Chapter 2
Estimating the Firing Rate

Shigeru Shinomoto

Abstract   Neuronal activity is measured by the number of stereotyped action potentials, called spikes, elicited in response to a stimulus or the behavioral conditions of an animal. Any nonparametric method for grasping the time-varying rate of spike firing contains a single parameter that controls the jaggedness of the estimated rate, such as the binsize of the time histogram or the bandwidth of the kernel smoother. In most neurophysiological studies, the parameter that determines the interpretation of neuronal activity has been selected subjectively by individual researchers. Recently, theories for objectively selecting the parameter have been developed. This chapter introduces the standard rate estimation tools, such as the peri-stimulus time histogram (PSTH), kernel density estimation, or Bayes estimation, and shows ways of selecting their parameters under the principles of minimizing the mean integrated squared error or maximizing the likelihood. We also sum up the methods in handy recipes that may be useful in practical data analysis.

2.1 Introduction

In the beginning of the last century, Edgar Adrian discerned that a neuron expresses the intensity of a stimulus in the frequency or occurrence rate of stereotyped action potentials (Adrian 1928; Rieke et al. 1997). Since then, the most basic protocol in neurophysiology has been computing the occurrence rate of neuronal action potentials, called spikes or firings, as the correlate of an animal's behavior (Gerstein and Kiang 1960; Johnson 1996; Dayan and Abbott 2001). The rate of neuronal firing, defined in the unit of frequency [Hz], or spikes per second [sp/s], can be computed

S. Shinomoto (✉)

Department of Physics, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto
606-8502, Japan
e-mail: shinomoto@scphys.kyoto-u.ac.jp
url: http://www.ton.scphys.kyoto-u.ac.jp/~shino/english.html

by simply counting the number of spikes and dividing it by the period of the ob-
servation interval. Thus the computation of the firing rate is a mere realization of
simple descriptive statistics, as exemplified by a census of the population.

However, to grasp the temporal modulation of neuronal activity in relation to
sequential behaviors of the animal, one has to go beyond the simple description
and treat a hard problem of inferential statistics in estimating an instantaneous rate
of firing. If neurons were generating temporally regular spikes, the instantaneous
spike rate or the firing frequency may be given by simply inverting the interspike
intervals (ISIs). However, this method leads to a messy diagram for highly irregular
spike trains recorded from cortical neurons in vivo. If, on the contrary, we spend
a long time collecting a large number of spikes for precise estimation of the firing
rate, we cannot grasp the fine temporal modulation. This seemingly naive issue may
arise in any sophisticated method. Any rate estimation tool has a (hyper)parameter,
such as the binsize of a time histogram and the bandwidth of a kernel smoother
that controls the jaggedness of the estimate. The estimated rate may become highly
fluctuating if the binsize or bandwidth is small, and constant in the opposite limit.
In neurophysiological studies, the parameter that critically determines the goodness
of rate estimation has mostly been selected subjectively by individual researchers.

Originally, neurons generate spikes according to their inputs in a mostly deter-
ministic manner, and there is no definite entity for the firing rate in the brain. The
endeavor to capture the rate of spike occurrence from discrete data can be viewed as
a process for compressing information. Rate estimation inevitably depends on the
method of information compression and is therefore not determined uniquely for a
set of data. Nevertheless, the range of plausible principles and estimation methods is
limited, and the rate estimated from a given set of data should not vary among prin-
cipled methods. In applying statistical principles, one views spikes as sample events
derived from a certain underlying probabilistic process and strives to estimate the
underlying probability from the spike data. In this respect, methods for selecting the
parameter based on statistically plausible principles, such as minimizing the mean
integrated squared error (MISE) or maximizing the likelihood, have recently been
developed.

In this chapter, we introduce several standard rate estimation tools, the peri-
stimulus time histogram (PSTH) (Gerstein and Kiang 1960; Abeles 1982), the ker-
nel smoother (Parzen 1962; Richmond et al. 1990; Nawrot et al. 1999), and the
Bayesian estimation method, and unfold the principles for optimizing rate estima-
tion methods. These methods are summed up in handy recipes that may be useful
for neurophysiologists in analyzing their own data. Package programs are available
at

http://www.ton.scphys.kyoto-u.ac.jp/~shino/toolbox/english.htm   or   http://www.
apst.spiketrain-analysis.org/.

2.2 Methods for Estimating the Firing Rate

Let us first describe how to construct a PSTH and a kernel density estimation, given
a set of spike data obtained from repeated trials.

