LETTER Communicated by Jonathan Victor

Estimating Spiking Irregularities Under Changing Environments

Keiji Miura
miura@ton.scphys.kyoto-u.ac.jp
Department of Physics, Kyoto University, Kyoto 606-8502, and Intelligent Cooperation and Control, PRESTO, JST, Chiba 277-8561, Japan

Masato Okada
okada@k.u-tokyo.ac.jp
Department of Complexity Science and Engineering, University of Tokyo, Chiba 277-8561; Intelligent Cooperation and Control, PRESTO, JST, Chiba 277-8561; and Brain Science Institute, RIKEN, Saitama 351-0198, Japan

Shun-ichi Amari
amari@brain.riken.jp
Brain Science Institute, RIKEN, Saitama 351-0198, Japan

We considered a gamma distribution of interspike intervals as a statistical model for neuronal spike generation. A gamma distribution is a natural extension of the Poisson process taking the effect of a refractory period into account. The model is specified by two parameters: a time-dependent firing rate and a shape parameter that characterizes spiking irregularities of individual neurons. Because the environment changes over time, observed data are generated from a model with a time-dependent firing rate, which is an unknown function. A statistical model with an unknown function is called a semiparametric model and is generally very difficult to solve. We used a novel method of estimating functions in information geometry to estimate the shape parameter without estimating the unknown function. We obtained an optimal estimating function analytically for the shape parameter independent of the functional form of the firing rate. This estimation is efficient without Fisher information loss and better than maximum likelihood estimation. We suggest a measure of spiking irregularity based on the estimating function, which may be useful for characterizing individual neurons in changing environments.

1 Introduction

The firing patterns of cortical neurons look very noisy (Holt, Softky, Koch, & Douglas, 1996), so probabilistic models are necessary to describe them (Cox & Lewis, 1966; Sakai, Funahashi, & Shinomoto, 1999; Tuckwell, 1988). For example, Baker and Lemon (2000) showed that the firing patterns recorded from motor areas can be explained using a continuous-time rate-modulated gamma process. Their model had a rate parameter ξ and a shape parameter κ related to spiking irregularity. ξ was assumed to be a function of time because it depended strongly on the behavior of the monkey. κ was assumed to be unique to individual neurons and constant over time.

The assumption that κ is unique to individual neurons is also supported by other studies (Shinomoto, Miura, & Koyama, 2005; Shinomoto, Miyazaki, Tamura, & Fujita, 2005; Shinomoto, Shima, & Tanji, 2003). However, these indirect supports are not conclusive. Therefore, we need to accurately estimate κ to make the assumption more reliable. If the assumption is correct, neurons may be identified by κ estimated from the spiking patterns, and κ may provide useful information about the function of a neuron. In other words, it may be possible to classify neurons according to functional firing patterns rather than static anatomical properties. Thus, it is very important to accurately estimate κ in the field of neuroscience.

In reality, however, it is very difficult to estimate all the parameters in the model from the observed spike data. The reason is that the unknown function for the time-dependent firing rate $\xi(t)$ has infinite degrees of freedom. This kind of estimation problem is called a *semiparametric model* (Bickel, Klaassen, Ritov, & Wellner, 1993; Groeneboom & Wellner, 1992; Pfanzagl, 1990; van der Vaart, 1998) and is generally very difficult to solve. Are there any ingenious methods of estimating κ accurately to overcome this difficulty?

Ikeda (2005) pointed out that the problem we need to consider is the semiparametric model. However, the problem remains unsolved. There is a method called estimating functions (Godambe, 1960, 1976, 1991; Mukhopadhyay, 2004) for semiparametric problems, and a general theory has been developed (Amari, 1987; Amari & Kawanabe, 1997; Amari & Kumon, 1988) from the viewpoint of information geometry (Amari, 1982, 1985, 1998; Amari, Kurata, & Nagaoka, 1992; Amari & Nagaoka, 2001; Murray & Rice, 1993). However, the method of estimating functions cannot be applied to our problem in its original form.

In this letter, we consider the semiparametric model suggested by Ikeda (2005) instead of the continuous-time rate-modulated gamma process. In this discrete-time rate-modulated model, the firing rate varies in time but assumes a fixed value during each interspike interval. This model is a mixture model and can represent various types of interspike interval distributions by adjusting its weight function. For our model, the difficulty of semiparametric models can be explained as follows. If one parameterizes the rate function to be estimated in a manner that does not make assumptions about its form, one needs one parameter for each spike. But then one has more parameters than data. Then there are more parameters than data points, unless there are repeated measures over the same time period. In spite of this difficulty, κ can be estimated by assuming that we have (at least) two observations at each rate and using the method of estimating functions for semiparametric models, in which we do not need to estimate the firing rate.

Various attempts have been made to solve semiparametric models. Neyman and Scott (1948) pointed out that the maximum likelihood method does not generally provide a consistent estimator when the number of parameters increases in proportion to that of observations. In fact, we show that maximum likelihood estimation for our problem is biased. Ritov and Bickel (1990) and Bickel et al. (1993) considered asymptotic attainability of information bound purely mathematically. However, their results are not practical for application to our problem. Amari and Kawanabe (1997) showed a practical method of estimating a finite number of parameters without estimating an unknown function. This is the method of estimating functions. If this method can be applied, then κ can be estimated consistently independently of the functional form of a firing rate.

In this article, we show that the model we consider here belongs to the class of the exponential form defined by Amari and Kawanabe (1997). However, an estimating function does not exist unless multiple observations are given for each firing rate ξ. We show that if multiple observations are given, the method of estimating functions can be applied. In that case, the estimating function of κ can be analytically obtained, and κ can be estimated consistently independently of the functional form of a firing rate. In general, estimation using estimating functions is not efficient. However, for our problem, this method yielded an optimal estimator in the sense of Fisher information (Amari & Kawanabe, 1997). That is, we obtained an efficient estimator whose mean-square error is asymptotically the smallest. The estimator generalizes well even when the assumptions of the model are violated. We suggest a measure of spiking irregularity based on the estimating function, which may be useful for characterizing individual neurons in the case where only a single observation is given for each firing rate.

2 Maximum Likelihood Estimation

2.1 Simple Case. We consider the following gamma distribution, which is defined as

$$q(T;\xi,\kappa) = \frac{(\xi\kappa)^{\kappa}}{\Gamma(\kappa)} T^{\kappa-1} e^{-\xi\kappa T}, \tag{2.1}$$

where the random variable T denotes an interspike interval. We generate interspike intervals from the distribution and align them to make a spike train. The mean and variance of the interspike intervals are

$$\begin{cases} Ex(T) = \frac{1}{\xi} \\ \mathrm{Var}(T) = \frac{1}{\xi^2\kappa}. \end{cases} \tag{2.2}$$

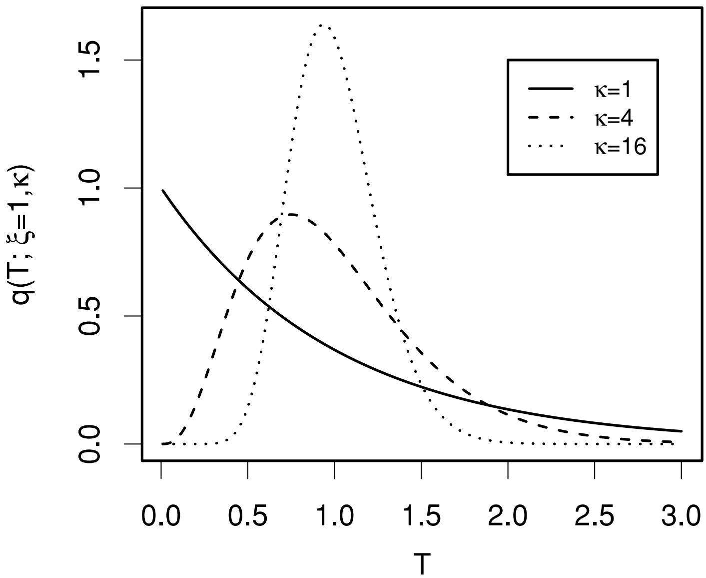

Figure 1: Probability densities of gamma distributions for various κ with $\xi = 1$.

