# Recipes for Selecting the Bin Size of a Histogram

**Hideaki Shimazaki**
Department of Physics, Graduate School of Science
Kyoto University, Kyoto 606-8502, Japan

December 11, 2006

# Contents

# Abstract

A histogram is a series of bar-graphs whose heights represent the number of events within class intervals called bins. In most literature, the bin width that critically determines the goodness of the fit of the histogram to the underlying rate or density function has been selected by an individual author in an unsystematic manner. The thesis proves, in the Poisson point process framework, that the bin width of a histogram can be selected as the one that minimizes the formula (2k-v)/width^2, where k and v are the mean and (biased) variance of the number of data points in the bins.

The resolution of the histogram increases, or the optimal bin size decreases, as the number of sampled data increases by repeating an experimental trial. It is notable that the optimal bin size may diverge if only a small number of experimental trials are available. In this case, any attempt to characterize the underlying function by a histogram will lead to a spurious result. Given a paucity of data, the thesis also provides a method that suggests how many more trials are needed until the set of data can be analyzed with the required resolution.

# Acknowledgements

The contents of this thesis were developed under the supervision of Prof. Shigeru Shinomoto. I should like to express my sincere thanks to Prof. Shinomoto, who welcomed a complete stranger like me into his group in the Nonlinear Dynamics Laboratory. His fair judgment on the quality of research has helped me to tackle significant scientific problems.

When I came to Kyoto three years ago, Shinsuke Koyama and Prof. Shinomoto had already made important theoretical findings regarding the scalings and phase transitions of the optimal histogram bin width. Shinsuke's pioneering work inspired me, and Prof. Shinomoto directed my attention to developing an empirical method to choose the bin size of a histogram.

I enjoyed discussion on the statistics with Takeaki Shimokawa and Ryota Kobayashi. Nanae Matsuno gave me many theoretical insights needed to complete this thesis. I learned a lot about nonlinear science through discussions with Kensuke Arai, Yoji Kawamura, and Hiroya Nakao. I acknowledge Kensuke Arai for proofreading our English manuscript submitted to journals. There were many memorable moments, including *off-campus* activities, that I shared with Kensuke and the colleagues of the nonlinear dynamics group.

Before I entered the Department of Physics at Kyoto University, I was a graduate student of the Department of Neuroscience at Johns Hopkins University. I should like to thank my former supervisor Prof. Ernst Niebur, who advised me to complete my Master's degree at Hopkins. Without his help, I would not have been able to complete Ph.D research here at Kyoto University.

Finally, I dedicate this thesis to my mother and father, who always but implicitly encouraged me to become a scientist.

During the first two years at Kyoto University, I received a scholarship from Japan Student Services Organization. In the last year, I was supported by the Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists.

# Chapter 1

# Introduction

We introduce the problem for selecting the bin size of a histogram. The background and significance of the solutions proposed in this thesis will be described. The theory and methods are also available as publications elsewhere(Shimazaki & Shinomoto, 2007a, 2007b).

## 1.1 Optimal bin size of a histogram

A histogram is a series of bar-graphs whose heights represent the number of events within intervals (see an example on page 6). These intervals, called bins or cells, are created by segmenting the data range of interest. Because of its simplicity, a histogram method is recommended by any elementary textbook of *Statistics*, and is used in scientific literature. A line-graph histogram is frequently used as well. In many applications, a histogram is used to make an inference on the underlying function that generated the data, such as a rate function or a probability density function.

When one wishes to estimate the underlying function by a histogram, the choice of the bin size critically determines the quality of the inference. On the one hand, with a bin size that is too large, one cannot represent details of the underlying function. On the other hand, with a bin size that is too small, the histogram fluctuates greatly and one

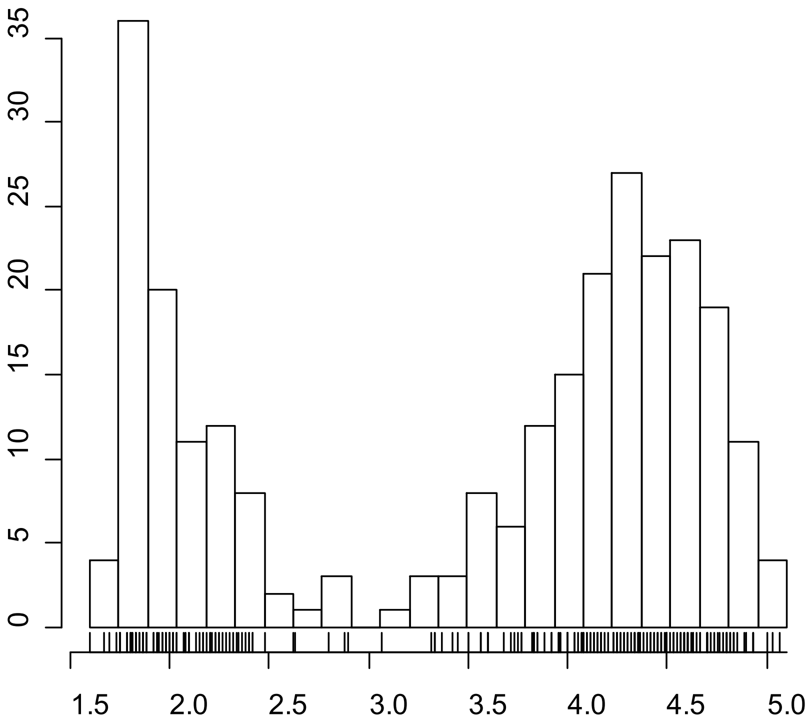


Figure 1.1: A frequency histogram for 272 observations on the duration for eruptions of the Old Faithful geyser in Yellowstone National Park (in minutes). The bin width of the histogram was obtained by the method proposed by this thesis (in page 12).

