

Optimizing Time Histograms for Non-Poissonian Spike Trains

Takahiro Omi

omitakahiro@gmail.com

Shigeru Shinomoto

shinomoto@scphys.kyoto-u.ac.jp

Department of Physics, Kyoto University, Kyoto, 606-8502, Japan

The time histogram is a fundamental tool for representing the inhomogeneous density of event occurrences such as neuronal firings. The shape of a histogram critically depends on the size of the bins that partition the time axis. In most neurophysiological studies, however, researchers have arbitrarily selected the bin size when analyzing fluctuations in neuronal activity. A rigorous method for selecting the appropriate bin size was recently derived so that the mean integrated squared error between the time histogram and the unknown underlying rate is minimized (Shimazaki & Shinomoto, 2007). This derivation assumes that spikes are independently drawn from a given rate. However, in practice, biological neurons express non-Poissonian features in their firing patterns, such that the spike occurrence depends on the preceding spikes, which inevitably deteriorate the optimization. In this letter, we revise the method for selecting the bin size by considering the possible non-Poissonian features. Improvement in the goodness of fit of the time histogram is assessed and confirmed by numerically simulated non-Poissonian spike trains derived from the given fluctuating rate. For some experimental data, the revised algorithm transforms the shape of the time histogram from the Poissonian optimization method.

1 Introduction ---

A rationale for estimating the neuronal firing rate in physiological studies lies in the presumption that neurons express information via the frequency of spike occurrences, obtained by dividing the number of spikes by the observation period (Adrian, 1928; Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997). To grasp the temporal modulation of the neuronal firing activity, a time histogram needs to be constructed by subdividing the observation period and counting the number of spikes in each bin (Gerstein & Kiang, 1960; Abeles, 1982). However, the shape of a time histogram depends considerably on the choice of the bin size. For instance, if the bin size is quite small, the time histogram fluctuates significantly and the

underlying spike rate cannot be discerned, whereas if the bin size is large, the time-dependent rate cannot be grasped. In neurophysiological literature, most researchers have subjectively selected a bin size that critically determines the goodness of fit of the time histogram to the experimental data.

Theories for optimizing the rate estimation have been suggested on various standards for the goodness of fit of the time histogram, such as the Kullback-Leibler divergence (Hall, 1990), the Hellinger distance (Kanazawa, 1993), the Bayesian principle (Endres, Oram, Schindelin, & Foldiak, 2008; Endres & Oram, 2010), and the mean integrated squared error (MISE) (Révész, 1968; Scott, 1979; Koyama & Shinomoto, 2004). It was recently proved that the expected MISE can be estimated from spike count statistics alone, without knowing the underlying rate process (Rudemo, 1982; Shimazaki and Shinomoto, 2007). A method for minimizing the expected MISE was rigorously derived assuming that the spikes are independently drawn from a given rate. When a peristimulus time histogram (PSTH) is constructed by superimposing numerous spike trains, mixed spikes are statistically independent, and the superimposed sequence can be approximated as Poissonian (Snyder, 1975; Daley & Vere-Jones, 1988; Kass, Ventura, & Brown, 2005).

However, individual spike trains bear non-Poissonian features such that the spike occurrence depends on the preceding spike (Kuffler, Fitzhugh, & Barlow, 1957; Baker & Lemon, 2000; Pillow, Paninski, Uzzell, Simoncelli, & Chichilnisky, 2005; Kostal & Lánský, 2006). In particular, individual neurons express specific firing patterns that may be characterized as regular, random, or bursty, which are reflected in the local variation of interspike intervals (ISIs) (Holt, Softky, Koch, & Douglas, 1996; Shinomoto, Shima, & Tanji, 2003; Shinomoto, Miyazaki, Tamura, & Fujita, 2005; Shinomoto et al., 2009; Miura, Okada, & Amari, 2006; Davies, Gerstein, & Baker, 2006; Nawrot et al., 2008). Thus, the non-Poissonian features should be considered when estimating the firing rate for a single spike train (Cunningham, Yu, Shenoy, & Sahani, 2008; Cunningham, Gilja, Ryu, & Shenoy, 2009; Shimokawa & Shinomoto, 2009). It is also critically important to consider the non-Poissonian features when analyzing the spike correlation across multiple spike trains. It has been shown that a simple Poissonian analysis produces false positives particularly for regular spike trains, which leads to an incorrect conclusion on the presence of spike correlations even between the independent spike trains (Grün, 2009; Louis, Gerstein, Grün, & Diesmann, 2010).

In this letter, we revise the method for selecting bin size by considering non-Poissonian features. By using simulation data, we evaluate the degree to which the revised method improves the goodness of fit according to the MISE. Furthermore, we apply the method to *in vivo* spike data to examine the manner in which the shape of a time histogram is transformed by the replacement of optimization algorithms.

2 Derivation of Bin Size Optimization Method

A time histogram is readily constructed by partitioning an observation period T into intervals of width Δ , counting the number of spikes k_i that fall into each (i th) bin, and drawing a bar at height k_i/Δ for $i = 1, 2, \dots, N$.

The optimization method aims to find the bin size Δ that minimizes the MISE between the time histogram, $\hat{\lambda}_t$, and the underlying rate, λ_t . Assuming that the spikes are sampled from a stochastic process, we use the expected MISE defined by the following formula,

$$\text{MISE} \equiv \frac{1}{T} \int_0^T E (\hat{\lambda}_t - \lambda_t)^2 dt, \quad (2.1)$$

where E refers to the expectation over different realizations of spikes under a given λ_t . Minimizing the MISE is equivalent to minimizing the cost function constructed by subtracting the optimization-free term, as given by

$$C(\Delta) \equiv \text{MISE} - \overline{(\lambda_t - \bar{\lambda}_t)^2}, \quad (2.2)$$

where $\bar{A} = \frac{1}{T} \int_0^T A dt$ represents the time average.

