frac{\sigma^2}{2\mu^2}\right]\right\}. \end{aligned} \tag{5.4}$$

In contrast to the case of sinusoidal rate fluctuation, this free energy never has multiple minima. The minimum of this function stays at $(\gamma, \kappa) = (0, \hat{\kappa}_c)$ until $\beta\tau \equiv \kappa^*\sigma^2\tau/\mu$ exceeds 1/2. By setting $\kappa = 1$ instead of estimating κ by the full optimization, the detection limit for the Poisson decoder can also be obtained as $\beta\tau \equiv \kappa^*\sigma^2\tau/\mu = (2\kappa^* - 1)/2$. The critical amplitude of the rate fluctuation σ_c for the OUP modulated process is scaled with $1/\sqrt{\kappa^*}$, which is the same as the case of sinusoidally modulated process (see Figure 1A, top panel). The analytical results are also compared with numerical simulations in Figure 1B. The numerical results are subject to a larger fluctuation than the sinusoidally modulated cases because the rate itself fluctuates greatly according to the stochastic process. The improvement in the estimation by replacing the Poissonian decoder by the non-Poissonian decoder can be seen from their ISEs (see Figure 2B). The conditions are similar to those for the sinusoidally modulated cases, including the higher sensitivity of the

non-Poisson decoder for regular spike trains (a sample spike train, d, in Figure 1) and the misinterpretation of the Poisson decoder for bursty spike trains (a sample spike train, c, in Figure 1).

6 Discussion

We have shown that the sensitivity for rate fluctuations is enhanced by a encoding rate with non-Poisson regular firing and decoding a spike train by taking account of non-Poisson firing characteristics. The smallest amplitude of detectable fluctuation decreases with $1/\sqrt{\kappa^*} = C_V$ of the ISI distribution. Because our derivation of the detectable fluctuation is based on the estimation of the fluctuation of the spike count in terms of the Fano factor, which is related to the coefficient of variation, or equation 4.17, this scaling of the threshold fluctuation with $1/\sqrt{\kappa^*}$ applies to any inhomogeneous non-Poisson, process including sinusoidally modulated and OUP modulated gamma interval processes (see Figure 1). Note that considerable care should be given to the use of the coefficient of variation, because C_V itself is fragile and tends to be biased upward in the presence of rate fluctuation (Shinomoto et al., 2005). The intrinsic firing irregularity κ could be estimated after eliminating rate fluctuation by rescaling the time axis (Reich et al., 1998; Oram et al., 1999; Koyama & Shinomoto, 2005; Nawrot et al., 2008; Shimokawa & Shinomoto, 2009) or through the local variation L_V introduced in Shinomoto, Shima, and Tanji (2003), which is robust against the rate fluctuation, as $1/\sqrt{\kappa^*} = \sqrt{2L_V/(3 - L_V)}$.

It should be emphasized that the information transmission efficiency is also dependent on the capacity of the decoder for detecting the firing irregularity and utilizing the knowledge in analyzing the spike train (Pillow et al., 2005). As has been demonstrated in Figure 2, the Poissonian rate decoder is unable to lower the detection threshold even if the sender is using non-Poisson regular spike trains. Our simulations show, for example, that for a value of $\kappa^* = 2.5$ consistent with the irregularity observed in the motor cortex (Shinomoto et al., 2009), the non-Poisson decoder can detect rate fluctuations of approximately 5 Hz, while the Poisson decoder requires fluctuations as large as 10 Hz. Furthermore, the Poisson decoder may detect spurious rate fluctuations when a bursty spike train is used to encode the rate. Interpretation of this phenomenon in the bursty case is not simple because the encoder and decoder should have some protocol in transmitting information when using bursty spikes, in which ISIs fluctuate greater than the case of the Poisson process.

Although we treated single neurons in this letter, our model can also account for interneuronal correlations by incorporating a covariate of other neurons' activity into the intensity function (Truccolo, Eden, Fellows, Donoghue, & Brown, 2005; Pillow et al., 2008).