ξ is the mean firing rate, and κ is called a shape parameter. $\kappa = 1$ corresponds to a Poisson process in which the instantaneous firing rate (hazard function) is constant over time independent of the previous firing time. In this case, a spike train looks irregular. When κ is large, a gamma distribution can be approximated by a normal distribution, whose variance decreases with increasing κ. In the limit of large κ, the interspike intervals become completely regular. Thus, κ is related to spiking irregularities. Figure 1 plots the probability densities of gamma distributions for various κ. We can scale T so that $\xi = 1$ because ξ always appears as ξT in $q(T; \xi, \kappa)$.

We assume that ξ changes over time under changing environments. One may assume that ξ is generated for each T from a probability density $k(\xi)$, whose functional form is unknown. Let $\xi^{(l)}$ be the lth firing rate and $T^{(l)}$ be the lth observation of an interspike interval. Thus, we have $N+1$ parameters $\{\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(N)}, \kappa\}$ and N observations $\{T^{(1)}, T^{(2)}, \ldots, T^{(N)}\}$. The purpose is to estimate κ that may be unique to individual neurons by the method of maximum likelihood estimation. Here we estimate all the parameters because we need all $\xi^{(l)}$'s to estimate κ.

The likelihood that $T^{(l)}$'s are generated from the gamma distribution with $\{\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(N)}, \kappa\}$ is given by

$$L = \prod_{l=1}^{N} q(T^{(l)}; \xi^{(l)}, \kappa). \tag{2.3}$$

In maximum likelihood estimation, we choose the parameter values that maximize the likelihood. Without loss of generality, we can consider the maximization of the log likelihood:

$$\log L = \sum_{l=1}^{N} \log q(T^{(l)}; \xi^{(l)}, \kappa). \tag{2.4}$$

The estimated parameters must satisfy

$$\frac{\partial}{\partial \kappa} \log L = \sum_{l=1}^{N} u(T^{(l)} : \xi^{(l)}, \kappa) = 0 \ \text{ and} \tag{2.5}$$

$$\frac{\partial}{\partial \xi^{(l)}} \log L = v(T^{(l)}; \xi^{(l)}, \kappa) = 0 \tag{2.6}$$

for all l, where the score functions for $q(T; \xi, \kappa)$ are defined as

$$u(T; \xi, \kappa) = \frac{\partial}{\partial \kappa} \log q(T; \xi, \kappa)$$
$$= 1 - \xi T + \log T + \log(\kappa \xi) - \phi(\kappa) \ \text{ and} \tag{2.7}$$

$$v(T; \xi, \kappa) = \frac{\partial}{\partial \xi} \log q(T; \xi, \kappa) = \frac{\kappa}{\xi} - \kappa T, \tag{2.8}$$

where the digamma function $\phi(\kappa)$ is defined using the gamma function $\Gamma(\kappa)$ as

$$\phi(\kappa) = \frac{\Gamma'(\kappa)}{\Gamma(\kappa)}. \tag{2.9}$$

The parameters are estimated by solving these equations as

$$\hat{\kappa} = \infty \ \text{ and} \tag{2.10}$$

$$\hat{\xi}^{(l)} = \frac{1}{T^{(l)}} \tag{2.11}$$

for all l.

This result can be understood intuitively as follows. When the mean μ and variance σ of a normal distribution are estimated from a single observation x, they are estimated as $\hat{\mu} = x$ and $\hat{\sigma} = 0$. Similarly, ξ and κ of a gamma distribution $q(T; \xi, \kappa)$ are estimated from a single observation T as $\hat{\xi} = \frac{1}{T}$ and $\hat{\kappa} = \infty$ corresponding to zero variance. Thus, two or more observations are required to estimate κ.

2.2 Cases with Multiple Observations for Each ξ. Next we consider the case where m observations are given for each $\xi^{(l)}$, which may be distributed according to $k(\xi)$. Let $\{T\} = \{T_1, \ldots, T_m\}$ be the m observations, which are generated from the same distribution specified by ξ and κ. We have N such observations $\{T^{(l)}\}, l = 1, \ldots, N$, with a common κ and different $\xi^{(l)}$. Thus, $\{T_1^{(l)}, \ldots, T_m^{(l)}\}$ are generated from the same firing rate $\xi^{(l)}$. Let us take one $\{T\}$. The probability model can be written as

$$p(\{T\}; \xi, \kappa) = \prod_{i=1}^{m} q(T_i; \xi, \kappa). \tag{2.12}$$

In this case, the score functions for $p(\{T\}; \xi, \kappa)$ become

$$\begin{aligned} u &= \frac{\partial}{\partial \kappa} \log p(\{T\}; \xi, \kappa) \\ &= \sum_{i=1}^{m} (1 - \xi T_i + \log T_i + \log(\kappa\xi) - \phi(\kappa)) \text{ and} \end{aligned} \tag{2.13}$$

$$v = \frac{\partial}{\partial \xi} \log p(\{T\}; \xi, \kappa) = \sum_{i=1}^{m} (-\kappa\xi + \kappa\xi^2 T_i). \tag{2.14}$$

Note that a score function is defined as the derivative of log likelihood with respect to a parameter. Then κ can be estimated by solving the equation as

$$\sum_{l=1}^{N} \sum_{i=1}^{m} (\log T_i^{(l)} + \log(\hat{\kappa}\hat{\xi}^{(l)}) - \phi(\hat{\kappa})) = 0, \tag{2.15}$$

where

$$\hat{\xi}^{(l)} = \frac{1}{\frac{1}{m}\sum_{i=1}^{m} T_i^{(l)}} \tag{2.16}$$

for all l.

As we show numerically later, the maximum likelihood estimator is biased even when an infinite number of observations is given ($N \to \infty$) for a fixed m. In general, in the case where the number of parameters is finite, the maximum likelihood estimator gives an asymptotically consistent estimator. However, as Neyman and Scott (1948) pointed out, when the number of parameters increases with increasing observations, the maximum likelihood estimator is not necessarily asymptotically consistent. To obtain an unbiased estimator of κ, we use the method of estimating functions in what follows.

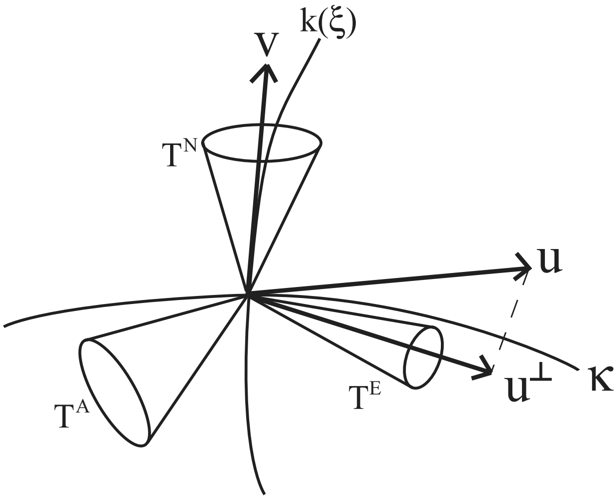

Figure 2: Schematic diagram for finding estimating functions from the viewpoint of information geometry. $u^{\perp}$ can be obtained by projecting u so that $u^{\perp}$ and v's are orthogonal to each other. In information geometry, score functions are represented as tangent vectors. The cone T^N represents the multidimensional space spanned by v's. The cone T^E represents the multidimensional space spanned by $u^{\perp}$'s. Although T^E is one-dimensional in our model, it can be multidimensional in general. The cone T^A represents the zero-mean functions of T that are orthogonal to both T^N and T^E. Note that in information geometry for semiparametric models, tangent vectors are functions and span an infinite-dimensional Hilbert space.

3 Theory of Estimating Functions

We introduce the information-geometric theory of estimating functions developed by Amari and Kawanabe (1997). Their method is based on the global geometry of families of probability distributions (see Figure 2) and provides a general method for finding unbiased estimators for semiparametric problems. However, here we summarize only necessary conditions for obtaining the estimator for our problem.

Let $p(x; \theta, k)$ be a probability density function of a random variable x, where the parameter of interest θ is a scalar and the nuisance parameter k is an infinite-dimensional parameter, typically a function. The purpose is to estimate θ consistently without estimating k.

A function $y(x, \theta)$ that does not depend on k is called an (unbiased) estimating function when it satisfies, for all k,

$$E_{\theta,k}[y(x,\theta)] = 0, \tag{3.1}$$

where $E_{\theta,k}$ denotes the expectation with respect to $p(x;\theta,k)$. When a nontrivial estimating function exists, we have an estimator $\hat{\theta}$ of θ by solving

$$\sum_{l=1}^{N} y(x^{(l)},\hat{\theta}) = 0, \tag{3.2}$$

where $\{x^{(1)},\ldots,x^{(N)}\}$ are N independent observations. As this sample average approximates the expectation with respect to the true probability, if the number of observations is large enough, θ can be estimated asymptotically.