By using bias-variance decomposition (see appendix A) Shimazaki and Shinomoto (2007) transformed the cost function into

$$C(\Delta) = 2\langle E(\hat{\theta}_i - \theta_i)^2 \rangle - \langle E(\hat{\theta}_i - \langle E\hat{\theta}_i \rangle)^2 \rangle, \quad (2.3)$$

where the brackets denote the average over all bins $\langle A_i \rangle \equiv \frac{1}{N} \sum_{i=1}^N A_i$, $\hat{\theta}_i$ is the height of the i th histogram bin given by k_i/Δ , and θ_i is the expected height of the histogram bin given by the unknown underlying rate λ_t ,

$$\theta_i \equiv \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} \lambda_t dt. \quad (2.4)$$

2.1 Poissonian Optimization Method. If individual spikes are drawn independently from an underlying rate, the number of spikes k_i counted in each (i th) bin obeys Poisson distribution, satisfying the condition that the spike count variance is equal to the mean, $E(k_i - E(k_i))^2 = Ek_i$. Because the height of the histogram bin is given by $\hat{\theta}_i = k_i/\Delta$, this variance-mean relationship implies

$$E(\hat{\theta}_i - \theta_i)^2 = \frac{1}{\Delta} E\hat{\theta}_i. \quad (2.5)$$

By inserting equation 2.5 into equation 2.3, the cost function can be represented as a function of the histogram estimator $\hat{\theta}_i$,

$$C(\Delta) = \frac{2}{\Delta} \langle E\hat{\theta}_i \rangle - \langle E(\hat{\theta}_i - \langle E\hat{\theta}_i \rangle)^2 \rangle. \quad (2.6)$$

By recovering the original terms for the spike count, the cost function is given by the simple formula

$$\hat{C}(\Delta) = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{2k_i - (k_i - \bar{k})^2}{\Delta^2} \right\}, \quad (2.7)$$

where $\bar{k} = \frac{1}{N} \sum_{i=1}^N k_i$ is the mean spike count. The optimal bin size is obtained by minimizing the cost function $\hat{C}(\Delta)$.

2.2 Non-Poissonian Optimization Method. We now revise the optimization algorithm so that it applies to the spike trains that bear non-Poissonian features for which the variance-mean relationship (see equation 2.5) is not valid. In general, the ratio of the variance to the mean of the event count is called the Fano factor F (Fano, 1947). We suggest revising equation 2.5 by using the Fano factor F_i in each bin as follows:

$$E(\hat{\theta}_i - \theta_i)^2 = F_i \frac{1}{\Delta} E\hat{\theta}_i. \quad (2.8)$$

With this new variance-mean relationship, the cost function is simply revised as

$$\hat{C}(\Delta) = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{2\hat{F}_i k_i - (k_i - \bar{k})^2}{\Delta^2} \right\}, \quad (2.9)$$

where \hat{F}_i is an estimator of the Fano factor. To estimate the Fano factor from a single spike train in each bin, we relate the Fano factor to the ISI variability by using the approximation $F \approx C_V^2$ (Cox, 1962; appendix B); where C_V is the coefficient of variation, which is defined as the standard deviation of the ISIs divided by the mean, $C_V \equiv \Delta\tau/\bar{\tau}$. In order to obtain a concrete estimation, we suggest two alternative algorithms based on this relationship.

Algorithm 1: First, we suggest a direct estimation of C_V in each bin according to the original definition. If a bin contains fewer than three spikes, then the ISI variability is not measurable, and we therefore suggest setting $\hat{F} = 1$, because the Fano factor approaches unity for an interval that contains few spikes (see appendix B). Taken together, we propose an algorithm denoted as algorithm 1 for estimating the Fano factor in each bin of a single

spike train:

$$\hat{F}_i \equiv \begin{cases} 1, & \text{if } k_i \leq 2, \\ C_V^2 = \left(\frac{\Delta\tau}{\bar{\tau}} \right)^2, & \text{otherwise} \end{cases}. \quad (2.10)$$

Algorithm 2: The coefficient of variation is known to be easily disturbed by rate fluctuation. The influence of rate fluctuation on the estimation of intrinsic firing irregularity can be eliminated by rescaling the time axis or by using the local variation L_V (Shinomoto et al., 2003, 2009; appendix C); that can be computed as $L_V = \frac{3}{k_i-2} \sum_{j=1}^{k_i-2} \left(\frac{\tau_j - \tau_{j+1}}{\tau_j + \tau_{j+1}} \right)^2$ using the ISIs $\{\tau_j\}$ that fall into each bin ($j = 1, 2, \dots, k_i - 1$). Here we relate the Fano factor F to L_V using the conversion relation, $C_V^2 = 2L_V / (3 - L_V)$, which is derived by assuming gamma processes (appendix D). Expanding upon algorithm 1, we propose an alternative algorithm denoted as algorithm 2:

$$\hat{F}_i \equiv \begin{cases} 1, & \text{if } k_i \leq 2, \\ \frac{2L_V}{3 - L_V}, & \text{otherwise} \end{cases}. \quad (2.11)$$

In the next section, we confirm by numerical simulation that algorithm 2 outperforms algorithm 1. Thus, we adopt algorithm 2 as the representative method; the computations are summarized by the following steps:

Computational Steps for Selecting Bin Size for a Non-Poissonian Spike Train.

1. Divide the observation period T into N bins having width Δ , and count the number of spikes k_i that enter the i th bin.
2. Estimate the Fano factor for each bin,

$$\hat{F}_i \equiv \begin{cases} 1, & \text{if } k_i \leq 2, \\ \frac{2L_V}{3 - L_V}, & \text{otherwise,} \end{cases}$$

where $L_V \equiv \frac{3}{k_i-2} \sum_{j=1}^{k_i-2} \left(\frac{\tau_j - \tau_{j+1}}{\tau_j + \tau_{j+1}} \right)^2$ is computed from the ISIs $\{\tau_j\}$ that fall into the bin.