Fig. 2.1 Methods for
estimating instantaneous rate.
A: An underlying spike rate
$r(t)$. **B**: Sample spike raster
diagrams for five trials.
C: A peri-stimulus time
histogram (PSTH).
D: A kernel density
estimation
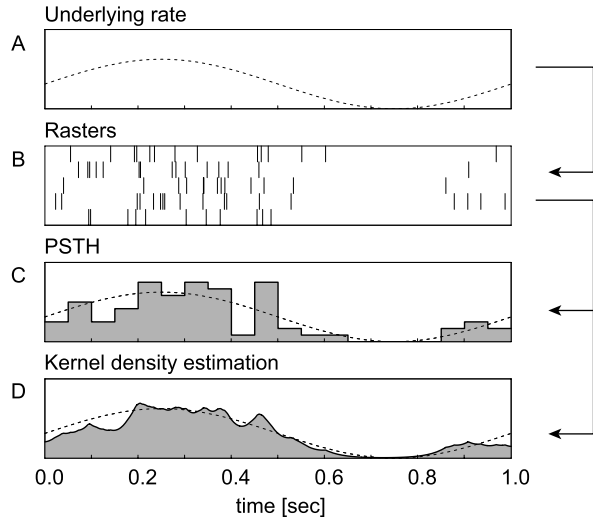
2.2.1 PSTH

A histogram is a realization of a simple counting method and can be constructed
from a set of spike trains, according to the following instructions (Figs. 2.1A–C).

A method for constructing a PSTH

1. Align n spike trains at the onset or offset of stimuli.
2. Divide an observation period T into intervals of width Δ.
3. Count spikes collected from all trials as k_i for ith bin.
4. Draw a bar at the height $k_i/(n\Delta)$ in a time period of $[(i-1)\Delta, i\Delta]$.
5. Repeat 3 and 4 for each bin from $i = 1$ to $N_b(= T/\Delta)$.

In many neurophysiological papers, the height of the PSTH is represented by
the raw spike count per bin. We recommend representing the height of the PSTH
in the units of spike rate for one trial, so that integration over the period T gives
the number of spikes averaged over trials. Due to this normalization procedure, the
heights of PSTHs obtained with different binsize Δ should be approximately equal,
and estimates can be compared across different binsizes.

2.2.2 The Kernel Density Estimation

A kernel density estimation can be obtained by blurring each spike with a kernel
function, according to the following instructions (Figs. 2.1A, B, and D).

A method for constructing a kernel density estimation

1. Align n spike trains at the onset or offset of stimuli.
2. At every spike, apply a kernel function $f_\Delta(t)$ of bandwidth Δ.
3. Divide the summed function by the number of trials n, or

$$\hat{r}(t) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N_s^j} f_\Delta(t - t_i^j), \tag{2.1}$$

where t_i^j is the time of the ith spike in the jth trial, and N_s^j is the total number of spikes recorded in the jth trial.

The height of the density estimation measures the spike rate per one trial, similar to the above-mentioned PSTH. The filtering kernel should satisfy the normalization to unit area, $\int f(t)\,dt = 1$. The kernel is normally assumed to be nonnegative, $f(t) \geq 0$, and to have a finite bandwidth defined by the variance that is normally finite, $\Delta^2 = \int t^2 f(t)\,dt < \infty$, and is often chosen to be symmetric, $f(t) = f(-t)$. Many kernel functions satisfy these conditions (Paulin 1992; Nawrot et al. 1999; Paulin and Hoffman 2001; Cherif et al. 2008): A rectangular "boxcar" kernel may give rise to a jagged density function with an appearance similar to that of the PSTH. The width of the boxcar should be $2\sqrt{3}\Delta$ in order to give the variance of Δ^2,

$$f_\Delta(t) = \begin{cases} \frac{1}{2\sqrt{3}\Delta} & \text{for } -\sqrt{3}\Delta \leq t \leq \sqrt{3}\Delta, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

A smooth density estimation can be obtained by using a smooth kernel such as the Gaussian function,

$$f_\Delta(t) = \frac{1}{\sqrt{2\pi}\,\Delta} \exp\left(-\frac{t^2}{2\Delta^2}\right). \tag{2.3}$$

Alternatively, emphasis may be put on the evidence for spike occurrence from the cusp of the exponential kernel,

$$f_\Delta(t) = \frac{1}{\sqrt{2}\Delta} \exp\left(-\sqrt{2}\left|\frac{t}{\Delta}\right|\right). \tag{2.4}$$

It is noteworthy that exponential functions turn out to be the optimal kernels under the MISE minimization principle for several underlying rates (Koyama and Shinomoto 2004).

Both the Gaussian kernel (2.3) and the exponential kernel (2.4) have infinite support. However, since they decay rapidly, they can be approximated as functions with compact support. For instance, the normalization condition to unit area is in practice not violated even if we cut them off outside the range of $\pm 5\Delta$; the integrated areas
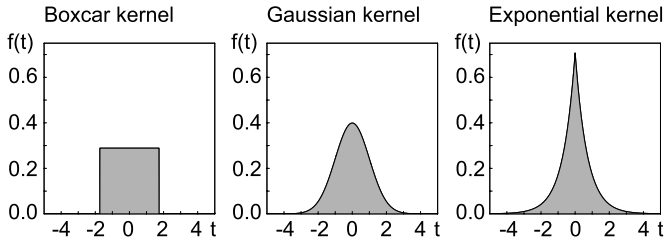
Fig. 2.2 The boxcar, Gaussian, and exponential kernels that give an identical bandwidth, or the variance $\int t^2 f(t)\, dt = 1$

$\int_{-5\Delta}^{5\Delta} f_\Delta(t)\, dt$ become 0.999999 and 0.999151, respectively, for the Gaussian and exponential kernels, thus practically losing nothing in both cases. (See Fig. 2.2.)