In the maximum likelihood estimation, the interest score function

$$u(x,\theta,k) = \frac{\partial}{\partial\theta}\log p(x;\theta,k) \tag{3.3}$$

played the role of an estimating function in equation 2.5 provided k is known. Note that an interest score function is defined as the derivative of log likelihood with respect to a parameter of interest, which we want to estimate. When k is unknown, the interest score function is not an estimating function in its original form. By differentiating equation 3.1 with respect to k, we get

$$E_{\theta,k}[v\ y(x,\theta)] = 0, \tag{3.4}$$

where

$$v(x,\theta,k) = \frac{\partial}{\partial k}\log p(x;\theta,k) \tag{3.5}$$

is the functional derivative (Fréchet derivative). We call v a nuisance score function because it is the derivative of log likelihood with respect to a "nuisance" parameter, which we do not need to estimate. We define an inner product as

$$a\cdot b = E_{\theta,k}[a\ b]. \tag{3.6}$$

Then equation 3.4 means that $y(x,\theta)$ must be orthogonal to all v's. Note that v is infinite-dimensional when k is a function.

Let us first consider an easier example where k is a scalar. In this case, we can construct $y(x,\theta)$, which is orthogonal to v by projection in the sense of probability as shown in Figure 2. The projection is given as

$$u^{\perp} = u - \frac{u\cdot v}{v\cdot v}v. \tag{3.7}$$

In fact, we have

$$E_{\theta,k}[u^{\perp}] = E_{\theta,k}[u] - \frac{u \cdot v}{v \cdot v} E_{\theta,k}[v] = 0 \text{ and} \tag{3.8}$$

$$u^{\perp} \cdot v = u \cdot v - \frac{u \cdot v}{v \cdot v} v \cdot v = 0, \tag{3.9}$$

where the expectations of the score functions are 0 as

$$E_{\theta,k}[u] = \frac{\partial}{\partial \theta} \int p(x; \theta, k) dx = 0. \tag{3.10}$$

When k is a function, v is infinite-dimensional, and the projection is very difficult to obtain in the closed form except for the case of the mixture model of an exponential form,

$$p(x; \theta, k) = \int q(x; \xi, \theta) k(\xi) d\xi, \tag{3.11}$$

where

$$q(x; \xi, \theta) = \exp\{\xi \cdot s(x, \theta) + r(x, \theta) - \psi(\theta, \xi)\}. \tag{3.12}$$

In this mixture model, $\{\xi^{(1)}, \xi^{(2)}, \ldots\}$ is an unknown sequence where ξ is independently and identically distributed (i.i.d.) according to a probability density function $k(\xi)$. Then lth observation $x^{(l)}$ is distributed according to $q(x^{(l)}; \xi^{(l)}, \theta)$. In effect, x is i.i.d. according to $p(x; \theta, k)$.

The nuisance score function for this model is given as follows. The small deviation of $k(\xi)$ in the direction of $a(\xi)$ can be represented by a curve $k(\xi, t)$ starting from $k(\xi)$,

$$k(\xi, t) = k(\xi) + t a(\xi), \tag{3.13}$$

where t $(0 \leq t < \epsilon)$ is the parameter of the curve. The nuisance score function in the direction of $a(\xi)$ is

$$v = \frac{d}{dt} \log p(x; \theta, k)|_{t=0} = \frac{\int a(\xi) \exp(\xi \cdot s - \psi(\theta, \xi)) d\xi}{\int k(\xi) \exp(\xi \cdot s - \psi(\theta, \xi)) d\xi}. \tag{3.14}$$

Note that the nuisance score functions depend on x only through $s(x, \theta)$. Thus, the vector space spanned by nuisance scores is generated by the random variable $s(x, \theta)$. In this case, we have an estimating function as

$$u^{I} = u - E_{\theta,k}[u|s], \tag{3.15}$$

where $E_{\theta,k}[u|s]$ is the conditional expectation of u conditioned on s. In fact, u^I is orthogonal to any function of $s(x, \theta)$ because

$$E_{\theta,k}[u^I\ f(s)] = \int E_{\theta,k}[u^I\ f(s)|s]p(s)ds = \int E_{\theta,k}[u^I|s]f(s)p(s)ds = 0. \tag{3.16}$$

It has been shown that the projected score function gives an optimal estimator when we estimate only one of many parameters. Thus, u^I gives an efficient estimating function for θ if it does not depend on $k(\xi)$ (Amari & Kawanabe, 1997; Bickel et al., 1993).

There is a special case when $\partial_\theta s$ is a function of s. In this case, the estimating function becomes simple as

$$\begin{aligned} u^I &= \{\partial_\theta s - E[\partial_\theta s|s]\} \cdot E_\xi[\xi|s] + \partial_\theta r - E[\partial_\theta r|s] \\ &= \partial_\theta r - E[\partial_\theta r|s], \end{aligned} \tag{3.17}$$

where

$$E_\xi[\xi|s] = \frac{\int \xi k(\xi)\exp(\xi \cdot s - \psi)d\xi}{\int k(\xi)\exp(\xi \cdot s - \psi)d\xi}. \tag{3.18}$$

4 Estimation by Estimating Functions

4.1 Simple Case. We considered the following statistical model of interspike intervals proposed by Ikeda (2005). Interspike intervals are distributed according to a gamma distribution whose mean firing rate changes over time. The mean firing rate ξ at each time is determined randomly according to an unknown probability density $k(\xi)$. To demonstrate that the model is of the exponential form defined by Amari and Kawanabe (1997), we define s, r, and ψ as

$$s(T, \kappa) = -\kappa T, \tag{4.1}$$

$$r(T, \kappa) = (\kappa - 1)\log(T), \quad \text{and} \tag{4.2}$$

$$\psi(\kappa, \xi) = -\kappa \log(\xi\kappa) + \log\Gamma(\kappa). \tag{4.3}$$

Here, T denotes an interspike interval. The model is described by

$$p(T; \kappa, k(\xi)) = \int q(T; \xi, \kappa)k(\xi)d\xi, \tag{4.4}$$

where

$$q(T;\xi,\kappa) = \frac{(\xi\kappa)^\kappa}{\Gamma(\kappa)} T^{\kappa-1} e^{-\xi\kappa T} = e^{\xi s(T,\kappa)+r(T,\kappa)-\psi(\kappa,\xi)}. \quad (4.5)$$

Note that this type of model is called a semiparametric model because it has both an unknown finite parameter κ, which is a scalar in this case, and a function $k(\xi)$.

To estimate κ without estimating $k(\xi)$, let us calculate the estimating function according to the method shown in the previous section. It was shown there that for the mixture distributions of an exponential form, the estimating function u^I is given by the projection of the score function, $u = \partial_\kappa \log p$, as

$$\begin{aligned} u^I(T,\kappa) &= u - E[u|s] \\ &= (\partial_\kappa s - E[\partial_\kappa s|s]) \cdot E[\xi|s] + \partial_\kappa r - E[\partial_\kappa r|s] \\ &= \partial_\kappa r - E[\partial_\kappa r|s], \end{aligned} \quad (4.6)$$

where the relation

$$E[\partial_\kappa s|s] = \frac{s}{\kappa} = -T = \partial_\kappa s \quad (4.7)$$

holds because the numbers of random variables T and s are the same. For the same reason,

$$E[\partial_\kappa r|s] = \log(T) = \partial_\kappa r. \quad (4.8)$$

Then,

$$u^I = 0. \quad (4.9)$$

This means that the set of estimating functions is empty. The restrictions imposed on u^I are the necessary conditions that the estimating function must satisfy by definition (Amari & Kawanabe, 1997). Therefore, we have proved that no estimating function of κ exists for the model.

Two or more random variables may be needed. We may generalize the model such that ξ_i's are not independently generated but are related. Let us consider the multivariate model described by

$$p(T_1,\ldots,T_n;\kappa,k(\xi_1,\ldots,\xi_n)) = \int \prod_{i=1}^{n} q(T_i;\xi_i,\kappa)k(\xi_1,\ldots,\xi_n)d\xi. \quad (4.10)$$

In this case, s_i, r, and ψ are defined as

$$s_i(T, \kappa) = -\kappa T_i, \tag{4.11}$$

$$r(T, \kappa) = (\kappa - 1)\sum_{i=1}^{n} \log(T_i), \quad \text{and} \tag{4.12}$$

$$\psi(\kappa, \xi) = -\kappa \sum_{i=1}^{n} \log(\xi_i \kappa) + n \log \Gamma(\kappa). \tag{4.13}$$

The numbers of random variables T_i's and s_i's are also the same. Then u^I becomes an empty set. This result implies that two or more observations are needed for each ξ.