3. Compute the average of $\{\hat{F}_i k_i\}$ and the variance of $\{k_i\}$ as

$$h \equiv \frac{1}{N} \sum_{i=1}^N \hat{F}_i k_i, \text{ and } v \equiv \frac{1}{N} \sum_{i=1}^N (k_i - \bar{k})^2,$$

where $\bar{k} \equiv \frac{1}{N} \sum_{i=1}^N k_i$.

4. Construct the cost function,

$$\hat{C}(\Delta) = \frac{2h - v}{\Delta^2}.$$

5. Repeat steps 1 through 4 while changing the bin size Δ to search for the optimal bin size that minimizes $\hat{C}(\Delta)$.

Note that the Poissonian optimization method is obtained by simply replacing step 2 with $\hat{F} = 1$. Also, computational complexity may be reduced drastically by skipping the individual estimation of L_V in each bin and replacing the value with the one computed for the entire spike train.

3 Performance of the Optimization Algorithms

3.1 Application to Simulation Data. We compare the performances of the original and revised algorithms by applying them to non-Poissonian spike sequences. For this purpose, we generate the spike trains by numerically simulating inhomogeneous gamma processes with a given rate function λ_t ($0 < t < T$) via inverse rescaling in the following steps. First, the ISIs $\{x_1, x_2, \dots, x_n\}$ are drawn independently from the gamma distribution function (Cox & Lewis, 1966),

$$f_\kappa(x) = \kappa(\kappa x)^{\kappa-1} e^{-\kappa x} / \Gamma(\kappa), \quad (3.1)$$

where $\Gamma(\kappa) \equiv \int_0^\infty x^{\kappa-1} e^{-x} dx$ is the gamma function. Here the non-Poissonian feature can be specified by the shape factor κ that determines the ISI variability. The number of ISIs n is determined as the smallest integer satisfying $\sum_{i=1}^{n+1} x_i > \int_0^T \lambda_t dt$. Second, the ISIs $\{x_1, x_2, \dots, x_n\}$ are consecutively arranged on the time axis to construct a non-Poissonian spike train of a uniform rate; the i th spike time is given by summing the previous ISIs as $y_i = \sum_{j=1}^i x_j$. Third, the time axis is rescaled with a given time-dependent rate (Berman, 1983; Ogata, 1988; Reich, Victor, & Knight, 1998; Oram, Wiener, Lestienne, & Richmond, 1999; Brown, Barbieri, Ventura, Kass, & Frank, 2002; Koyama & Shinomoto, 2005) such that the original spike times $\{y_1, y_2, \dots, y_n\}$ are transformed into rescaled times by $t_i \equiv \Lambda^{-1}(y_i)$, where $\Lambda^{-1}(y)$ is the inverse of the function $\Lambda(t) \equiv \int_0^t \lambda_s ds$. By using this method, we can parametrically control the non-Poissonian spiking features without impairing the given rate process. By using the simulated spike trains derived from the given rate λ_t , the performance of any rate estimation algorithm can be evaluated by computing the integrated squared error (ISE) between the underlying rate λ_t and the time histogram $\hat{\lambda}_t$ constructed from the spike train, as defined by $\text{ISE} \equiv \int_0^T (\hat{\lambda}_t - \lambda_t)^2 dt$.

Figure 1 exemplifies three types of spike trains derived from an identical sinusoidally modulated rate; they bear different non-Poissonian

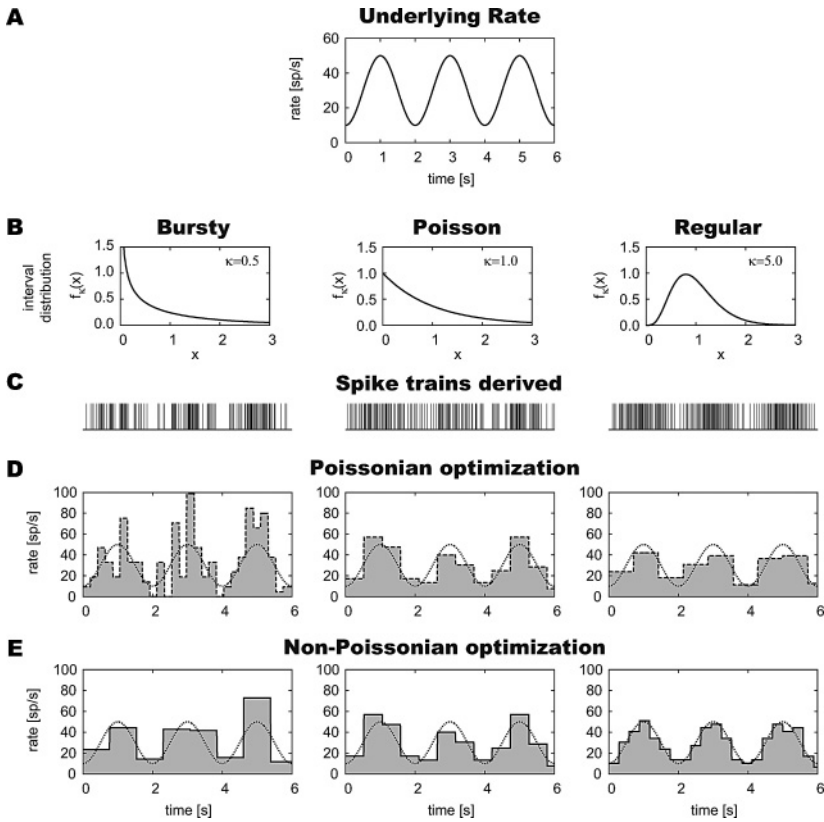


Figure 1: Time histograms determined by Poissonian and non-Poissonian optimization methods. (A) Underlying spike rate, λ_t . (B) Three ISI distribution functions. We selected the gamma distribution functions of the shape factors $\kappa = 0.5, 1$, and 5 , which generate bursty, Poissonian random, and regular spike trains, respectively. (C) Spike trains derived from an identical sinusoidal rate process using different ISI distribution functions. (D) Time histograms constructed using the bin size determined by the Poissonian optimization method (Shimazaki & Shinomoto, 2007). (E) Non-Poissonian optimization method (algorithm 2).

features that may be called bursty, Poissonian random, or regular according to the shape parameter of the gamma distribution function, $\kappa < 1$, $\kappa = 1$, and $\kappa > 1$, respectively. For respective spike trains, the time histograms optimized by the present algorithm (algorithm 2) are compared with those constructed by the original Poissonian method (Shimazaki & Shinomoto, 2007), illustrating the improved performance achieved by the revision.