To derive the formula for the free energy (negative log marginal likelihood), equation 4.18, we assumed that the timescale of the rate fluctuation

is longer than the mean ISI so that the firing rate in each ISI does not change drastically. The analytical results obtained under this assumption were in good agreement with our numerical simulations (see Figures 1A and 1B).

Cortical neurons in vivo have been approximated as Poisson spike generators, but a recent nonstationary analysis has revealed that individual neurons are signaling with non-Poisson firing, whose characteristics are specific to individual neurons and appear to depend on the function of the cortical area; the data examined indicated that neuronal firing is regular in the primary and higher-order motor areas, random in the visual areas, and bursty in the prefrontal area (Shinomoto et al., 2009). This implies that firing patterns may play an important role in function-specific computation. In accordance with our finding, the regular firing in the motor-related areas can be understood as efficient in transmitting transient motor commands. The firing rate of neuronal populations in the motor area encodes motor commands, and information can be successfully decoded by the population vector algorithm (Georgopoulos, Kettner, & Schwartz, 1986, 1989). The decoding efficiency can be improved by incorporating the variability of neuronal firing into decoders using a probabilistic framework (Brockwell, Rojas, & Kass, 2004; Brockwell, Schwartz, & Kass, 2007; Koyama, Chase et al., 2010; Wu, Gao, Bienenstock, Donoghue, & Black, 2006). Most previous work, however, has assumed Poisson statistics for spiking variability. Using non-Poisson firing models could improve decoding efficiency.

Ma, Beck, Latham, and Pouget (2006) suggested a hypothesis that the Poisson-like statistics in the responses of populations of cortical neurons may represent probability distributions over the stimulus and a broad class of Bayesian inference can be implemented by simple linear combinations of populations of neural activity. A crucial assumption made in their hypothesis is that the covariance matrix of spike count is proportional to the mean spike count, based on the mathematical fact that the variance is equal to the mean for Poisson firings. It should be noted that their argument does not directly apply to real neuronal spike trains, which are of a non-Poisson nature. However, if the firing irregularity is maintained in individual neurons, as has been exemplified by time-rescaling to a non-Poisson renewal process, their argument still applies to such neuronal firing, in which the proportionality between the variance and mean holds, with a coefficient different from unity.

There are many alternative methods for estimating fluctuating firing rates (Kass, Ventura, & Brown, 2005), including an optimized histogram and kernel density estimator chosen to minimize mean squared error (Shimazaki & Shinomoto, 2007, 2010; Omi & Shinomoto, 2011). Shintani and Shinomoto (2012) showed that the detectable limit of rate fluctuations is given by an identical formula for the optimized histogram and the Bayesian decoder if a spike train is derived from an inhomogeneous Poisson process. It would be interesting to know whether this result can be generalized to an inhomogeneous non-Poisson spike trains.

We still do not know why random firings are used in the higher-order cortical areas. Though several hypotheses have been proposed for explaining the function of irregular firing (van Vreeswijk & Sompolinsky, 1996; Roudi & Latham, 2007), it is open to future research.

Appendix: Derivation of the Free Energy

In this appendix, we present the detail of derivation of the free energy (see equation 4.18). Under the condition $\sigma/\mu \ll 1$, the Lagrangian can be approximated to the range quadratic with respect to x and $\dot{x}$ as

$$\begin{aligned} L(\dot{x}, x) &= \frac{1}{2\gamma^2}\dot{x}^2 + \kappa x - \kappa\{\mu + \sigma g(t) + \sqrt{[\mu + \sigma g(t)]/\kappa^*}\xi(t)\}\log\left(1 + \frac{x}{\mu}\right) \\ &\approx \frac{1}{2\gamma^2}\dot{x}^2 - \kappa\frac{\sigma g(t) + \sqrt{\mu/\kappa^*}\xi(t)}{\mu}x + \frac{\kappa}{2\mu}x^2, \end{aligned} \tag{A.1}$$

where we have ignored the term $o((\sigma/\mu)^{3/2})$. The solution of the Euler-Lagrange equation, 4.12, representing the MAP estimate of the rate fluctuation, is given by

$$\hat{x}(t) = \frac{\gamma}{2}\sqrt{\frac{\kappa}{\mu}}\int_0^T e^{-\gamma\sqrt{\frac{\kappa}{\mu}}|t-s|}\left\{\sigma g(s) + \sqrt{\mu/\kappa^*}\xi(s)\right\}ds. \tag{A.2}$$