4.2 Cases with Multiple Observations for Each ξ. Let us consider the case where we have multiple observations for each ξ. Here, a consistent estimator of κ exists.

Let $\{T\} = \{T_1, \ldots, T_m\}$ be the m observations that have the same firing rate ξ. The probability model can be written as

$$p(\{T\}; \kappa, k(\xi)) = \int \prod_{i=1}^{m} q(T_i; \xi, \kappa) k(\xi) d\xi, \tag{4.14}$$

where

$$\prod_{i=1}^{m} q(T_i; \xi, \kappa) = \prod_{i=1}^{m} \frac{(\xi\kappa)^{\kappa}}{\Gamma(\kappa)} T_i^{\kappa-1} e^{-\xi\kappa T_i} = e^{\xi \cdot s(\{T\},\kappa) + r(\{T\},\kappa) - \psi(\kappa,\xi)}. \tag{4.15}$$

We define s, r, and ψ as

$$s(\{T\}, \kappa) = -\kappa \sum_{i=1}^{m} T_i, \tag{4.16}$$

$$r(\{T\}, \kappa) = (\kappa - 1)\sum_{i=1}^{m} \log(T_i), \quad \text{and} \tag{4.17}$$

$$\psi(\kappa, \xi) = -m\kappa \log(\xi\kappa) + m \log \Gamma(\kappa). \tag{4.18}$$

Then the estimating function is given by

$$\begin{aligned} u^I(\{T\}, \kappa) &= u - E[u|s] \\ &= (\partial_\kappa s - E[\partial_\kappa s|s]) \cdot E[\xi|s] + \partial_\kappa r - E[\partial_\kappa r|s] \end{aligned}$$

$$
\begin{aligned}
&= \partial_\kappa r - E[\partial_\kappa r|s] \\
&= \sum_{i=1}^{m} \log(T_i) - \sum_{i=1}^{m} E[\log(T_i)|s] \\
&= \sum_{i=1}^{m} \log(T_i) - mE[\log(T_1)|s],
\end{aligned} \tag{4.19}
$$

where we used

$$
E[\partial_\kappa s|s] = \frac{s}{\kappa} = \partial_\kappa s. \tag{4.20}
$$

The last equality in equation 4.19 holds because of the permutation symmetry among T's. The conditional expectation of $\log T_1$ is given as (see the appendix)

$$
E[\log(T_1)|s] = \log\left(-\frac{s}{\kappa}\right) - \phi(m\kappa) + \phi(\kappa), \tag{4.21}
$$

where the digamma function is defined as

$$
\phi(\kappa) = \frac{\Gamma'(\kappa)}{\Gamma(\kappa)}. \tag{4.22}
$$

Note that $E[\log(T_1)|s]$ does not depend on the unknown function $k(\xi)$. Thus, we have

$$
u^I(\{T\}, \kappa) = \sum_{i=1}^{m} \log(T_i) - m \log\left(\sum_{i=1}^{m} T_i\right) + m\phi(m\kappa) - m\phi(\kappa). \tag{4.23}
$$

The form of u^I can be understood as follows. If we scale T as $t = \xi T$, we have $E[t] = 1$. Then we can show that u^I does not depend on ξ because

$$
\sum_{i=1}^{m} \log(T_i) - m \log\left(\sum_{i=1}^{m} T_i\right) = \sum_{i=1}^{m} \log(t_i) - m \log\left(\sum_{i=1}^{m} t_i\right). \tag{4.24}
$$

This implies that we can estimate κ without estimating ξ.
κ can be estimated consistently from N independent sets of observations, $\{T^{(l)}\} = \{T_1^{(l)}, \ldots, T_m^{(l)}\}, l = 1, \ldots, N$, as the value of κ that solves

$$
\sum_{l=1}^{N} u^I(\{T^{(l)}\}, \hat{\kappa}) = 0. \tag{4.25}
$$

In fact, the expectation of u^I is 0 independent of $k(\xi)$:

$$E[u^I] = \int E[u - E[u|s]|s]p(s)ds = 0 \tag{4.26}$$

u^I yields an efficient estimating function. An efficient estimator is one whose variance attains the Cramér-Rao lower bound asymptotically. Thus, there is no estimator of κ whose mean square estimation error is smaller than that given by u^I. As u^I does not depend on $k(\xi)$, it is the optimal estimating function whatever $k(\xi)$ is, or whatever the sequence $\xi^{(1)}, \ldots, \xi^{(N)}$ is.

Maximum likelihood estimation for this problem gives an estimating function as

$$u^{MLE} = \sum_{i=1}^{m} \log(T_i) + m \log(\hat{\xi}) + m \log \kappa - m\phi(\kappa), \tag{4.27}$$

where

$$\frac{1}{\hat{\xi}} = \frac{1}{m} \sum_{i=1}^{m} T_i. \tag{4.28}$$

u^{MLE} is similar to u^I but different in terms of the constant

$$u^{MLE} - u^I = m \log(m\kappa) - m\phi(m\kappa). \tag{4.29}$$

As a result, the maximum likelihood estimator $\hat{\kappa}$ is biased.

Figure 3 shows the biases for the maximum likelihood estimation and the proposed estimation. The maximum likelihood estimation is biased even when an infinite number of observations are given while the estimating function is asymptotically unbiased. In the numerical calculation, we used the model with $\kappa = 4$ and $m = 2$. An interspike interval with firing rate ξ can be generated as follows. First, a normalized interspike interval t is generated according to the standard gamma distribution with $\xi = 1$. Second, an interspike interval T is obtained by dividing t by ξ: $T = t/\xi$. Note that $\log T = \log t - \log \xi$. Then, as shown in equation 4.24, y does not depend on ξ and $f(\xi)$ at all. Therefore, we fixed ξ to be 1 without loss of generality. The figure is obtained as follows. We generated T's according to the standard gamma distribution and substituted them into equation 4.27 to estimate κ. We repeated the estimation many times ($n = 10^4$) for each number of observations and calculated the mean (bias) and the quartiles of $\hat{\kappa}$. Note that the result does not depend on $k(\xi)$.

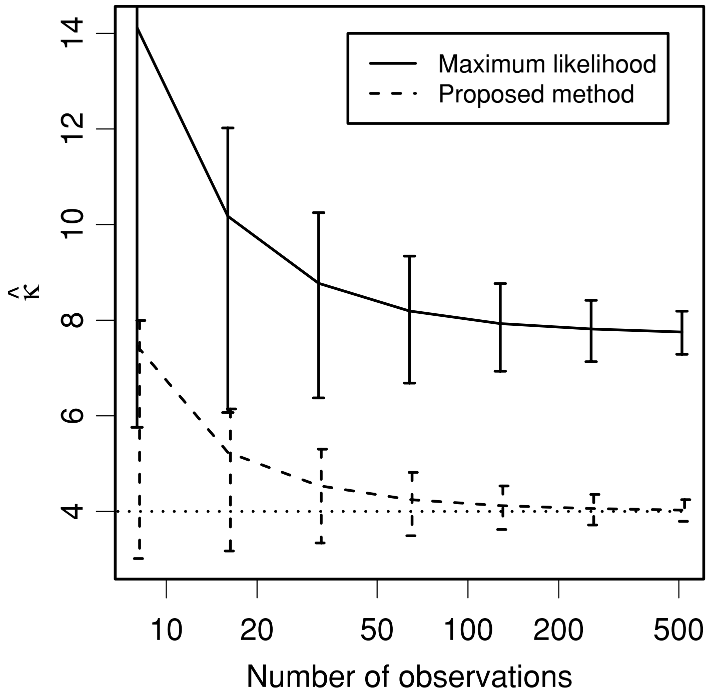

Figure 3: Biases of $\hat{\kappa}$ for maximum likelihood estimation and proposed method for $m = 2$. The dotted line represents the true value, $\kappa = 4$. The error bars represent the quartiles for repeated trials. The estimation was repeated 10^4 times for each number of observations. The maximum likelihood estimation is biased even when an infinite number of observations are given while the estimating function is asymptotically unbiased.