2.3 Methods for Optimizing the Rate Estimation

For a single set of spike trains, either a PSTH or a kernel density estimation is not given uniquely, and its shape depends greatly on the choice of the binsize or bandwidth. Figure 2.3 exemplifies three PSTHs constructed from an identical set of spike trains. If the binsize is too small, the time histogram fluctuates greatly, and one cannot discern the underlying spike rate, whereas if it is too large, one cannot capture the time dependence of the rate. There would be an appropriate binsize or bandwidth for each set of spike sequences, based on the goodness-of-fit of a PSTH or a kernel density estimation.

Several plausible principles exist for the goodness-of-fit of the estimator. Here, we introduce two of them and demonstrate formulae that may be practically useful in the application to spike data. One is the principle of minimizing the distance between the estimated rate and the unknown underlying rate measured in the MISE, and another is the principle of maximizing the likelihood function for a given set of data.
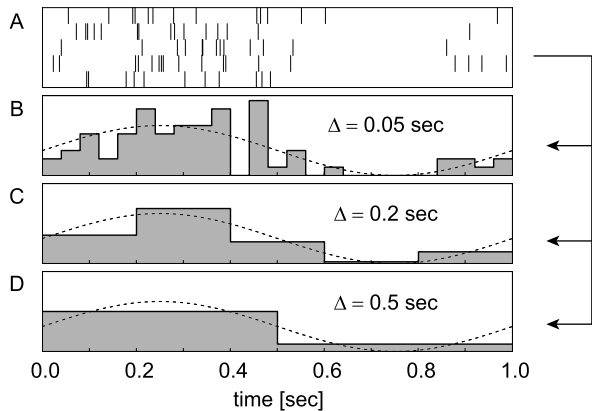

Fig. 2.3 Time histograms with various binsizes. **A**: Spike data given in Fig. 2.1. **B–D**: The time histograms constructed with small, medium, and large binsizes. *Dashed line* represents the underlying rate $r(t)$ from which the spikes were derived

2.3.1 MISE Principle

First, we select the binsize or the bandwidth parameter Δ such that the PSTH or the kernel rate estimator $\hat{r}_\Delta(t)$ best approximates the underlying rate $r(t)$ under the principle of minimizing the MISE,

$$\text{MISE}(\Delta) \equiv \frac{1}{T} \int_0^T E\big(\hat{r}_\Delta(t) - r(t)\big)^2 dt, \tag{2.5}$$

where E refers to the expectation over different realizations of point events, given $r(t)$.

The definition of MISE contains the underlying rate $r(t)$ itself, which is unknown by nature. Nevertheless, using the spike data as a clue, it is possible to minimize the expected MISE with respect to the parameter Δ (Rudemo 1982; Shimazaki and Shinomoto 2007a, 2007b, 2010). Note that the optimization method can be derived rigorously without assuming anything about the time-dependent rate $r(t)$, such as the continuity of the rate. The only assumption needed in deriving the rule is that spikes are drawn from an inhomogeneous Poisson process in which spikes are independently derived from an underlying rate $r(t)$. Though spikes recorded from a biological neuron correlate in each sequence (Shinomoto et al. 2003, 2005, 2009), spikes collected from a number of independent trials are statistically independent, and the superimposed sequence can be approximated as a single inhomogeneous Poisson process (Snyder 1975; Daley and Vere-Jones 2003; Kass et al. 2005).

2.3.1.1 MISE Optimization of PSTH

Leaving the derivation of the rule for minimizing MISE for the PSTH to the literature (Shimazaki and Shinomoto 2007a), we introduce the method of selecting the binsize Δ in a form of a simple recipe in the following.

A method for selecting the binsize of PSTH

1. Compute the average $\bar{k}$ and the variance v of the spike counts $\{k_i\}$ for all bins, $i = 1, 2, \ldots, N_b$,

$$\bar{k} \equiv \frac{1}{N_b} \sum_{i=1}^{N_b} k_i, \tag{2.6}$$

$$v \equiv \frac{1}{N_b} \sum_{i=1}^{N_b} (k_i - \bar{k})^2, \tag{2.7}$$

and evaluate the cost function (Shimazaki and Shinomoto 2007a),

$$C(\Delta) = \frac{2\bar{k} - v}{(n\Delta)^2}, \tag{2.8}$$

where n is the number of trials, and Δ is the binsize.