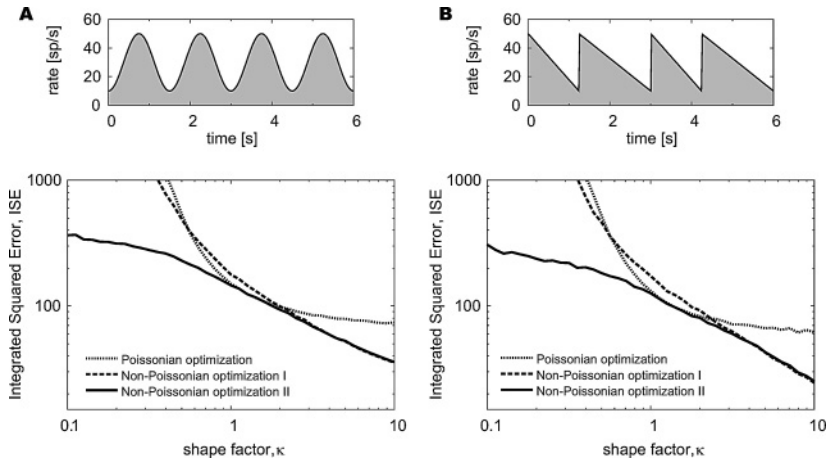


Figure 2: Goodness-of-fit performance of the Poissonian and non-Poissonian optimization method. Integrated squared error (ISE) between the underlying rate and the time histograms constructed for the spike trains characterized with a shape factor κ of the ISI distribution. The dotted, dashed, and solid lines represent the ISEs for the Poissonian (Shimazaki & Shinomoto, 2007) and algorithms 1 and 2 of the non-Poissonian optimization method, respectively. $\kappa < 1$, $\kappa = 1$, and $\kappa > 1$ correspond to bursty, Poissonian random, and regular, respectively. (A) Sinusoidally regulated rate process. (B) Sawtooth rate fluctuation whose teeth are mutually incommensurate $1 : \sqrt{2}$.

Note that the optimized bin size shows the opposite behavior between the Poissonian and non-Poissonian optimization methods. For the regular spike train, bin size obtained by the non-Poissonian optimization method is smaller than by the Poissonian optimization method. This is because the metric of the local variation L_V detected the regular noiseless firing, and, accordingly, the optimization method could obtain the reliable estimate using fewer spikes. For the bursty spike train, the firing irregularity is higher than the Poisson randomness, and, accordingly, the optimization method required more spikes for the reliable rate estimation. In contrast, the Poissonian optimization method chose the smaller bin size for the bursty spike train. This is because the Poissonian optimization method interpreted the bursty firing as a signal rather than noise.

Figure 2 compares the performances of two variant algorithms of the revised method and the original Poissonian algorithm represented by a small ISE between histograms and the given underlying rate. Algorithm 1 improves the performance over the Poissonian algorithm for the non-Poissonian spike trains whose shape factor κ deviates considerably from 1 (Poissonian); however, the performance is slightly inferior in the near-Poissonian range. In contrast, algorithm 2 always performs better than

both the Poissonian algorithm and algorithm 1. Therefore, we adopt algorithm 2 as an appropriate algorithm for the non-Poissonian optimization method.

Here we tested two types of rate fluctuation: sinusoidally regulated and sawtooth rate fluctuations. Both cases show that the ISE decreases with κ , reflecting the tendency of the spike signals to become less noisy with an increase of the shape factor. In both cases of any regularity κ , algorithm 2 is advantageous over other algorithms.

3.2 Application to Biological Data. Next, we apply the optimization methods to actual biological data to examine the manner in which the shape of the time histogram is transformed by replacing the optimization methods. Here, we employ the commonly available spike trains of a neuron in area MT of a monkey responding to a random dot stimulus (Britten, Shadlen, Newsome, & Movshon, 1992, 2004). Figure 3 depicts sample spike sequences and compares the time histograms constructed using the original Poissonian method and the non-Poissonian method for each spike train. It is observed that for some spike trains, the revised optimization method considerably transforms the shape of the time histogram.

In the section 3.1, we tested the effectiveness of the optimization methods using numerically simulated data by directly comparing the resulting histogram to a given underlying rate. However, with regard to the experimental data, the underlying rate is unknown. Thus, we attempt here to verify the suitability of the optimization methods by performing a likelihood cross-validation, although it is more indirect than numerical simulation. Specifically, from a time histogram constructed from a single spike train, we estimate the average likelihood of other spike trains. Among the 15 spike trains from repetitive presentation of an identical stimulus (see Figure 3A), seven of these samples resulted in an estimate of the same bin size for Poissonian and non-Poissonian optimization methods. Among the remaining eight spike trains for which the two methods gave different time histograms, six samples gave a greater value for the cross-validated likelihood for the non-Poissonian method, while two samples gave a smaller value. This result implies that in our spike data, the non-Poissonian method is superior with respect to the likelihood of cross-validation criteria. In addition, we compute the ISE between the individual optimized histogram and the average rate of spike occurrences of the 15 spike trains. By using the optimal kernel method proposed by Shimazaki and Shinomoto (2010), we estimated the average rate of spike occurrences (see Figure 3B, also depicted as the dashed line in Figures 3D and 3E). Among the eight spike trains, seven samples gave a smaller ISE for the non-Poissonian method, and one sample gave a larger ISE. This also implies the superiority of the non-Poissonian method. However, it should be noted that the likelihood cross-validation analysis or the ISE analysis is based on the assumption that the spike train is repeatedly derived from the same underlying rate.