The action integral along the MAP path $S[\hat{x}(t)]$ and the contribution of the quadratic derivation R in equation 4.14 are obtained analytically. Utilizing the Euler-Lagrange equation, we can rewrite $S[\hat{x}(t)]$ as

$$\begin{aligned} \int_0^T L(\dot{\hat{x}}, \hat{x})\,dt &= \int_0^T \left\{\frac{1}{2\gamma^2}\dot{\hat{x}}^2 - \kappa\frac{\sigma g(t) + \sqrt{\mu/\kappa^*}\xi(t)}{\mu}\hat{x} + \frac{\kappa}{2\mu}\hat{x}^2\right\}dt \\ &= \int_0^T \left\{\frac{1}{2\gamma^2}\frac{d}{dt}(\dot{\hat{x}}\hat{x}) - \kappa\frac{\sigma g(t) + \sqrt{\mu/\kappa^*}\xi(t)}{2\mu}\hat{x}\right\}dt. \end{aligned} \tag{A.3}$$

The first term on the right-hand side representing the end point effect is negligible compared to the second term, whose contribution is of the order of $T \gg 1$. $S[\hat{x}(t)]$ can be obtained explicitly by inserting the MAP solution, equation A.2, into the second term of equation A.3 as

$$\int_0^T L(\dot{\hat{x}}, \hat{x})\,dt = -\frac{\gamma}{4}\sqrt{\frac{\kappa}{\mu}}T\left\{\frac{\kappa}{\kappa^*} + 2\frac{\kappa\sigma^2}{\mu}\int_0^\infty \phi(u)e^{-\gamma\sqrt{\frac{\kappa}{\mu}}u}du\right\}. \tag{A.4}$$

Substituting the Lagrangian, equation A.1, into equation 4.15, R is obtained as

$$R = \int \exp\left[-\frac{1}{2}\int_0^T \left(\frac{1}{\gamma^2}\dot{\eta}^2 + \frac{\kappa}{\mu}\eta^2\right)dt\right] D\{\eta(t)\}. \tag{A.5}$$

An integration by part leads to

$$\int_0^T \left(\frac{1}{\gamma^2}\dot{\eta}^2 + \frac{\kappa}{\mu}\eta^2\right)dt = \int_0^T \eta\left(-\frac{1}{\gamma^2}\partial_t^2 + \frac{\kappa}{\mu}\right)\eta\, dt, \tag{A.6}$$

where we have used the boundary condition $\eta(0) = \eta(T) = 0$. Let $\{\vartheta_i(t)\}$ be a complete set of orthogonal eigenfunctions of $(-\frac{1}{\gamma^2}\partial_t^2 + \frac{\kappa}{\mu})$ vanishing at the boundaries, and $\{\theta_i\}$ be its eigenvalues. Then $\eta(t)$ can be expressed as $\eta(t) = \sum_i a_i\vartheta_i(t)$. Accordingly, the measure $D\{\eta(t)\}$ is transformed to $D\{\eta(t)\} = C\prod_i (2\pi)^{-\frac{1}{2}} da_i$, where C is a constant chosen so that the integral over this measure corresponds to the Wiener integral, equation 3.3, and we find

$$\begin{aligned} R &= C\prod_i \int_{-\infty}^{\infty} \frac{da_i}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\theta_i a_i^2\right) \\ &= C\prod_i \theta_i^{-\frac{1}{2}} \equiv C\det\left(-\frac{1}{\gamma^2}\partial_t^2 + \frac{\kappa}{\mu}\right)^{-\frac{1}{2}}. \end{aligned} \tag{A.7}$$