2. Repeat 1 by moving the initial binning position to take an average of $C(\Delta)$.
3. Repeat 2 while changing Δ to draw the cost as a function of the bin size.
4. The MISE optimal binsize Δ^* is given by the one that minimizes $C(\Delta)$.

Note here that the variance v is not the unbiased variance, but the biased variance as defined by (2.7). The averaging of $C(\Delta)$ over the initial binning positions is useful for limiting the possible fluctuation due to the finiteness of data. Figure 2.4 displays the raw and averaged cost functions computed for the set of spike data given in Fig. 2.3, demonstrating significant fluctuation for a raw cost function.

By applying the optimization method to spike trains whose intensity does not modulate greatly in time, it might come to pass that the original cost function $C(\Delta)$ computed for n spike sequences does not have a minimum or has a minimum at a binsize comparable to the observation period T. This implies that the data are insufficient for estimating the time-dependent rate. The divergence of Δ^* implies that any PSTH of a finite binsize captures a spurious rate and therefore is worse than simply drawing a constant rate, in the sense of MISE (Koyama and Shinomoto 2004). This cost function evaluation may be useful in justifying the estimation of the time-dependent rate, in particular for discussing the presence or absence of oscillatory activity.

Because a shortage of data underlies the divergence of the optimized binsize, one would consider carrying out more experiments to obtain reliable rate estimation. In making the experimental plan, one may want to estimate how many experiments should be added to secure the PSTH resolution one deems sufficient. This can be done through extrapolating the cost function $C(\Delta)$ obtained with a given number of trials n to the case of a different number of trials m, as
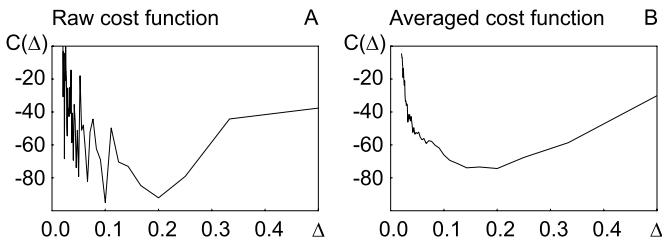

Fig. 2.4 The cost functions $C(\Delta)$ obtained for the set of spike data given in Fig. 2.1B. **A**: Raw cost function obtained for a single set of counting statistics ($\Delta^* = 0.1$). **B**: The cost function averaged over the initial binning positions ($\Delta^* = 0.2$)

$$C(\Delta|m) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{\bar{k}}{n\Delta^2} + C(\Delta), \tag{2.9}$$

where $\bar{k}$ is the average spike count (2.6) obtained from n sequences. Even if the optimal binsize Δ^* diverged for a given number of spike sequences, implying the incapability of drawing a meaningful PSTH, one can still estimate the number of additional experimental trials needed to make the optimal binsize finite, justifying the evaluation of the time-dependent rate from the data.

It is also possible to extend the estimation method from the bar-graph PSTH to the line-graph PSTH, which generally provides a superior goodness-of-fit (Koyama and Shinomoto 2004). We do not go into the details of the line-graph histogram optimization, which is more complicated than the optimization method presented here (Shimazaki and Shinomoto 2007a).

2.3.1.2 MISE Optimization of Kernel Density Estimation

The same principle of minimizing the MISE can also be applied to the kernel density estimation (Shimazaki and Shinomoto 2007b, 2010). Leaving the derivation of the rule to the literature, we introduce the method of selecting the bandwidth Δ in a form of a simple recipe in the following.

A method for selecting the bandwidth of the kernel density estimator

1. Compute the cost function (Shimazaki and Shinomoto 2007b, 2010)

$$C(\Delta) = \frac{1}{n^2}\sum_{i,j}\phi_\Delta(t_i - t_j) - \frac{2}{n^2}\sum_{i\neq j}f_\Delta(t_i - t_j), \tag{2.10}$$

   where $\phi_\Delta(t_i - t_j) \equiv \int f_\Delta(s - t_i)f_\Delta(s - t_j)\,ds$.
2. Repeat 1 while changing Δ to draw the cost as a function of the bandwidth.
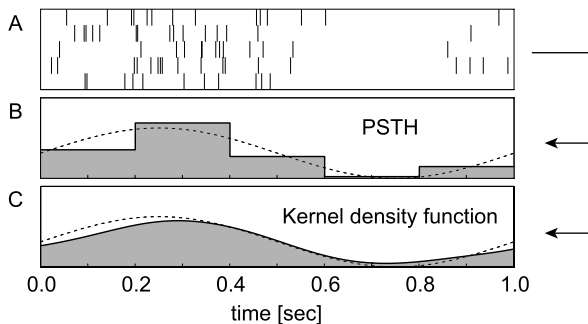3. The MISE optimal bandwidth Δ^* is given by the one that minimizes $C(\Delta)$.