We examined whether our estimator works when κ is close to zero. Figure 4 plots the biases for $\kappa = 0.5$. The detail of Figure 4 is the same as that of Figure 3. Figure 4 demonstrates that the proposed estimator also works for $\kappa = 0.5$.

5 Cases Where the Firing Rate Is Continuously Modulated

So far, we have considered only the cases in which two or more consecutive firing rates are the same. In this section, we remove this assumption and consider more general cases where consecutive firing rates are not necessarily the same but the firing rate continuously changes slowly. Although the assumption of the statistical model is violated, we try to estimate κ by the proposed method heuristically. We

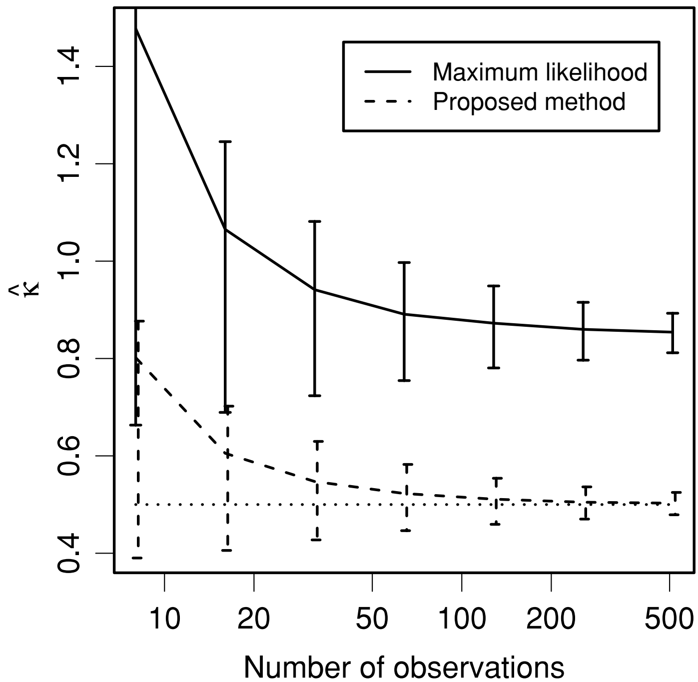

Figure 4: Biases of $\hat{\kappa}$ for maximum likelihood estimation and proposed method for $m = 2$. The dotted line represents the true value, $\kappa = 0.5$. The error bars represent the quartiles for repeated trials. The estimation was repeated 10^4 times for each number of observations. The maximum likelihood estimation is biased even when an infinite number of observations are given while the estimating function is asymptotically unbiased. The proposed estimator works even if κ is close to zero.

compare the estimating functions with various m and the measure of spiking irregularity, which we will introduce based on the estimating function.

5.1 Measure of Spiking Irregularity. In this section, we introduce a practical measure of spiking irregularity for experimental data based on the estimating function. The new measure may be useful in the case where the firing rate continuously changes slowly. For experimental data, the assumption that consecutive m interspike intervals have the same or similar ξ is most probable for $m = 2$. Therefore, we set $m = 2$ in the estimating function. Let $\{T_1, T_2, \ldots, T_N\}$ form a single spike train, where T_i denotes the ith interspike interval. Let N be odd. There are two types of possible estimating equations depending on the choice of the starting point: (1)

$$\frac{1}{\frac{N-1}{2}} \sum_{i=1}^{\frac{N-1}{2}} \frac{1}{2} \log \left(\frac{4T_{2i-1}T_{2i}}{(T_{2i-1} + T_{2i})^2} \right) - \log 2 + \phi(2\hat{\kappa}) - \phi(\hat{\kappa}) = 0, \tag{5.1}$$

where $\xi_1 = \xi_2, \xi_3 = \xi_4, \ldots, \xi_{N-2} = \xi_{N-1}$ are assumed, and (2)

$$\frac{1}{\frac{N-1}{2}} \sum_{i=1}^{\frac{N-1}{2}} \frac{1}{2} \log \left(\frac{4T_{2i}T_{2i+1}}{(T_{2i} + T_{2i+1})^2} \right) - \log 2 + \phi(2\hat{\kappa}) - \phi(\hat{\kappa}) = 0, \tag{5.2}$$

where $\xi_2 = \xi_3, \xi_4 = \xi_5, \ldots, \xi_{N-1} = \xi_N$ are assumed. We take the average of these equations as

$$\frac{1}{N-1} \sum_{i=1}^{N-1} \frac{1}{2} \log \left(\frac{4T_i T_{i+1}}{(T_i + T_{i+1})^2} \right) - \log 2 + \phi(2\hat{\kappa}) - \phi(\hat{\kappa}) = 0. \tag{5.3}$$

We estimate κ by solving this equation for $\hat{\kappa}$ numerically. To avoid troublesome numerical iterations, we suggest using part of the estimating equation

$$S_I \equiv -\frac{1}{N-1} \sum_{i=1}^{N-1} \frac{1}{2} \log \left(\frac{4T_i T_{i+1}}{(T_i + T_{i+1})^2} \right), \tag{5.4}$$

as a measure of spiking irregularity. $\phi(2\kappa) - \phi(\kappa)$ is monotonic, so that it is clear that we can easily solve equation 5.3 for κ. The correspondence between $\hat{\kappa}$ and S_I is shown in Figure 5.

Note that we assumed $\xi_1 = \xi_2$, $\xi_2 = \xi_3$, and so on. Therefore, unless all ξ_i's are the same, $\hat{\kappa}$ is biased. However, when the firing rate changes slowly enough, $\hat{\kappa}$ is approximately correct, as we will show in the example below. In addition, S_I is similar to the measure of local variation L_V, which is known to be useful for cell classification (Shinomoto, Miyazaki, et al., 2005; Shinomoto et al., 2003). Then, S_I may also be useful for cell classification.

The measure of local variation (Shinomoto et al., 2003),

$$L_V = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{3(T_i - T_{i+1})^2}{(T_i + T_{i+1})^2} = 3 - \frac{12}{N-1} \sum_{i=1}^{N-1} \frac{T_i T_{i+1}}{(T_i + T_{i+1})^2} \tag{5.5}$$

looks similar to the measure of spiking irregularity S_I. In fact, there exists a tight inequality between these measures. By using Jensen's inequality, we

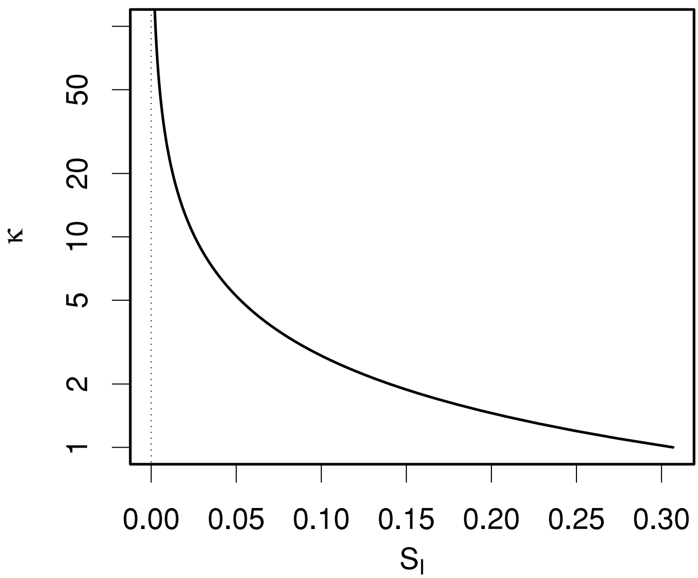

Figure 5: Correspondence between $\hat{\kappa}$ and S_I. The lower bound of S_I is 0. For $\kappa = 1$, S_I is $1 - \log 2 = 0.307$.

obtain

$$-\frac{1}{N-1}\sum_{i=1}^{N-1}\frac{1}{2}\log\left(\frac{4T_iT_{i+1}}{(T_i+T_{i+1})^2}\right) \geq -\frac{1}{2}\log\left(\frac{1}{N-1}\sum_{i=1}^{N-1}\frac{4T_iT_{i+1}}{(T_i+T_{i+1})^2}\right)$$
$$= -\frac{1}{2}\log\left(1-\frac{L_V}{3}\right). \tag{5.6}$$

Thus, L_V gives a lower bound of S_I. Since S_I and $\hat{\kappa}$ are inversely related, as shown in Figure 5, L_V gives an upper bound on $\hat{\kappa}$. This relation may be useful when only the value of L_V is available in the existing literature. Note that this relation holds for any spike trains independent of their statistical models and for any N.