From equations 3.3 and A.7, we also obtain $C\det(-\frac{1}{\gamma^2}\partial_t^2)^{-\frac{1}{2}} = 1/\sqrt{2\pi\gamma^2 T}$. Thus, R is obtained as

$$R = \frac{1}{\sqrt{2\pi\gamma^2 T}}\left[\frac{\det\left(-\frac{1}{\gamma^2}\partial_t^2 + \frac{\kappa}{\mu}\right)}{\det\left(-\frac{1}{\gamma^2}\partial_t^2\right)}\right]^{-\frac{1}{2}} = \frac{1}{\sqrt{2\pi\gamma^2 T}}\left[\frac{\varphi_1(T)}{\varphi_2(T)}\right]^{-\frac{1}{2}}. \tag{A.8}$$

It has been proved that the determinants can be computed by solving the associated differential equations (Coleman, 1988; Kleinert, 2009; Gelfand & Yaglom, 1960):

$$\begin{aligned} \left(-\frac{1}{\gamma^2}\partial_t^2 + \frac{\kappa}{\mu}\right)\varphi_1(t) = 0, \quad \varphi_1(0) = 0, \quad \frac{d\varphi_1(0)}{dt} = 1, \\ -\frac{1}{\gamma^2}\partial_t^2\varphi_2(t) = 0, \quad \varphi_2(0) = 0, \quad \frac{d\varphi_2(0)}{dt} = 1. \end{aligned} \tag{A.9}$$

These differential equations are solved as $\varphi_1(t) = \frac{1}{\gamma}\sqrt{\frac{\mu}{\kappa}}\sinh\gamma\sqrt{\frac{\kappa}{\mu}}t$ and $\varphi_2(t) = t$, from which R is obtained as

$$R = \left(\pi\gamma\sqrt{\frac{\mu}{\kappa}}\right)^{-\frac{1}{2}} \exp\left(-\frac{\gamma}{2}\sqrt{\frac{\kappa}{\mu}}T\right), \tag{A.10}$$

for $T \gg 1$.

In order to evaluate the likelihood function of the gamma distribution, equation 4.7, we need to evaluate $\frac{1}{n}\sum_i \log(t_i - t_{i-1})$. Let $\{t_i^{(\lambda)} - t_{i-1}^{(\lambda)}\}$ be a set of ISIs derived from the gamma distribution with the rate λ and n_λ be the number of the ISIs in this set. Then we obtain

$$\frac{1}{n_\lambda}\sum_{i=1}^{n_\lambda}\log\left(t_i^{(\lambda)} - t_{i-1}^{(\lambda)}\right) \to \psi(\kappa^*) - \log\kappa^* - \log\lambda, \quad \text{as } n_\lambda \to \infty,$$

where $\psi(\kappa)$ is the digamma function. On the other hand, $n_\lambda/n \to \lambda p(\lambda)d\lambda/\mu$ as $n \to \infty$ from the law of large number, where $p(\lambda)$ is a stationary distribution of $\lambda(t)$. Using these, we obtain

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log\left(t_i - t_{i-1}\right) = \int_0^\infty \left[\psi(\kappa^*) - \log\kappa^* - \log\lambda\right]\frac{\lambda p(\lambda)}{\mu}d\lambda.$$

By expanding up to the second order with respect to σ/μ, the above equation can be evaluated as

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log(t_i - t_{i-1}) = \psi(\kappa^*) - \log\kappa^* - \log\mu - \frac{\sigma^2\phi(0)}{2\mu^2}. \tag{A.11}$$

When equation A.11 is substituted into equation 4.7, the log likelihood of the gamma distribution is obtained as

$$\frac{1}{T}\mathcal{L} = \mu\left\{\log\mu - \kappa + \kappa\log\kappa - \log\Gamma(\kappa)\right.$$
$$\left. + (\kappa - 1)\left[\psi(\kappa^*) - \log\kappa^* - \frac{\sigma^2\phi(0)}{2\mu^2}\right]\right\}. \tag{A.12}$$

Summing the action integral, equation A.4, the contribution of the quadratic derivation, equation A.10, and the likelihood function, equation A.12, we obtain the free energy, equation 4.18.