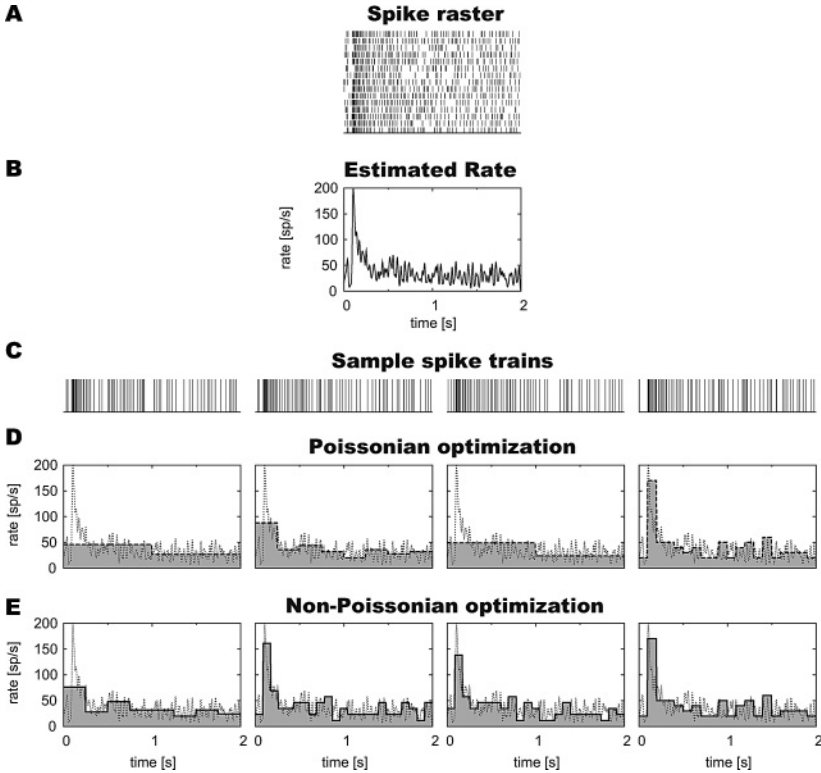


Figure 3: Application of Poissonian and non-Poissonian optimization methods to the spike data of an MT neuron (#j015 in Britten et al., 2004). (A) Spike trains obtained by the repeated presentation of a visual stimulus. (B) Average frequency of spikes estimated with the kernel density estimate (Shimazaki & Shinomoto, 2010, fixed bandwidth optimization). (C) Single spike trains sampled from the ensemble. (D) Time histograms constructed with the bin size determined by the Poissonian optimization method (Shimazaki & Shinomoto, 2007). (E) Non-Poissonian optimization method (algorithm 2).

4 Discussion

We revised the method of selecting the bin size of the time histogram so that it applies to the non-Poissonian spike trains. The revision replaces the Poissonian variance-mean relationship with the general non-Poissonian relationship given by the Fano factor F . To make the revision practicable, we proposed an estimation of the Fano factor in each bin from the ISI variability of a single spike train. We suggested two algorithms for the non-Poissonian optimization method and tested them for their goodness-of-fit

performance by using the spike trains numerically derived from an inhomogeneous gamma process. Algorithm 1, which employs the raw coefficient of variation C_V , generally improves the goodness of fit for the non-Poissonian data; however, its performance is slightly inferior to that of the original Poissonian optimization method for the near-Poissonian spike trains. This is presumably because C_V fluctuates considerably under the influence of rate fluctuation and thus degrades the optimization. To improve this imperfection, we suggested an alternative algorithm, algorithm 2, which employs the local variation metric L_V that robustly detects the non-Poissonian feature despite the rate fluctuation. Algorithm 2 exhibits goodness-of-fit superior to both algorithm 1 and the original Poissonian optimization method throughout the entire range of the shape parameter of the gamma ISI distribution.

In theoretical neuroscience, neuronal firing has also been approximated by the multiplicative intensity model in which neuronal refractoriness is considered by introducing a hazard function (Aalen, 1978). This effect is similar to considering a gamma process of a regular range, $\kappa > 1$. By using spike trains numerically generated by the multiplicative intensity model, we also confirmed the validity of algorithm 2 (data not shown). Furthermore, we applied the non-Poissonian optimization method to actual biological data and demonstrated that the revised algorithm transforms the shape of the time histogram from the Poissonian optimization method.

However, it should be noted that the revised method does not cover all types of non-Poissonian firing. The extension of the Poissonian optimization method to the non-Poissonian range is based on the assumption of a renewal process in which the ISIs are drawn independently from an identical distribution function, although its timescale may be modulated slowly in time. Thus, the method is not valid for spike trains with a strong serial correlation of arbitrary order.

Rate estimation inevitably depends on the method and the principle it is based on. Therefore, it is not uniquely determined for a set of data (Bowman, 1985). Nevertheless, the range of the plausible principles and the estimation methods is limited, and the rate estimated from a given set of data should not vary among the principled methods. For instance, it was shown that the bin size that minimizes the Hellinger distance is asymptotically equal to the bin size that minimizes the Kullback-Leibler distance and also to the bin size selected according to the Akaike information criterion, in the limit of large sample sizes (Kanazawa, 1993). We also argue that the MISE minimization for a PSTH and the marginal likelihood maximization for the Bayes estimation both render the same degree of detectability for the presence of rate modulation (Shinomoto, 2010). Thus, the discussion based on the MISE principle is not entirely exceptional and may share a common trait with various reasonable principles.

It has been reported recently that by analyzing biological data, researchers decoded animal behavior more efficiently via the correlation code, which lies in the serial correlations within a spike train, rather than by the standard “rate code” analysis (Jacobs et al., 2009). In this letter, we have demonstrated that the firing rate estimation can be improved by considering non-Poissonian firing characteristics. These two conclusions are not necessarily conflicting but may represent two sides of the same coin. The “correlation code” may indicate the specific interspike correlation reflecting the non-Poissonian firing characteristics, and in both cases, the endeavor to capture the non-Poissonian characteristics adds information to the simple rate decoded under the Poissonian assumption. The similarities and differences between those approaches are worth investigating.