cannot discern the underlying function. For each set of data, there is an appropriate bin size which needs to be chosen based on the goodness of the fit of the histogram to the underlying function.

The most frequently used criteria to measure the goodness of the fit of a histogram to the underlying function is the mean integrated squared error (MISE),

$$\mathrm{MISE} \equiv \int E_f \left\{\hat{f}_n(x) - f(x)\right\}^2 dx. \tag{1.1}$$

Here $f(x)$ is the underlying function, and $\hat{f}_n(x)$ is a histogram constructed from the data of size $n$. $E_f$ indicates expectation over an ensemble of the histograms $\hat{f}_n(x)$, given $f(x)$. The thesis provides a method to select a bin size that minimizes the MISE. The difficulty of the present problem comes from the fact that the underlying function $f(x)$ is not known.

## 1.2 Background and significance of the proposed methods

Theories on the optimal bin size of a histogram were mostly developed in the framework of the density estimation (Izenman, 1991). There are classical theories that treat how the optimal bin width scales with the total number of data points $n$. It was proven that the optimal bin size scales as $n^{-1/3}$ with regard to the bar-graph density estimator (Révész, 1968; Scott, 1979). By using a Gaussian distribution as a reference, Scott suggested to use the bin width of $\Delta^* = 3.49\,\sigma\,n^{-1/3}$, where $\sigma$ is a standard deviation of the samples. This approach is generally referred to as a plug-in method. Some other extensions of Scott's method can be found in references (Freedman & Diaconis, 1981; Wand, 1997). There are studies on an asymptotically optimal bin size of a histogram based on other standards. Devroye and Gyröfi studied the mean integrated absolute error (Devroye & Györfi, 1985). Kanazawa's approach is based on the Hellinger distance (Kanazawa, 1993). Yet other criteria are the

Akaike's Information Criterion (Taylor, 1987) and the Kullback-Leibler divergence (Hall, 1990).

In the course of my graduate study, I became aware that an empirical method for choosing a histogram bin width was first presented by Rudemo in 1982 (Eq.2.8 in *Scandinavian Journal of Statistics* 9: 65-78). His formula is not well known by rest of the statistical community, and is often ignored in the review papers on the histogram bin width selection. I will demonstrate in Appendix C.1 that, if the total number of data points sampled by experiments is predetermined, our general formula reproduces his result. I also constructed the method of bin size selection for a line-graph histogram under this condition, which will be described in Appendix C.2.

All the previous studies assume that the sample size of a histogram can be strictly controlled: a preset number of samples are randomly picked up from a population. In many practical situations, however, the sample size is not precisely controlled by an experimentalist. Rather, indeterminate number of samples is obtained under a carefully designed experimental protocol with a predetermined time/space constraint. A predominant, but not exclusive, example is the data of the timing of event occurrence in a given observation period. It is thus preferable to have a theory that accords with these experimental designs that frequently appear in practice.

In this thesis, we provide a method for selecting a histogram bin width for the data obtained through such experimental designs. The method is applicable to selecting a size of a cell of a 2-dimensinal histogram as well. In addition to a bar-graph histogram, we also developed a method for selecting the bin size of a line-graph histogram, which is superior to a bar-graph in the goodness of the fit to the underlying function.

When only a small number of experimental trials are available to make a histogram, the estimated optimal bin size may become comparable to the range of the sampled data, implying that by constructing a histogram, it is likely that one obtains spurious results for the estimation (Koyama & Shinomoto, 2004). Because a shortage of data underlies this divergence, one can carry out more experiments to obtain a reliable estimation. The thesis also provides the method that can suggest how many more experimental trials should be performed in order to obtain a meaningful histogram with the required accuracy.

It was pointed out by Koyama and Shinomoto (Koyama & Shinomoto, 2004), and proved in general in Chapter 3, that the optimal bin size scales differently depending on the smoothness of the underlying function. As an application of the proposed method, we also show that the scaling relations of the optimal bin size that appears for a large number of spike sequences can be examined from a relatively small amount of data. The degree of the smoothness of an underlying rate process can be estimated by this method. The naive assumptions made on the underlying function should not be accepted uncritically because it is often possible to examine them from the data.

The organization of the thesis is as follows.

In Chapter 2, we provide user-friendly recipes i) for selecting the bin size of a histogram and ii) for estimating the minimum number of experimental trials required for constructing a histogram.