Note here that the computational cost can be significantly reduced by neglecting unnecessary parts in the double summation in (2.10), based on the practical compactness of either the Gaussian or the exponential kernel. In addition, the convoluted function is given simply as $\phi_\Delta(t) = 1/(2\sqrt{\pi}\Delta)\exp(-t^2/4\Delta^2)$, for the Gaussian kernel, $f_\Delta(t) = 1/(\sqrt{2\pi}\Delta)\exp(-t^2/2\Delta^2)$.

Kernel density estimation with a smooth kernel function is generally much superior to PSTH in its goodness-of-fit to the underlying rate, as demonstrated in Fig. 2.5.

2.3.1.3 TIPS

Application programs for optimizing the PSTH and kernel density estimation methods under the MISE principle were provided by Hideaki Shimazaki at

Fig. 2.5 Comparison of the optimized PSTH and optimized kernel density estimator. **A**: Spike data given in Fig. 2.1. **B**: The time histograms of the optimal binsize $\Delta^* = 0.2$. **C**: The Gaussian kernel density estimation of the optimal bandwidth $\Delta^* = 0.1$

http://www.ton.scphys.kyoto-u.ac.jp/~shino/toolbox/sshist/hist.html or http://www.apst.spiketrain-analysis.org/

and

http://www.ton.scphys.kyoto-u.ac.jp/~shino/toolbox/sskernel/kernel.html or http://www.apst.spiketrain-analysis.org/.

By simply copying and pasting a data set of spike times to the application, one can obtain the optimal binsize or bandwidth Δ^* for the PSTH or the kernel, and a list of $\{t, \hat{r}_{\Delta^*}(t)\}$, with which to draw the PSTH or kernel density function. One may also download MATLAB codes for these optimization algorithms.
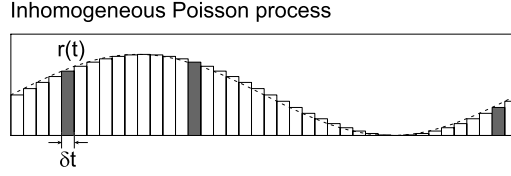
2.3.2 Likelihood Principle

Another standard optimization principle is maximizing the likelihood. This can be done by treating a rate function $r(t)$ as a distribution function of spike times. If one does not have a prior idea for a particular set of rate functions, one may construct a rate estimator nonparametrically. However, one still needs to determine a (hyper)parameter for the smoothness or jaggedness of the estimator. Here, we introduce a relatively simple Bayesian method (Koyama and Shinomoto 2005; Koyama et al. 2007; Shimokawa and Shinomoto 2009) and show the ways to determine a hyperparameter that plays a role similar to the bandwidth of the filtering kernel.

2.3.2.1 Empirical Bayes Method of Rate Estimation

Let us first consider a time-dependent Poisson process in which spikes are derived randomly in time from a given underlying rate $r(t)$. In this process, the probability for spikes to occur at $\{t_i\} \equiv \{t_1, t_2, \ldots, t_{N_s}\}$ in the period of $t \in [0, T]$ is given by

$$p(\{t_i\} \mid r(t)) = \left[\prod_{i=1}^{N_s} r(t_i)\right] \exp\left(-\int_0^T r(t)\,dt\right). \tag{2.11}$$

Inhomogeneous Poisson process

r(t)

δt

Here the exponential term is the survivor function that represents the probability that spikes have not occurred in the interval (Cox and Lewis 1966; Daley and Vere-Jones 2003): The probability for a spike to occur in each short interval of δt is $r(t)\delta t \ll 1$ (see Fig. 2.6), and the probability of having no spike from the time t_1 to t_2 is given by

$$\lim_{\delta t \to 0} \prod_m \big(1 - r(t_m)\delta t\big) = \exp\left(-\int_{t_1}^{t_2} r(t)\,dt\right).$$

We invert the arguments of the conditional probability (2.11) so that the unknown underlying rate is inferred from the spikes observed. This "inverse probability" can be obtained using the Bayes formula

$$p_\beta\big(r(t) \mid \{t_i\}\big) = \frac{p(\{t_i\} \mid r(t))\, p_\beta(r(t))}{p_\beta(\{t_i\})}. \tag{2.12}$$

As a Bayesian prior distribution of $r(t)$, we incorporates the tendency of the estimated rate to be flat by penalizing the large gradient, $|dr(t)/dt|$,

$$p_\beta\big(r(t)\big) \propto \exp\left[-\beta \int_0^T \left(\frac{dr(t)}{dt}\right)^2 dt\right], \tag{2.13}$$

where β is a hyperparameter representing the "flatness"; the estimated rate becomes insensitive/sensitive to individual spike occurrences as β is large/small. The flatness may be replaced by the smoothness by changing the first-order derivative to the second-order derivative. The result of the inference is known to be robust to the order of differentiation in the prior distribution (Nemenman and Bialek 2002). The probability of having spikes at $\{t_i\} \equiv \{t_1, t_2, \ldots, t_{N_s}\}$ is given by the "marginal likelihood function" or the "evidence,"