Appendix A: Derivation of Cost Function

Here, we derive a formula for the cost function, equation 2.3, by tracing the derivation in Shimazaki and Shinomoto (2007).

By splitting the observation period into bins of size Δ , the MISE can be divided into parts,

$$\text{MISE} \equiv \frac{1}{T} \int_0^T E(\hat{\lambda}_t - \lambda_t)^2 dt = \frac{1}{N} \sum_{i=1}^N \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} E(\hat{\theta}_i - \lambda_t)^2 dt, \quad (\text{A.1})$$

where $\hat{\theta}_i \equiv k_i/\Delta$ is the height of the histogram at the i th bin. The summand on the right-hand side can be decomposed into two parts,

$$\frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} E(\hat{\theta}_i - \theta_i + \theta_i - \lambda_t)^2 dt = E(\hat{\theta}_i - \theta_i)^2 + \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} (\lambda_t - \theta_i)^2 dt, \quad (\text{A.2})$$

where

$$\theta_i \equiv \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} \lambda_t dt = E(\hat{\theta}_i). \quad (\text{A.3})$$

The second term of equation A.2 can be further decomposed into

$$\frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} (\lambda_t - \langle \theta_i \rangle + \langle \theta_i \rangle - \theta_i)^2 dt = \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} (\lambda_t - \langle \theta_i \rangle)^2 dt - (\theta_i - \langle \theta_i \rangle)^2, \quad (\text{A.4})$$

where brackets represent the average over all bins, $\langle A_i \rangle \equiv \frac{1}{N} \sum_{i=1}^N A_i$, and, accordingly,

$$\langle \theta_i \rangle \equiv \frac{1}{N} \sum_{i=1}^N \theta_i = \frac{1}{N} \sum_{i=1}^N \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} \lambda_t = \bar{\lambda}_t. \quad (\text{A.5})$$

By incorporating equations A.2 and A.4 in equation A.1, we obtain

$$\text{MISE} = \langle E(\hat{\theta}_i - \theta_i)^2 \rangle - \langle (\theta_i - \langle \theta_i \rangle)^2 \rangle + \overline{(\lambda_t - \bar{\lambda}_t)^2}. \quad (\text{A.6})$$

By removing the last term that is independent of the bin size, the cost function defined in equation 2.2 is obtained as

$$\begin{aligned} C(\Delta) &\equiv \text{MISE} - \overline{(\lambda_t - \bar{\lambda}_t)^2} \\ &= \langle E(\hat{\theta}_i - \theta_i)^2 \rangle - \langle (\theta_i - \langle \theta_i \rangle)^2 \rangle. \end{aligned} \quad (\text{A.7})$$

Because θ_i is given by the unobservable underlying rate λ_t , the second term should be rewritten in terms of the observable estimator $\hat{\theta}_i$. For this purpose, bias-variance decomposition is employed:

$$E(\hat{\theta}_i - \theta_i)^2 = (\theta_i - \langle \theta_i \rangle)^2 + E(\hat{\theta}_i - \theta_i)^2, \quad (\text{A.8})$$

where the first and second terms on the right-hand side, respectively, represent the (bias)² and variance. By replacing $\langle \theta_i \rangle$ on the left-hand side with $\langle E\hat{\theta}_i \rangle$ and plugging this relationship into equation A.7, the cost function is transformed into equation 2.3:

$$C(\Delta) = 2\langle E(\hat{\theta}_i - \theta_i)^2 \rangle - \langle E(\hat{\theta}_i - \langle E\hat{\theta}_i \rangle)^2 \rangle. \quad (\text{A.9})$$

The variance $E(\hat{\theta}_i - \theta_i)^2$ in the first term is replaced with the observable mean $E(\hat{\theta})$ through the variance-mean relationship given by equation 2.5 or 2.8.

Appendix B: Estimating the Fano Factor from a Single Spike Train

The Fano factor measures the variability in spike count using the ratio of the variance to the mean for the number of spikes that fall into an interval. This original definition requires numerous spike trains, but it can be estimated from a single spike train in two cases.

For a long sequence of spikes derived from a stationary renewal process, the Fano factor F is approximated by the square of the coefficient of variation C_V of the ISIs (Cox, 1962):

$$\lim_{D \rightarrow \infty} F = C_V^2 = \left(\frac{\Delta\tau}{\bar{\tau}} \right)^2, \quad (\text{B.1})$$

where D is the duration of observation and $\Delta\tau$ and $\bar{\tau}$ are the standard deviation and the mean of the ISIs, respectively. The relationship can be proven as follows. For a stationary renewal process in which the ISIs $\{\tau_i\}$ are drawn independently from an identical distribution function, the probability that the number of spikes in the interval $N(D)$ is smaller than m is identical to the probability that the sum of m ISIs is longer than D :

$$\text{Prob}(N(D) < m) = \text{Prob}(\tau_1 + \tau_2 + \cdots + \tau_m > D). \quad (\text{B.2})$$

The central limit theorem holds in the limit of large m ; the sum of the ISIs is approximately normally distributed with the mean $m\bar{\tau}$ and the variance $m(\Delta\tau)^2$. This is represented as

$$\text{Prob}(\tau_1 + \tau_2 + \cdots + \tau_m > D) = \frac{1}{\sqrt{2\pi}} \int_{\frac{D-m\bar{\tau}}{\sqrt{m(\Delta\tau)^2}}}^{\infty} dx \exp\left(-\frac{x^2}{2}\right). \quad (\text{B.3})$$

By combining equations B.3 and B.2 and introducing a new variable, $y = (D - m\bar{\tau})/\sqrt{D}$, we obtain

$$\text{Prob}(N(D) < D/\bar{\tau} - y\sqrt{D}/\bar{\tau}) = \frac{1}{\sqrt{2\pi}} \int_{\frac{y}{\Delta\tau} \sqrt{\frac{\bar{\tau}}{1-y/\sqrt{D}}}}^{\infty} dx \exp\left(-\frac{x^2}{2}\right). \quad (\text{B.4})$$

In the limit of the long duration D , the lower bound of the integral becomes $y\sqrt{\bar{\tau}}/\Delta\tau$. Thus, $N(D)$ is asymptotically normally distributed with the mean $D/\bar{\tau}$ and the variance $D(\Delta\tau)^2/\bar{\tau}^3$. Therefore, the Fano factor is approximated as $F \approx (\Delta\tau)^2/\bar{\tau}^2$, leading to equation B.1.