Acknowledgments

S.K. was supported by JSPS KAKENHI grant number 24700287. T.O. was supported by JSPS Research Fellowships for Young Scientists. S.S. was supported by Grants-in-Aid for Scientific Research from MEXT Japan (20300083, 23115510) and the Global COE Program "The Next Generation of Physics, Spun from Universality and Emergence."

References

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

Brockwell, A. E., Rojas, A. L., & Kass, R. E. (2004). Recursive Bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology*, *91*, 1899–1907.

Brockwell, A. E., Schwartz, A. B., & Kass, R. E. (2007). Statistical signal processing and the motor cortex. *Proceedings of the IEEE*, *95*, 882–898.

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2001). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, *14*, 325–346.

Bruce, A. B., & Saad, D. (1994). Statistical mechanics of hypothesis evaluation. *Journal of Physics A: Mathematical and General*, *27*, 3355–3364.

Burns, B. D., & Pritchard, R. (1964). Contrast discrimination by neurons in the cat's visual cerebral cortex. *Journal of Physiology*, *175*, 445–463.

Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. (2nd ed.). London: Chapman and Hall.

Coleman, S. (1988). *Aspects of symmetry*. Cambridge: Cambridge University Press.

Cox, D. R. (1962). *Renewal theory*. London: Chapman and Hall.

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

Cunningham, J. P., Yu, B. M., Shenoy, K. V., & Sahani, M. (2008). Inferring neural firing rates from spike trains using gaussian processes. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Neural information processing systems*, *20* (pp. 329–336). Cambridge, MA: MIT Press.

Daley, D., & Vere-Jones, D. (1988). *An introduction to the theory of point processes*. New York: Springer-Verlag.

Davies, R. M., Gerstein, G. L., & Baker, S. N. (2006). Measurement of time-dependent changes in the irregularity of neural spiking. *Journal of Neurophysiology*, *96*, 906–918.

Fano, U. (1947). Ionization yield of radiations. II. The fluctuations of the number of ions. *Physical Review*, *72*, 26–29.

Feynman, R. P., & Hibbs, A. R. (1965). *Quantum mechanics and path integrals*. New York: McGraw-Hill.

Gelfand, I. M., & Yaglom, A. M. (1960). Integration in functional spaces and applications in quantum physics. *Journal of Mathematical Physics*, *1*, 48–69.

Georgopoulos, A., Kettner, R., & Schwartz, A. (1986). Neuronal population coding of movement direction. *Science*, *233*, 1416–1419.

Georgopoulos, A., Kettner, R., & Schwartz, A. (1989). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neural population. *Journal of Neuroscience*, *8*, 2928–2937.

Grun, S. (2009). Data-driven significance estimation for precise spike correlation. *Journal of Neurophysiology*, *101*, 1126–1140.

Jacobs, A. L., Fridman, G., Douglas, R. M., Alam, N. M., Latham, P. E., Prusky, G. T., et al. (2009). Ruling out and ruling in neural codes. *Proceedings of the National Academy of Sciences*, *106*, 5936–5941.

Karatzas, I., & Shreve, S. E. (1997). *Brownian motion and stochastic calculus* (2nd ed.). New York: Springer.

Kass, R. E., & Raftery, A. (1995). Bayes factors. *Journal of American Statistical Association*, *90*, 773–795.

Kass, R. E., Tierney, L., & Kadane, J. B. (1991). Laplace's method in Bayesian analysis. In N. Flourny & R. K. Tsutakawa (Eds.), *Statistical multiple integration* (pp. 89–99). Providence, RI: American Mathematical Society.

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

Kass, R. E., Ventura, V., & Brown, E. N. (2005). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*, *94*, 8–25.