5.2 AR Model. As an example of a rate-modulated case, let us consider an AR model in which ξ_i's are given as

$$\log \xi_{i+1} = e^{-\frac{1}{\tau}} \log \xi_i + \Delta\sqrt{1-e^{-\frac{2}{\tau}}}\sigma_{i+1}, \tag{5.7}$$

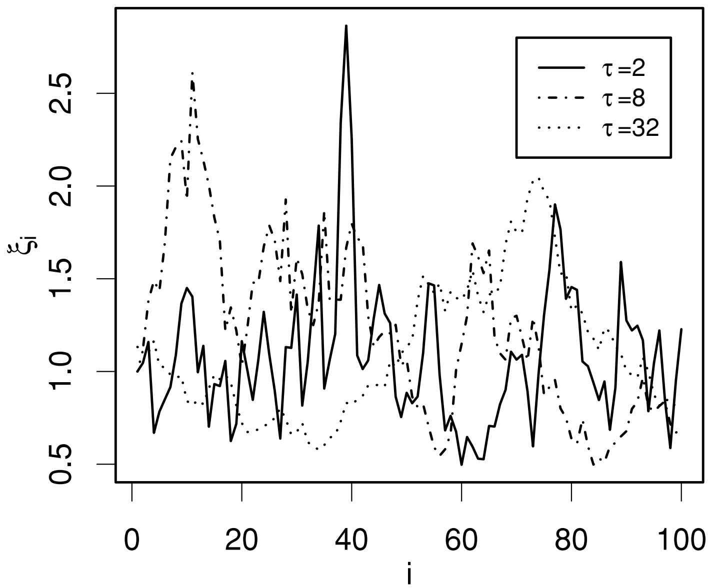

Figure 6: Sample processes of ξ for AR model with $\Delta = 0.3$. The larger the time constant, τ, the longer the correlations.

where σ_i's are independently and identically distributed according to the standard normal distribution with mean 0 and variance 1. Here, we assumed that $\log \xi_i$'s obey a gaussian process so that ξ_i's are nonnegative. τ represents the time correlation length of $\log \xi$:

$$\langle \log \xi_{i+j} \log \xi_i \rangle = \Delta^2 e^{-\frac{j}{\tau}}. \tag{5.8}$$

The coefficient $\sqrt{1 - e^{-\frac{2}{\tau}}}$ is multiplied to the noise so that the variances of $\log \xi_i$'s are Δ independent of τ. Thus, we have two free parameters, τ and Δ. Figure 6 illustrates sample paths for $\Delta = 0.3$. The figure shows that the larger the time constant τ, the longer the correlation of ξ.

Next, we generated interspike intervals by using these ξ_i's and estimated κ from them. Figure 7 plots the results of numerical calculation. The dotted line represents the true value, $\kappa = 4$. S_I denotes the estimation given by S_I using equation 5.3. Note that what is plotted is not S_I, but rather the estimate of κ obtained putting S_I through the function shown in Figure 5. $m = 2$ denotes the estimation given by the estimating function with $m = 2$ using equation 5.1. Note that $m = 2$ refers to estimating κ without averaging over the two types of pairing: $\xi_1 = \xi_2, \xi_3 = \xi_4, \ldots$ and $\xi_2 = \xi_3, \xi_4 = \xi_5, \ldots$ as S_I does. The figure shows that biases exist even if the number of observations

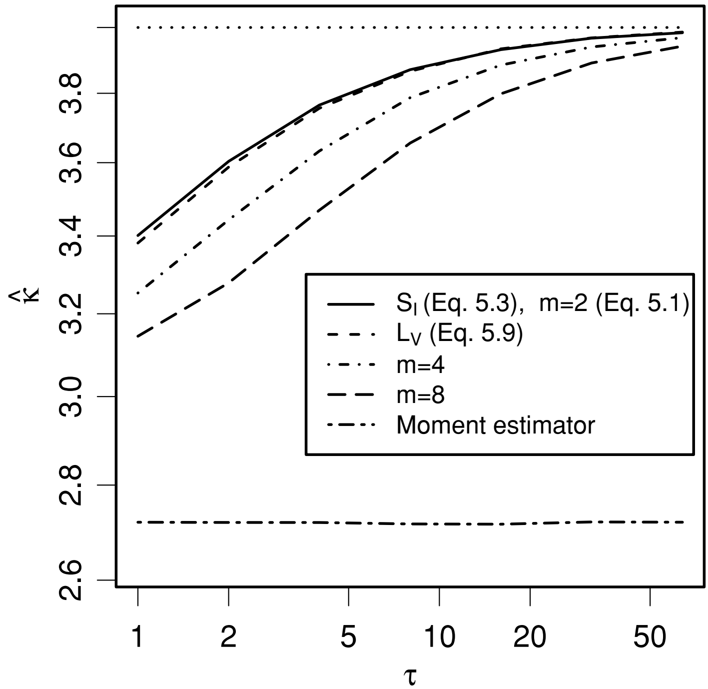

Figure 7: Asymptotic biases for AR model with $\Delta = 0.3$. The dotted line represents the true value, $\kappa = 4$. $m = 2$ denotes the estimation obtained using the estimating function with $m = 2$. S_I denotes the estimation given by S_I. $m = 2$ and S_I give the same result. L_V denotes the estimation given by L_V. Moment estimator denotes the estimation by the moment estimator. The number of observations is $N = 10^7$. The biases decrease with increasing τ. $m = 2$ (and S_I) always gives the smallest bias.

is infinite, although they decrease with increasing τ. The bias increases with increasing m. It is not the case that there is an optimal m depending on τ, but the estimating function with $m = 2$ always gives the smallest bias. The bias given by the estimating function with $m = 2$ and that given by S_I are the same because the AR model does not distinguish ξ_i's with even and odd i. Note that the biases in Figure 7 are much smaller than those in Figure 3 for the maximum likelihood estimation. Although we fixed $\Delta = 0.3$ in Figure 7, the bias increases as Δ increases.

We also estimated κ by the moment estimator:

$$\hat{\kappa} = Mean(T)^2/\text{Variance}(T).$$

If the firing rate is constant over time, the mean and variance are given by equations 2.2, and κ can be estimated correctly. Figure 7 shows that the moment estimator is always the worst independent of τ. In fact, because the moment estimator assumes that the firing rate is constant over time, it does not properly capture the spiking irregularity κ when the firing rate changes over time.

κ can also be estimated by using L_V:

$$\hat{\kappa} = \frac{3}{2L_V} - \frac{1}{2}. \tag{5.9}$$

Note that the expectation value of L_V for the gamma distribution is $\overline{L_V} = \frac{3}{2\kappa+1}$ (Shinomoto et al., 2003). Figure 7 shows that the estimate of κ obtained by putting L_V through equation 5.9 is slightly more biased than that given by S_I. We also compared the variances directly in the limit of large τ where there is no bias. The variance of the estimate given by L_V is larger than that given by S_I although the difference is only about 3%.

Figure 8 shows how bias and variability of the estimator for the AR model depend on the number of observations. The true value is $\kappa = 4$. The estimation by S_I gives a smaller bias than the estimating function with $m = 2$ for finite observations. This can be intuitively understood as follows. The number of terms summed in S_I is about twice as large as that in the estimating function with $m = 2$. Then, because the bias decreases with the number of observations, the bias given by S_I is smaller. The bias for the estimating function with $m = 4$ is the smallest in a certain range of the number of observations. In fact, the bias for $m = 4$ becomes 0 for a certain number of observations, because the asymptotic bias is always negative while the bias due to finite observations is always positive, as shown in Figure 3. However, it is not a good way to fix the number of observations because the variance of the estimator decreases as the numbers of observations increases, as we show next.

So far, we have considered only the biases of the estimators. However, it is also important to know the trial-to-trial variability of the estimators. The smaller the variability, the better the estimate. The error bars in Figure 8 show the quartiles of the estimators of κ for repeated trials. The estimation was repeated 10^4 times for each number of observations. While the variability decreases with increasing m, the bias increases with increasing m. Thus, there is a trade-off between bias and variability. The variability of the estimator given by S_I is smaller than that given by the estimating function with $m = 2$.