In the opposite limit in which the interval D is narrow, few spikes are counted by chance with a small probability. This process can be approximated as Poissonian, in which the Fano factor is unity:

$$\lim_{D \rightarrow 0} F = 1. \quad (\text{B.5})$$

The transition of the Fano factor from 1 to C_V^2 with the observation interval D is obtained by performing numerical simulation in a manner similar to Nawrot (2010): the Fano factor is computed for an ensemble of

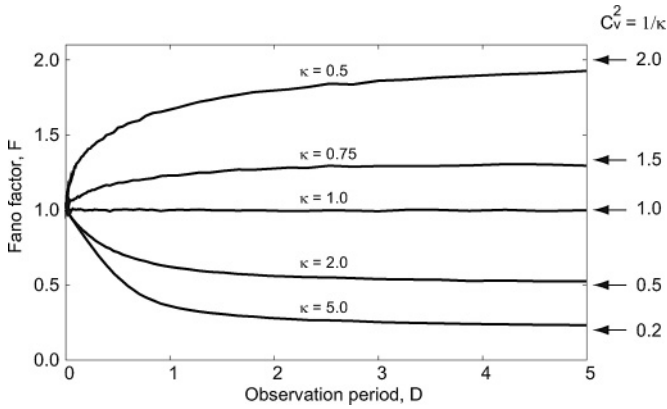


Figure 4: Dependence of the Fano factor on the observation period D . The Fano factor is computed from 10^5 spike trains derived from a uniform gamma process with the mean rate unity. The asymptotic values $C_V^2 = 1/\kappa$ for the large limit of D for various processes are indicated by arrows.

spike trains derived from a uniform gamma process with a mean rate unity. Figure 4 shows that the Fano factor is close to unity for a short interval that contains few spikes, and it swiftly approaches the asymptotic value of $C_V^2 = 1/\kappa$ in an interval typically containing a few spikes. This result may support our step for using $F = 1$ when there are fewer than three spikes.

Appendix C: Relating the Fano Factor to ISI Variability in the Presence of Rate Fluctuation

The method of estimating the Fano factor from the ISI variability of a single spike train is derived in appendix B assuming that the spikes were drawn from a stationary renewal process. Here we extend the method to nonstationary processes in which the rate fluctuates in time.

If the inhomogeneous spike rate $\lambda(t)$ is known for a given spike train, then we can convert it to a uniform spike train by transforming the original spike times $\{t_1, \dots, t_k\}$ to another sequence $\{s_1, \dots, s_k\}$ by $s_i = \Lambda(t_i)$, defined by

$$\Lambda(t) = \int_0^t \lambda(t') dt'. \quad (\text{C.1})$$

If time rescaling recovers stationarity without breaking the renewal characteristics, the theory relating the Fano factor to the ISI variability applies to

the new sequence $\{s_1, \dots, s_k\}$, because any time rescaling does not change the number of spike count.

Several methods for performing a suitable time rescaling have been proposed (Reich et al., 1998; Oram et al., 1999; Barbieri, Quirk, Frank, Wilson, & Brown, 2001; Smith & Brown, 2003; Koyama & Shinomoto, 2005; Shinomoto & Koyama, 2007; Shimokawa & Shinomoto, 2009; Shimokawa, Koyama, & Shinomoto, 2010; Nawrot, 2010), but they generally require complicated algorithms, and therefore high computational costs. It should be noted that a simple metric of the local variation L_V spontaneously realizes time rescaling in the estimation of the variability of consecutive intervals (Shinomoto et al., 2003, 2009; Zhao, Omi, Matsuno, & Shinomoto, 2010): In the definition of $L_V = \frac{3}{k-2} \sum_{j=1}^{k-2} \left(\frac{\tau_j - \tau_{j+1}}{\tau_j + \tau_{j+1}} \right)^2$, the summand computes the cross-correlation of the consecutive ISIs τ_i and τ_{i+1} , each rescaled with the instantaneous spike rate $(\tau_j + \tau_{j+1})/2$, as represented by

$$\left(\frac{\tau_j - \tau_{j+1}}{\tau_j + \tau_{j+1}} \right)^2 = 1 - \left(\frac{2\tau_j}{\tau_j + \tau_{j+1}} \right) \left(\frac{2\tau_{j+1}}{\tau_j + \tau_{j+1}} \right). \quad (\text{C.2})$$

Owing to this time-rescaling mechanism, L_V robustly detects the intrinsic firing characteristic against the rate fluctuation.

Appendix D: Relating L_V to C_V

Although different variability metrics are not uniquely related, we can obtain a practical conversion formula by allowing them to measure the same set of spike trains. For uniform gamma processes in which the ISIs are drawn from the gamma distribution functions of the shape factor κ (equation 3.1), we can compute the expectation value of the local variation L_V ,

$$E[L_V] = \frac{3}{\Gamma(\kappa)^2} \int_0^\infty dx_1 \int_0^\infty dx_2 (x_1 x_2)^{\kappa-1} e^{-x_1-x_2} \left(\frac{x_1 - x_2}{x_1 + x_2} \right)^2 \quad (\text{D.1})$$

$$= \frac{3}{2^{2\kappa-1} \Gamma(\kappa)^2} \int_0^\infty dt \int_{-t}^t ds (t^2 - s^2)^{\kappa-1} e^{-t} \left(\frac{s}{t} \right)^2 \quad (\text{D.2})$$

$$= \frac{3}{2\kappa + 1}. \quad (\text{D.3})$$

The expectation value of C_V can also be obtained by computing the mean $\bar{\tau}$ and the standard deviation $\Delta\tau$ for the gamma distribution,

$$E[C_V] = \frac{1}{\sqrt{\kappa}}. \quad (\text{D.4})$$

Thus, the two metrics are parametrically related through the gamma distribution functions of various shape parameters as

$$C_V^2 = \frac{2L_V}{3 - L_V}. \quad (\text{D.5})$$

Acknowledgments

This study was supported in part by Grant-in-Aid for Scientific Research to S.S. from the MEXT Japan (20300083) and the Grant-in-Aid for the Global COE Program, The Next Generation of Physics, Spun from Universality and Emergence, from the MEXT of Japan. T.O. is supported by JSPS Research Fellowships for Young Scientists. We greatly appreciate that K. H. Britten, M. N. Shadlen, W. T. Newsome, and J. A. Movshon have made their data publicly available.