$$p_\beta\big(\{t_i\}\big) = \int D\{r(t)\}\, p\big(\{t_i\} \mid r(t)\big)\, p_\beta\big(r(t)\big), \tag{2.14}$$

where $\int D\{r(t)\}$ represents a functional integration over all possible paths of the unknown underlying rate $r(t)$. The method of selecting the hyperparameter according to the principle of maximizing the marginal likelihood function is called the Empirical Bayes method (Good 1965; Akaike 1980; MacKay 1992; Carlin and Louis 2000). The marginalization path integral (2.14) for a given set of spike data $\{t_i\}$ can be carried out by the Expectation Maximization (EM) method (Dempster et al. 1977; Smith and Brown 2003) or the Laplace approximation (Koyama and Paninski 2009). With the hyperparameter determined as $\beta = \beta^*$, we can further obtain the maximum a posteriori (MAP) estimate of the rate $\hat{r}(t)$, so that their posterior distribution,

$$p_{\beta^*}\big(r(t) \mid \{t_i\}\big) \propto p\big(\{t_i\} \mid r(t)\big)\, p_{\beta^*}\big(r(t)\big), \tag{2.15}$$

Fig. 2.7 Comparison of the
optimized kernel density
estimator and Empirical
Bayes rate estimator. **A**: Spike
data given in Fig. 2.1. **B**: The
kernel density estimation of
the optimal binsize $\Delta^* = 0.1$.
C: The MAP rate estimate of
the Empirical Bayes method
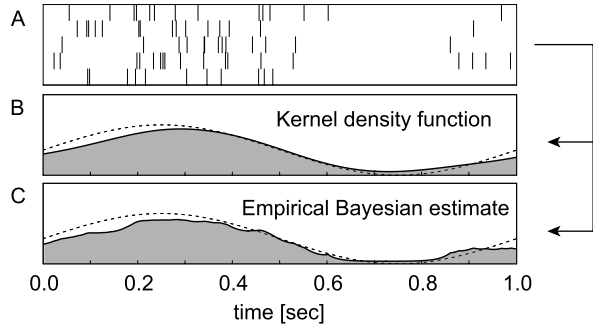
is maximized. The MAP estimate resembles the kernel filtering method in that the
estimated rate $\hat{r}(t)$ is lifted at every spike. The Bayesian MAP rate estimate gives
a similar result to the optimized kernel density estimator for a given set of data
(Fig. 2.7). With the flatness constant β large/small, the rate estimation becomes
insensitive/sensitive to individual spike occurrences. Here we outline the algorithms
for maximizing the marginalized likelihood (2.14) by using the EM method, and
Kalman filtering and smoothing or "point-process filtering" (Brown et al. 1998) for
a given set of spike trains.

An Empirical Bayes method for estimating the rate

1. Perform the marginalization path integral (2.14) using a state-space model
   or the hidden Markov model, in which the rates and spikes are treated as
   states and observations, respectively defined at every time a spike occurred.
2. Optimize the hyperparameter by the EM method.
3. Obtain the expected values and the expected variances of the latent vari-
   ables with Kalman filtering and smoothing.

2.3.2.2 TIPS

An application program that helps the readers to analyze their own neuronal data
with the Empirical Bayesian rate estimation was provided by Takeaki Shimokawa
at

http://www.ton.scphys.kyoto-u.ac.jp/~shino/toolbox/ssBayes/bayes.html or http://
www.apst.spiketrain-analysis.org/.

By simply copying and pasting a data set of spike times, one may obtain the MAP
estimate of the rate and a list of $\{t, \hat{r}(t)\}$, with which one can draw the Bayesian

estimation of the rate. The extension to the simultaneous estimation of the rate and the instantaneous irregularity is given at

http://www.ton.scphys.kyoto-u.ac.jp/~shino/toolbox/ssNeCo09/page_SULAB2.html or http://www.apst.spiketrain-analysis.org/

which are based on the theory developed in Shimokawa and Shinomoto (2009).

2.4 Discussion

In this chapter, we introduced typical rate estimation methods, the PSTH, kernel density estimation, and Bayesian estimation. We also demonstrated optimization of the PSTH and the kernel method under the principle of minimizing the MISE and optimizing the Bayesian rate estimation under the principle of maximizing the likelihood function.