Thus, S_I gives a relatively good estimator of κ for the AR model. In particular, as far as the bias is concerned, S_I looks optimal among the estimators we considered here. We believe that S_I generally works well in many models, including the Markov model in which the firing rate changes slowly because the Markov model does not distinguish even and odd i for T_i.

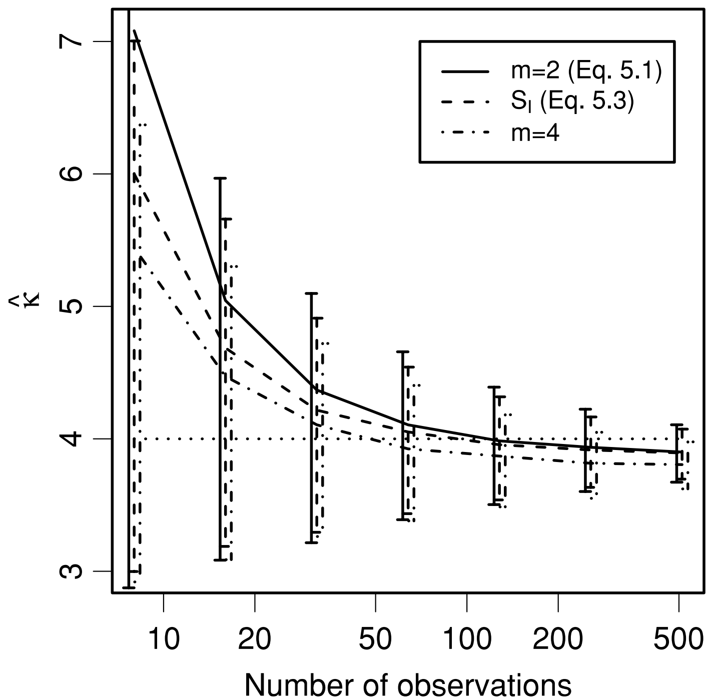

Figure 8: Bias and variability for AR model with finite observations. The dotted line represents the true value, $\kappa = 4$. $\Delta = 0.3$ and $\tau = 8$. $m = 2$ denotes the estimation obtained using the estimating function with $m = 2$. SI denotes the estimation by S_I. The error bars represent the quartiles for repeated trials. The estimation was repeated 10^4 times for each number of observations. The bias for S_I is smaller than that for $m = 2$ for finite observations.

6 Summary and Discussion

We estimated the shape parameter κ of the semiparametric model suggested by Ikeda (2005) without estimating the firing rate ξ. The maximum likelihood estimator is not consistent for this problem because the number of nuisance parameters ξ_l increases with increasing observations of interspike interval T. We showed that the model is of the exponential form defined by Amari and Kawanabe (1997) and can be analyzed by a method of estimating functions for semiparametric models. We found that an estimating function does not exist unless multiple observations are given for each firing rate ξ. If multiple observations are given, the method of estimating functions can be applied. In that case, the estimating function of κ can be analytically obtained, and κ can be estimated consistently independent of the functional form of the firing rate $k(\xi)$. In general, the estimating function is not efficient.

However, this method provided an optimal estimator in the sense of Fisher information for our problem. That is, we obtained an efficient estimator. We suggested the measure of spiking irregularity based on the estimating function, which may be useful for characterizing individual neurons in the case where only a single observation is given for each firing rate.

Various measures for spiking randomness have been used in previous studies. The coefficient of variation C_V is the global variance normalized by the mean interspike interval (Holt et al., 1996):

$$C_V = \frac{\sqrt{\mathrm{Var[T]}}}{\overline{T}}. \tag{6.1}$$

Although C_V increases with increasing irregularity, it also becomes large when the firing rate changes over time. Thus, we cannot distinguish these two cases according to C_V. The local variation of interspike intervals L_V is locally normalized and relatively independent of the firing rate change (Shinomoto, Miyazaki et al., 2005; Shinomoto et al., 2003).

However, it is an ad hoc measure and has no corresponding parameter in statistical models. The measure of spiking irregularity S_I that we introduced in this article is an estimator of κ, relatively independent of the firing rate change. For these reasons, we suggest using S_I as a measure of spiking irregularity.

In this article, we focused on the semiparametric model, in which the firing rate may vary for each interspike interval. However, a better fit of experimental data could be obtained by using other models, depending on situations. For instance, the firing rate can be assumed to be a function of continuous time (Baker & Lemon, 2000; Brown, Barbieri, Ventura, Kass, & Frank, 2002) or early stages of sensory cortex can be explained by more deterministic models such as a noisy leaky integrate-and-fire model (Reich, Victor, & Knight, 1998). The selection of an appropriate model (Brown et al., 2002; Reich et al., 1998) is very important and will depend on the recorded area and its state.

It is important to know to what extent the proposed estimator is robust in the sense that if the ISI distribution is close to a gamma distribution but not precisely gamma. We believe that the robustness can be evaluated numerically or analytically. However, it is beyond the scope of this letter. We leave it for future work.

Appendix: Calculation of $E[\log(T_1)|s]$ in Equation 4.21

To calculate $E[\log(T_1)|s]$ in equation 4.21 let us use Bayes theorem:

$$p(T|s) = \frac{p(T,s)}{\int p(T,s)dT} = \frac{p(T,s)}{p(s)}. \tag{A.1}$$

By repeating beta integrals,

$$\int_0^1 x^{a-1}(1-x)^{b-1}dx = B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{(a-1)!(b-1)!}{(a+b-1)!}, \qquad \text{(A.2)}$$

we obtain

$$\begin{aligned}
& p(s) \\
&= \int \delta\left(s + \kappa\sum_{i=1}^{m} T_i\right)\prod_{i=1}^{m}(q(T_i;\xi,\kappa)dT_i)k(\xi)d\xi \\
&= \int q\left(-\frac{s}{\kappa} - \sum_{i=1}^{m-1} T_i\right)\prod_{i=1}^{m-1}(q(T_i;\xi,\kappa)dT_i)\frac{k(\xi)}{\kappa}d\xi \\
&= \int \left[\frac{(\xi\kappa)^\kappa}{\Gamma(\kappa)}\left(-\frac{s}{\kappa} - \sum_{i=1}^{m-1} T_i\right)^{\kappa-1} e^{-\xi\kappa(-\frac{s}{\kappa}-\sum_{i=1}^{m-1}T_i)}\right] \\
&\quad \times \prod_{i=1}^{m-1}\left(\frac{(\xi\kappa)^\kappa}{\Gamma(\kappa)}T_i^{\kappa-1}e^{-\xi\kappa T_i}dT_i\right)\frac{k(\xi)}{\kappa}d\xi \\
&= \int \left[\left(-\frac{s}{\kappa} - \sum_{i=1}^{m-2} T_i\right)^{\kappa-1}\left(1 - \frac{T_{m-1}}{-\frac{s}{\kappa} - \sum_{i=1}^{m-2} T_i}\right)^{\kappa-1}\right] \\
&\quad \times \prod_{i=1}^{m-1}\left(T_i^{\kappa-1}dT_i\right)\frac{e^{s\xi}(\xi\kappa)^{m\kappa}}{\Gamma(\kappa)^m}\frac{k(\xi)}{\kappa}d\xi \\
&= \int \left[\int\left(1 - \frac{T_{m-1}}{-\frac{s}{\kappa} - \sum_{i=1}^{m-2} T_i}\right)^{\kappa-1} T_{m-1}^{\kappa-1}dT_{m-1}\right] \\
&\quad \times \left(-\frac{s}{\kappa} - \sum_{i=1}^{m-2} T_i\right)^{\kappa-1}\prod_{i=1}^{m-2}\left(T_i^{\kappa-1}dT_i\right)\frac{e^{s\xi}(\xi\kappa)^{m\kappa}}{\Gamma(\kappa)^m}\frac{k(\xi)}{\kappa}d\xi \\
&= \int B(\kappa,\kappa)\left(-\frac{s}{\kappa} - \sum_{i=1}^{m-2} T_i\right)^{2\kappa-1}\prod_{i=1}^{m-2}\left(T_i^{\kappa-1}dT_i\right)\frac{e^{s\xi}(\xi\kappa)^{m\kappa}}{\Gamma(\kappa)^m}\frac{k(\xi)}{\kappa}d\xi \\
&= \dots \\
&= \int \left[\int\left(-\frac{s}{\kappa} - T_1\right)^{(m-1)\kappa-1} T_1^{\kappa-1}dT_1\right]\prod_{i=1}^{m-2} B(i\kappa,\kappa)\frac{e^{s\xi}(\xi\kappa)^{m\kappa}}{\Gamma(\kappa)^m}\frac{k(\xi)}{\kappa}d\xi \\
&= \prod_{i=1}^{m-1} B(i\kappa,\kappa)\frac{(-s)^{m\kappa-1}}{\Gamma(\kappa)^m}\int \xi^{m\kappa}e^{s\xi}k(\xi)d\xi. \qquad \text{(A.3)}
\end{aligned}$$