References

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics*, 6, 701–726.
- Abeles, M. (1982). Quantification, smoothing, and confidence limits for single-units' histograms. *Journal of Neuroscience Methods*, 5, 317–325.
- Adrian, E. D. (1928). *The basis of sensation: The action of the sense organs*. New York: Norton.
- Baker, S. N., & Lemon, R. N. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *Journal of Neurophysiology*, 84, 1770–1780.
- 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.
- Berman, M. (1983). Comment on "Likelihood analysis of point processes and its applications to seismological data" by Ogata. *Bulletin Internatl. Stat. Instit.*, 50, 412–418.
- Bowman, A. W. (1985). A comparative study of some kernel-based nonparametric density estimators. *Journal of Statistical Computation and Simulation*, 21, 313–327.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12, 4745–4765.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (2004). Responses of single neurons in macaque MT/V5 as a function of motion coherence in stochastic dot stimuli. *Neural Signal Archive*, NSA2004.1
- 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 Computation*, 14, 325–346.
- Cox, D. R. (1962). *Renewal theory*. London: Chapman and Hall.

- Cox, D. R., & Lewis, P.A.W. (1966). *Statistical analysis of series of events*. 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. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*, 20. Cambridge, MA: MIT Press.
- Daley, D. J., & 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.
- Endres, D., & Oram, M. (2010). Feature extraction from spike trains with Bayesian binning: “Latency is where the signal starts.” *Journal of Computational Neuroscience*, 29, 149–169.
- Endres, D., Oram, M., Schindelin, J., & Foldiak, P. (2008). Bayesian binning beats approximate alternatives: Estimating peristimulus time histograms. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*, 20. Cambridge, MA: MIT Press.
- Fano, U. (1947). Ionization yield of radiations, II. The fluctuations of the number of ions. *Physical Review*, 72, 26–29.
- Gerstein, G. L., & Kiang, N.Y.S. (1960). An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophysical Journal*, 1, 15–28.
- Grün, S. (2009). Data-driven significance estimation for precise spike correlation. *Journal of Neurophysiology*, 101, 1126–1140.
- Hall, P. (1990). Akaike’s information criterion and Kullback-Leibler loss for histogram density estimation. *Probability Theory and Related Fields*, 85, 449–467.
- 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. *Journal of Neurophysiology*, 75, 1806–1814.
- 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 of the United States of America*, 106, 5936–5941.
- Kass, R. E., Ventura, V., & Brown, E. N. (2005). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*, 94, 8–25.
- Kanazawa, Y. (1993). Hellinger distance and Akaike’s information criterion for the histogram. *Statistics and Probability Letters*, 17, 293–298.
- Kostal, L., & Lánský, P. (2006). Classification of stationary neuronal activity according to its information rate. *Network: Computation in Neural Systems*, 17, 193–210.
- 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.

- Kuffler, S. W., Fitzhugh, R., & Barlow, H. B. (1957). Maintained activity in the cat's retina in light and darkness. *Journal of General Physiology*, 40, 683–702.
- Louis, S., Gerstein, G. L., Grün, S., & Diesmann, M. (2010). Surrogate spike train generation through dithering in operational time. *Frontiers in Computational Neuroscience*, 4, 127.
- Miura, K., Okada, M., & Amari, S. (2006). Estimating spiking irregularities under changing environments. *Neural Computation*, 18, 2359–2386.
- Nawrot, M. P. (2010). Analysis and interpretation of interval and count variability in neural spike trains. In S. Grün & S. Rotter (Eds.), *Analysis of parallel spike trains*. New York: Springer.
- 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. *Journal of American Statistical Association*, 83, 9–27.
- 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.
- 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*, 25, 11003–11013.
- 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.
- Révész, P. (1968). *The laws of large numbers*. Orlando, FL: Academic Press.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9, 65–78.
- Scott, D. W. (1979). Optimal and data-based histograms. *Biometrika*, 66, 605–610.
- 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.
- Shimokawa, T., Koyama, S., & Shinomoto, S. (2010). A characterization of the time-rescaled gamma process as a model for spike trains. *Journal of Computational Neuroscience*, 29, 183–191.
- Shinomoto, S. (2010). *Estimating the firing rate*. In S. Grün & S. Rotter (Eds.), *Analysis of parallel spike trains*. New York: Springer.
- 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., & Koyama, S. (2007). A solution to the controversy between rate and temporal coding. *Statistics in Medicine*, 26, 4032–4038.

- Shinomoto, S., Miyazaki, Y., & Tamura, H., & Fujita, I. (2005). Regional and laminar differences in vivo firing patterns of primate cortical neurons. *Journal of Neurophysiology*, 94, 567–575.
- Shinomoto, S., Shima, K., & Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural Computation*, 15, 2823–2842.
- Smith, A. C., & Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Computation*, 15, 965–991.
- Snyder, D. L. (1975). *Random point processes*. Hoboken, NJ: Wiley.
- Zhao, X., Omi, T., Matsuno, N., & Shinomoto, S. (2010). A non-universal aspect in the temporal occurrence of earthquakes. *New Journal of Physics*, 12, 063010.

Received December 16, 2010; accepted June 13, 2011.