Kleinert, H. (2009). *Path integrals in quantum mechanics, statistics, polymer physics, and financial markets* (5th ed.). Singapore: World Scientific.

Koyama, S., Chase, S. M., Whitford, A. S., Velliste, M., Schwartz, A. B., & Kass, R. E. (2010). Comparison of decoding algorithms in open-loop and closed-loop performance. *Journal of Computational Neuroscience*, *29*, 73–87.

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

Koyama, S., Perez-Bolde, L. C., Shalizi, C. R., & Kass, R. E. (2010). Approximate methods for state-space models. *Journal of American Statistical Association*, *105*, 170–180.

Koyama, S., Shimokawa, T., & Shinomoto, S. (2007). Phase transitions in the estimation of event rate: A path integral analysis. *Journal of Physics A: Mathematical and General*, *40*, F383–F390.

Koyama, S., & Shinomoto, S. (2004). Histogram bin width selection for time-dependent Poisson processes. *Journal of Physics A: Mathematical and General*, *37*, 7255–7265.

Koyama, S., & Shinomoto, S. (2005). Empirical Bayes interpretations of random point events. *Journal of Physics A: Mathematical and General*, *38*, L531–L537.

Lundstrom, B. N., & Fairhall, A. L. (2006). Decoding stimulus variance from a distributional neural code of interspike intervals. *Journal of Neuroscience*, *26*, 9030–9037.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438.

MacKay, D.J.C. (1992). Bayesian interpolation. *Neural Computation*, *4*, 415–447.

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

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J American Statistical Association*, *83*, 9–27.

Olypher, A. V., Lansky, P., & Fenton, A. A. (2002). Properties of the extra-positional signal in hippocampal place cell discharge derived from the overdispersion in location-specific firing. *Neuroscience*, *111*, 553–566.

Omi, T., & Shinomoto, S. (2011). Optimizing time histograms for non-Poisson spike trains. *Neural Computation*, *23*, 3125–3144.

Oram, M. W., Wiener, M. C., Lestienne, R., & Richmond, B. J. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *Journal of Neurophysiology*, *81*, 3021–3033.

Perkel, D. H., & Bullock, T. H. (1968). Neural coding. *Neurosciences Research Program Bulletin*, *6*, 221–348.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, *23*, 11003–11013.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., et al. (2008). Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature*, *454*, 995–999.

Ramsay, J., & Silverman, B. W. (2010). *Functional data analysis* (2nd ed.). New York: Springer.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

Ratliff, F., Hartline, H. K., & Lange, D. (1968). Variability of interspike intervals in optic nerve fibers of limulus. *Proceedings of the National Academy of Science of the United States of America*, *60*, 464–469.

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

Roudi, Y., & Latham, P. E. (2007). A balanced memory network. *PLoS Computational Biology*, *3*, 1679–1700.

Schulman, L. S. (2005). *Techniques and applications of path integration*. New York: Dover.

Shimazaki, H., & Shinomoto, S. (2007). A method for selecting the bin size of a time histogram. *Neural Computation*, *19*, 1503–1527.

Shimazaki, H., & Shinomoto, S. (2010). Kernel bandwidth optimization in spike rate estimation. *Journal of Computational Neuroscience*, *29*, 171–182.

Shimokawa, T., & Shinomoto, S. (2009). Estimating instantaneous irregularity of neuronal firing. *Neural Computation*, *21*, 1931–1951.

Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., et al. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Computational Biology*, *5*, e1000433.

Shinomoto, S., Miura, K., & Koyama, S. (2005). A measure of local variation of inter-spike intervals. *BioSystems*, *79*, 67–72.

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

Shintani, T., & Shinomoto, S. (2012). Detection limit of rate fluctuations in inhomogeneous Poisson processes. *Physical Review E*, *85*, 041139.

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

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & 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*, 1074–1089.

van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, *274*, 1724–1726.

Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P., & Black, M. J. (2006). Bayesian population decoding of motor cortical activity using a Kalman filter. *Neural Computation*, *18*, 80–118.

Received April 5, 2012; accepted October 18, 2012.