A Recipe for Optimizing Time Histograms for non-Poissonian Spike Trains

Takahiro Omi (PY), and Shigeru Shinomoto
Department of Physics, Kyoto University
E-mail: omitakahiro@gmail.com

Abstract—We propose a method for selecting the appropriate bin size of a time histogram for a non-Poissonian spike train, so that the mean integrated squared error between the time histogram and the unknown underlying rate is minimized. Effectiveness of the method is demonstrated via numerically simulated non-Poissonian spike trains derived from the given fluctuating rate.

Keywords—Time histogram, Non-Poissonian feature

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. To grasp the temporal modulation of the neuronal firing activity, a time histogram is constructed by subdividing the observation period and counting the number of spikes in each bin. However, the shape of a time histogram depends considerably on the choice of the bin size. For instance, if the bin size is considerably small, the time histogram fluctuates significantly and the underlying spike rate cannot be discerned; whereas if the bin size is substantially 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.

A rigorous method for selecting the appropriate bin size was recently derived so that the mean integrated squared error (MISE) between the time histogram and the unknown underlying rate is minimized [1]. This derivation assumes that spikes are independently drawn from a given rate. However, individual spike trains bear non-Poissonian features such that the spike occurrence depends on the preceding spike [2], which inevitably deteriorates the optimization. In this contribution, we suggest revising the method for selecting the bin size by considering the possible non-Poissonian features [3].

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

The optimization method aims to find the bin size Δ that minimizes the MISE between the time histogram, λ̂t, and the underlying rate, λt. Assuming that the spikes are sampled from a stochastic process, we use the expected MISE defined by the following formula

\[ \text{MISE} = \frac{1}{T} \int_0^T E(\hat{\lambda}_t - \lambda_t)^2 \, dt, \]  

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

\[ C(\Delta) = \text{MISE} - \left( \lambda_t - \overline{\lambda} \right)^2, \]  

where \( \overline{\lambda} = \frac{1}{T} \int_0^T \lambda \, dt \) represents the time average.

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

\[ C(\Delta) = 2 \left( E(\hat{\theta}_i - \theta_i)^2 \right) - \left( E(\hat{\theta}_i - E(\hat{\theta}_i))^2 \right). \]  

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 ith histogram bin, given by \( k_i/\Delta \), and \( \theta_i \) is the expected height of the histogram bin given by the unknown underlying rate \( \lambda_t \).

\[ \theta_i = \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} \lambda \, dt. \]  

In general, the ratio of the variance to the mean of the event count is called the Fano factor F. For Poisson process, the Fano factor takes 1. The first term of equation (3) can be transformed by using the Fano factor F_i in each bin as follows:

\[ E(\hat{\theta}_i - \theta_i)^2 = F_i \frac{1}{\Delta} E\hat{\theta}_i. \]  

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

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

where \( \bar{k} = \frac{1}{N} \sum_{i=1}^{N} k_i \) is the mean spike count, and \( \hat{F}_i \) 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 \( F \approx C_V^2 \) [4], where \( C_V \) is the coefficient of variation, which is defined as the standard deviation of the ISIs divided by the mean, \( C_V \equiv \Delta \tau / \bar{r} \). However, 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 $L_V$ that can be computed as $L_V = \frac{3}{\kappa - 2} \sum_{j=1}^{k_i - 2} \left( \frac{\tau_{j+1} - \tau_{j}}{\tau_{j} + \tau_{j+1}} \right)^2$ using the ISIs $\{\tau_j\}$ that fall into each bin ($j = 1, 2, \ldots, k_i - 1$) [2]. Here we relate the Fano factor $\hat{F}$ to $L_V$ using the conversion relation, $C_V^2 = 2L_V/(3 - L_V)$, which is derived by assuming gamma processes. On the other hand, if a bin contains less than three spikes, then the ISI variability is not measurable, therefore we suggest setting $\hat{F} = 1$, because the Fano factor approaches unity for an interval that contains few spikes. Taken together, we propose an algorithm for estimating the Fano factor in each bin of a single spike train:

$$\hat{F}_i \equiv \begin{cases} 1, & \text{if } k_i \leq 2, \\ \frac{2L_V}{3 - L_V}, & \text{otherwise}, \end{cases} \quad (7)$$

With Eq. (6) and (7), the cost function can be estimated solely from the spike times. The optimal bin size is obtained by minimizing cost function $C(\Delta)$. The computations of this method are summarized by the following steps.

1. Divide the observation period $T$ into $N$ bins having width $\Delta$, 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 $L_V \equiv \frac{3}{\kappa - 2} \sum_{j=1}^{k_i - 2} \left( \frac{\tau_{j+1} - \tau_{j}}{\tau_{j} + \tau_{j+1}} \right)^2$ 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,

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

where $\bar{k} \equiv \frac{1}{N} \sum_{i=1}^{N} k_i$.
4. Construct the cost function, $C(\Delta) = \frac{2h - v}{2\sigma^2}$.
5. Repeat steps (i) through (iv) while changing the bin size $\Delta$ to search for the optimal bin size that minimizes $C(\Delta)$.

[Note] The Poissonian optimization method is obtained by simply replacing step (ii) with $\hat{F} = 1$.

3 Performance of the Optimization Methods

Here, we compare the performances of the original and revised algorithms by applying them to non-Poissonian spike sequences generated numerically by simulating inhomogeneous gamma processes in the following steps. First, ISIs are drawn independently from the gamma distribution function,

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

where $\Gamma(\kappa) = \int_0^{\infty} x^{\kappa-1} e^{-x} dx$ is the gamma function. Here the non-Poissonian feature can be specified by the shape factor $\kappa$ that determines the ISI variability. Second, the ISIs are consecutively arranged on the time axis to construct a non-Poissonian spike train having fixed rate. Third, the time axis is rescaled with a given time-dependent rate $\{\lambda_t\}$.

Figure 1 exemplifies three types of spike trains derived from an identical sinusoidally modulated rate; they bear different non-Poissonian 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 method are compared with those constructed by the original Poissonian method, illustrating the improved performance achieved by the revision. We also confirmed the effectiveness of the new method not only for simulated spike trains but also for actual spike trains [3].

References