Similarly, by Taylor expansion of $\log(T_1)$ about $T_1 = -\frac{s}{\kappa}$, we get

$$
\begin{aligned}
&E[\log(T_1)|s] \\
&= \int \log(T_1)\delta\left(s+\kappa\sum_{i=1}^{m}T_i\right)\prod_{i=1}^{m}(q(T_i;\xi,\kappa)dT_i)k(\xi)d\xi\,\frac{1}{p(s)} \\
&= \int \log(T_1)\left(-\frac{s}{\kappa}-T_1\right)^{(m-1)\kappa-1} T_1^{\kappa-1}dT_1\prod_{i=1}^{m-2}B(i\kappa,\kappa) \\
&\quad\times\frac{e^{s\xi}(\xi\kappa)^{m\kappa}}{\Gamma(\kappa)^m}\frac{k(\xi)}{\kappa}d\xi\,\frac{1}{p(s)} \\
&= \log\left(-\frac{s}{\kappa}\right)+\int\sum_{j=1}^{\infty}\frac{-1}{j}B((m-1)\kappa+j,\kappa)\left(-\frac{s}{\kappa}\right)^{m\kappa-1}\prod_{i=1}^{m-2}B(i\kappa,\kappa) \\
&\quad\times\frac{e^{s\xi}(\xi\kappa)^{m\kappa}}{\Gamma(\kappa)^m}\frac{k(\xi)}{\kappa}d\xi\,\frac{1}{p(s)} \\
&= \log\left(-\frac{s}{\kappa}\right)-\sum_{j=1}^{\infty}\frac{1}{j}\frac{B((m-1)\kappa+j,\kappa)}{B((m-1)\kappa,\kappa)}.
\end{aligned}
\tag{A.4}
$$

Next we show that

$$
\sum_{j=1}^{\infty}\frac{1}{j}\frac{B((m-1)\kappa+j,\kappa)}{B((m-1)\kappa,\kappa)} = \phi(m\kappa)-\phi(\kappa), \tag{A.5}
$$

where the digamma function is defined as

$$
\phi(\kappa) = \frac{\Gamma'(\kappa)}{\Gamma(\kappa)}. \tag{A.6}
$$

Let κ be an integer. We define I_l as

$$
I_l = \sum_{j=1}^{\infty}\frac{1}{j}\frac{(m\kappa-1)(m\kappa-2)\cdots((m-1)\kappa+l)}{(m\kappa+j-1)(m\kappa+j-2)\cdots((m-1)\kappa+j+l)}. \tag{A.7}
$$

Then the infinite series can be rewritten as

$$
\sum_{j=1}^{\infty}\frac{1}{j}\frac{B((m-1)\kappa+j,\kappa)}{B((m-1)\kappa,\kappa)} = I_0. \tag{A.8}
$$

By repeatedly using the following equation,

$$I_l = I_{l+1} - \frac{1}{\kappa + l - 1}, \tag{A.9}$$

we get

$$I_0 = I_{\kappa-1} - \sum_{l=0}^{\kappa-2} \frac{1}{\kappa - 1 - l} = \sum_{j=1}^{\infty} \frac{1}{j} \frac{m\kappa - 1}{m\kappa + j - 1} - \sum_{l=1}^{\kappa-1} \frac{1}{l} = \sum_{l=1}^{m\kappa-1} \frac{1}{l} - \sum_{l=1}^{\kappa-1} \frac{1}{l}. \tag{A.10}$$

The last equality follows by telescoping of the infinite sum.

By using the formula for harmonic series (Havil, 2003),

$$\sum_{l=1}^{n} \frac{1}{l} = \gamma + \phi(n+1), \tag{A.11}$$

where Euler's constant is $\gamma = 0.57721\cdots$, we get

$$\sum_{j=1}^{\infty} \frac{1}{j} \frac{B((m-1)\kappa + j, \kappa)}{B((m-1)\kappa, \kappa)} = \phi(m\kappa) - \phi(\kappa). \tag{A.12}$$

Although we assumed that κ is an integer during the proof, the numerical calculation shows that the result also holds for noninteger κ. Thus, we obtain

$$E[\log(T_1)|s] = \log(-\frac{s}{\kappa}) - \phi(m\kappa) + \phi(\kappa). \tag{A.13}$$

Note that $E[\log(T_1)|s]$ does not depend on the unknown function $k(\xi)$.

Acknowledgments

This work was supported in part by grants from the Japan Society for the Promotion of Science (Nos. 14084212 and 16500093).

References

Amari, S. (1982). Differential geometry of curved exponential families—curvatures and information loss. *Ann. Statist.*, *10*, 357–385.

Amari, S. (1985). *Differential-geometrical methods in statistics*. New York: Springer-Verlag.

Amari, S. (1987). Dual connections on the Hilbert bundles of statistical models. In C. T. J. Dodson (Ed.), *Geometrization of statistical theory*. Lancaster: University of Lancaster, Department of Mathematics.

Amari, S. (1998). Natural gradient works efficiently in learning. *Neural Comput.*, *10*, 251–276.

Amari, S., & Kawanabe, M. (1997). Information geometry of estimating functions in semi-parametric statistical models. *Bernoulli*, *3*, 29–54.

Amari, S., & Kumon, M. (1988). Estimation in the presence of infinitely many nuisance parameters—geometry of estimating functions. *Ann. Statist.*, *16*, 1044–1068.

Amari, S., Kurata, K., & Nagaoka, H. (1992). Information geometry of Boltzmann machines. *IEEE Trans. on Neural Networks*, *3*, 26–271.

Amari, S., & Nagaoka, H. (2001). *Methods of information geometry*. Providence, RI: American Mathematical Society.

Baker, S. N., & Lemon, R. N. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J. Neurophysiol.*, *84*, 1770–1780.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., & Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Baltimore, MD: Johns Hopkins University Press.

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput.*, *14*, 325–346.

Cox, D. R., & Lewis, P. A. W. (1966). *The statistical analysis of series of events*. London: Methuen.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, *31*, 1208–1211.

Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika*, *63*, 277–284.

Godambe, V. P. (Ed.). (1991). *Estimating functions*. New York: Oxford University Press.

Groeneboom, P., & Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Basel: Birkhauser.

Havil, J. (2003). *Gamma: Exploring Euler's constant*. Princeton, NJ: Princeton University Press.

Holt, G. R., Softky, W. R., Koch, C., & Douglas, R. J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *J. Neurophysiol.*, *75*, 1806–1814.

Ikeda, K. (2005). Information geometry of interspike intervals in spiking neurons. *Neural Comput.*, *17*, 2719–2735.

Mukhopadhyay, P. (2004). *An introduction to estimating functions*. Harrow: Alpha Science International.

Murray, M. K., & Rice, J. W. (1993). *Differential geometry and statistics*. New York: Chapman and Hall.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, *32*, 1–32.

Pfanzagl, J. (1990). *Estimation in semiparametric models*. Berlin: Springer-Verlag.

Reich, D. S., Victor, J. D., & Knight, B. W. (1998). The power ratio and the interval map: Spiking models and extracellular recordings, *J. Neurosci.*, *18*, 10090–100104.

Ritov, Y., & Bickel, P. J. (1990). Achieving information bounds in non and semiparametric models. *Ann. Statist.*, *18*, 925–938.

Sakai, Y., Funahashi, S., & Shinomoto, S. (1999). Temporally correlated inputs to leaky integrate-and-fire models can reproduce spiking statistics of cortical neurons. *Neural Netw.*, *12*, 1181–1190.

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

Shinomoto, S., Miyazaki, Y., Tamura, H., & Fujita, I. (2005). Regional and laminar differences in in vivo firing patterns of primate cortical neurons. *J. Neurophysiol.*, *94*, 567–576.

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

Tuckwell, H. C. (1988). *Introduction to theoretical neurobiology: Vol 2, nonlinear and stochastic theories*. Cambridge: Cambridge University Press.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.

Received August 19, 2005; accepted April 12, 2006.