Those who intended to practically perform rate estimation may waver regarding the choice of the multiple methods of rate estimation and multiple principles for parameter optimization to use. There is in fact no oracle that may select one of the different optimization principles, and therefore one may choose any of them. Nevertheless, one may expect that reasonable optimization principles provide similar rate estimations for a given set of spike data. It is interesting in this respect to examine whether the different methods may give qualitatively different conclusions for the detectability of rate modulation. For a small number of spike trains derived from an underlying rate $r(t)$ that fluctuates moderately in time, the optimal binsize may diverge, indicating the incapability of detecting a time-dependent rate. For a doubly stochastic process in which the rate is modulated according to the Ornstein–Uhlenbeck process, we may analytically obtain the conditions for the divergence of the optimized binsize Δ^* (Koyama and Shinomoto 2004) and the optimized flatness constant β^* (Koyama et al. 2007). It is noteworthy that both parameters Δ^* and β^* diverge for the identical condition, $n\sigma^2\tau/\mu \leq 1/2$, where the μ, σ^2, and τ are the mean, variance, and timescale of rate fluctuation, respectively. This implies that the different principles such as MISE minimization for a PSTH and marginal likelihood maximization for the Bayes estimation may render similar detectability for rate modulation.

Though these optimization principles may provide reasonable rate estimations, there may still be room for improvement. The practical problem we did not consider in this chapter is the nonstationarity of the rate, such that neuronal activity exhibits abrupt changes in response to a stimulus given to an animal. In such a case, it would be more efficient to modulate the hyperparameter according to the time-varying circumstances. Methods have been proposed for locally adaptive determination of hyperparameters (Abramson 1982; Sain and Scott 1996, 2002; Loader 1999a, 1999b; DiMatteo et al. 2001; Kass et al. 2003; Endres et al. 2008; Shimazaki and Shinomoto 2010).

When applying any sophisticated optimization method to neurophysiological data, one might feel the optimized binsize or bandwidth to be larger than expected.

In practice, the PSTH binsize chosen in neurophysiological literature has tended to be much smaller than the optimal one. This would be because we are inclined to see the details rather than to ignore unnecessary parts. Therefore, the choice lies with the researcher to not use any of the optimization principles. However, there are circumstances in which objective optimization principles are essential. The most important application of the optimal rate estimation method would be the detectability of a significant rate modulation, such as oscillation in neuronal firing. The presence of rate modulation may be judged from the finiteness of the optimal binsize or bandwidth and the systematic modulation of the estimated rate.

Another important application of rate estimation is the analysis of the intrinsic regularity of neuronal firing: It has been revealed that neuronal firing bears non-Poisson aspects; the spike occurrence depends on the preceding spike. The neuronal firing patterns have been analyzed by paying attention to the variability of ISIs (Kuffler et al. 1957; Gerstein and Mandelbrot 1964; Stein 1965; Baker and Lemon 2000; Barbieri et al. 2001; Oram et al. 1999; Kostal and Lansky 2006). However, ISI analysis is vulnerable to firing rate fluctuation, which tends to broaden the ISI distribution and makes the regular sequence evaluated closer to Poissonian randomness. The analysis of firing patterns includes an essential difficulty in its definition of firing regularity. An apparently random sequence of spikes can be interpreted as being derived either irregularly in time from a constant rate or regularly from a fluctuating rate. To determine which interpretation is more plausible in any given case, we introduced a metric measuring the ISI variability rescaled locally in time (Shinomoto et al. 2003, 2005, 2009) and also considered rescaling the sequence of spikes by the instantaneous firing rate (Reich et al. 1998; Koyama and Shinomoto 2005; Nawrot et al. 2008; Shimokawa and Shinomoto 2009). To carry out the time rescaling correctly, one needs to accurately estimate the instantaneous rate. Inversely, the information on firing regularity may be used in improving the firing rate estimation (Cunningham et al. 2008). In this way, the rate and the regularity are complementary aspects of a single spike train; this interesting issue will be discussed in other chapters in more detail (Chaps. 1, 3).

References

Abeles M (1982) Quantification, smoothing, and confidence-limits for single-units histograms. J Neurosci Methods 5:317–325

Abramson I (1982) On bandwidth variation in kernel estimates—a square root law. Ann Statist 10:1217–1223

Adrian ED (1928) The basis of sensation: the action of the sense organs. Christophers, London

Akaike H (1980) Likelihood and Bayes procedure. In: Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds) Bayesian statistics. University Press, Valencia, p 143

Baker SN, Lemon RN (2000) Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. J Neurophysiol 84:1770–1780

Barbieri R, Quirk MC, Frank LM, Wilson MA, Brown EN (2001) Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. J Neurosci Methods 105:25–37

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. J Neurosci 18:7411–7425

Carlin BP Louis TA (2000) Bayes and empirical bayes methods for data analysis, 2nd edn. Chapman and Hall, New York

Cherif S, Cullen KE, Galiana HL (2008) An improved method for the estimation of firing rate dynamics using an optimal digital filter. J Neurosci Methods 173:165–181

Cox DR, Lewis PAW (1966) The statistical analysis of series of events. Wiley, New York

Cunningham JP, Yu BM, Shenoy KV, Sahani M (2008) Inferring neural firing rates from spike trains using Gaussian processes. Adv Neural Inf Process Syst 20:329–336

Daley D, Vere-Jones D (2003) An introduction to the theory of point processes, vol. 1: Elementary theory and methods, 2nd edn. Springer-Verlag, New York

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

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Roy Statist Soc Ser B 39:1–38

DiMatteo I, Genovese CR, Kass RE (2001) Bayesian curve-fitting with free-knot splines. Biometrika 88:1055–1071

Endres D, Oram M, Schindelin J, Földiák P (2008) Bayesian binning beats approximate alternatives: estimating peri-stimulus time histograms. Adv Neural Inf Process Syst 20:393–400

Gerstein GL, Kiang, NYS (1960) An approach to the quantitative analysis of electrophysiological data from single neurons. Biophys J 1:15–28

Gerstein GL, Mandelbrot B (1964) Random walk models for the spike activity of a single neuron. Biophys J 4:41–68

Good IJ (1965) The estimation of probabilities: an essay on modern Bayesian methods. MIT Press, Cambridge

Johnson DH (1996) Point process models of single-neuron discharges. J Comput Neurosci 3:275–299

Kass RE, Ventura V, Cai C (2003) Statistical smoothing of neuronal data. Network Comput Neural Syst 14:5–15

Kass RE, Ventura V, Brown EN (2005) Statistical issues in the analysis of neuronal data. J Neurophysiol 94:8–25

Kostal L, Lansky P (2006) Classification of stationary neuronal activity according to its information rate. Network Comput Neural Syst 17:193–210

Koyama S, Shinomoto S (2004) Histogram bin-width selection for time-dependent point processes. J Phys A Math Theor 37:7255–7265

Koyama S, Shinomoto S (2005) Empirical Bayes interpretations of random point events. J Phys A Math Theor 38:L531–L537

Koyama S, Shimokawa T, Shinomoto S (2007) Phase transitions in the estimation of event rate: a path integral analysis. J Phys A Math Theor 40:F383–F390

Koyama S, Paninski L (2009) Efficient computation of the maximum a posteriori path and parameter estimation in integrate-and-fire and more general state-space models. J Comput Neurosci doi:10.1007/s10827-009-0179-x

Kuffler SW, Fitzhugh R, Barlow HB (1957) Maintained activity in the cat's retina in light and darkness. J Gen Physiol 40:683–702

Loader CR (1999a) Bandwidth selection: classical or plug-in? Ann Statist 27:415–438

Loader CR (1999b) Local regression and likelihood. Springer-Verlag, New York

MacKay DJC (1992) Bayesian interpolation. Neural Comput 4:415–447

Nawrot M, Aertsen A, Rotter S (1999) Single-trial estimation of neuronal firing rates: from single-neuron spike trains to population activity. J Neurosci Methods 94:81–92

Nawrot MP, Boucsein C, Rodriguez-Molina V, Riehle A, Aertsen A, Rotter S (2008) Measurement of variability dynamics in cortical spike trains. J Neurosci Methods 169:374–390

Nemenman I, Bialek W (2002) Occam factors and model-independent Bayesian learning of continuous distributions. Phys Rev E 65:026137

Oram MW, Wiener MC, Lestienne R, Richmond BJ (1999) Stochastic nature of precisely timed spike patterns in visual system neuronal responses. J Neurophysiol 81:3021–3033

Parzen E (1962) Estimation of a probability density-function and mode. Ann Math Statist 33:1065

Paulin MG (1992) Digital filters for firing rate estimation. Biol Cybern 66:525–531

Paulin MG, Hoffman LF (2001) Optimal filtering rate estimation. Neural Networks 14:877–881

Reich DS, Victor JD, Knight BW (1998) The power ratio and the interval map: spiking models and extracellular recordings. J Neurosci 18:10090–10104

Richmond BJ, Optican LM, Spitzer H (1990) Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. J Neurophysiol 64:351–369

Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1997) Spikes: exploring the neural code. MIT Press, Cambridge

Rudemo M (1982) Empirical choice of histograms and kernel density estimators. Scand J Statist 9:65–78

Sain S, Scott D (1996) On locally adaptive density estimation. J Amer Statist Assoc 91:1525–1534

Sain S, Scott D (2002) Zero-bias locally adaptive density estimators. Scand J Statist 29:441–460

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

Shimazaki H, Shinomoto S (2007b) Kernel width optimization in the spike-rate estimation. Budelli R, Caputi A, and Gomez L (eds) Neural coding 2007, pp 143–146

Shimazaki H, Shinomoto S (2010) Kernel bandwidth optimization in spike rate estimation. J Comput Neurosci, published on line. doi:10.1007/s10827-009-0180-4

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

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

Shinomoto S, Miyazaki Y, Tamura H, Fujita I (2005) Regional and laminar differences in in vivo firing patterns of primate cortical neurons. J Neurophysiol 94:567–575

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

Smith AC, Brown EN (2003) Estimating a state-space model from point process observations. Neural Comput 15:965–991

Snyder D (1975) Random point processes. Wiley, New York

Stein RB (1965) A theoretical analysis of neuronal variability. Biophys J 5:173–194