In Chapter 3, theories behind i) the method of the bin size selection and ii) the method for estimating the minimum number of experimental trials required for constructing a histogram will be described. We also derive the theoretical values of i) the minimum number of experimental trials required for constructing a histogram and ii) the asymptotic optimal bin size for a large number of experimental trials.

In Chapter 4, the empirical methods for i) selecting the bin size, ii) estimating the number of experimental trials required for histogram construction, and iii) estimating the scaling exponents of the optimal bin size are tested by comparing the results with those obtained by the theoretical analysis.

In Appendix A, we generalize the theory developed in Chapter 3. The method for selecting the bin size developed in this appendix is applicable to a histogram with a dimension larger than one. In addition, it is applicable to selecting the bin size of a histogram which is constructed to estimate any function related to the sampled point events.

In Appendix B, we provide a theory and a recipe for creating a line-graph histogram. In addition, a theory and a recipe to estimate the minimum number of experimental trials needed to construct a line-graph histogram will be described.

In Appendix C, the method for selecting the bin size of a bar-graph and a line-graph histogram constructed from a fixed number of samples is described.

# Chapter 2

# Methods

In this chapter, we provide a user-friendly recipe for selecting the bin size of a bar-graph histogram of the data obtained by repeating an identical experimental trial. We also provide the method to estimate the minimum number of experimental trials required to construct a bar-graph histogram. Recipes for a line-graph histogram are presented in Appendix B.2.

## 2.1 Selection of the bin size

A bar-graph histogram is constructed simply by counting the number of events that belong to each bin. We divide the data range of interest into $N$ intervals with each length given by $\Delta$. The number of events accumulated from all $n$ repeated trials in the $i$th interval is counted as $k_i$.

The method to obtain an optimal bin size of a histogram is summarized as 'Algorithm 1' on page 12. In this method, the cost function is computed by using the mean and variance of event counts within the bins with width $\Delta$. The bin size that minimizes this cost function should be selected.

In the same way as in a one-dimensional histogram, a bin size of a $d$-dimensional histogram can be selected by following Algorithm 1.

**Algorithm 1: Bin size selection for a bar-graph histogram**

(i) Divide the data range of interest into $N$ bins of size $\Delta$†, and count the number of events $k_i$ from all $n$ repeated trials that enter the $i$th bin.

(ii) Calculate the mean and variance of the number of events $\{k_i\}$ as,

$$k \equiv \frac{1}{N}\sum_{i=1}^{N} k_i, \text{ and } v \equiv \frac{1}{N}\sum_{i=1}^{N}(k_i - k)^2.$$

(iii) Compute the cost function,

$$C_n(\Delta) = \frac{2k - v}{(n\Delta)^2}.$$

(iv) Repeat i through iii while changing the bin size $\Delta$. Find $\Delta^*$ that minimizes $C_n(\Delta)$.

† For a histogram with a dimension larger than one, $\Delta$ is a volume of a bin.

In this case, the bin size $\Delta$ is a volume of a $d$-dimensional cube. The mean and variance are computed from the event counts in all the bins that segment a total space. A $d$-dimensional histogram will be treated in Appendix A.

We also constructed the method for selecting the bin size of a line-graph histogram, which is summarized as 'Algorithm 3' in Appendix B.2.

## 2.2 Extrapolation of the bin size

The number of experimental trials plays an important role in determining the resolution of the histogram. We propose the method which can suggest how many more experimental trials should be performed in order to construct a meaningful bar-graph histogram with the resolution we deem sufficient. The method is summarized in Algorithm 2 on page 2.2. In this method, the cost function (in Algorithm 1) computed to obtain the bin size of a histogram for the currently available number of experimental trials is modified to construct a cost function for any number of experimental trials. The bin size that minimizes this 'extrapolated' cost function is the estimated optimal bin size for the desired number of experimental trials.

For a very small number of data, the optimal bin size may become comparable to the observed data range. In this case, we propose to construct the extrapolated cost functions for a larger number of experimental trials than the current number of trials. By examining the bin sizes that minimize these extrapolated cost functions, one can find the number of experimental trials above which the estimated bin sizes become smaller than the observed data range. In this way, the method summarized in Algorithm 2 could tell how many more experimental trials should be performed to obtain a meaningful histogram.

The extrapolation method for a line-graph histogram is summarized as 'Algorithm 3' in Appendix B.3.

**Algorithm 2: Extrapolation method for a bar-graph histogram**

A Construct the extrapolated cost function,

$$C_m(\Delta|n) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{k}{n\Delta^2} + C_n(\Delta),$$

using the mean $k$ and variance $v$ of the number of events obtained from $n$ trials. $C_n(\Delta)$ is the cost function computed for $n$ trials with the Method 1.

B Search for $\Delta_m^*$ that minimizes $C_m(\Delta|n)$. $\Delta_m^*$ is the expected optimal bin size for $m$ trials.

C Repeat A and B while changing $m$, and plot $1/\Delta_m^*$ vs $1/m$ to search for the critical value $1/m = 1/\hat{n}_c$ above which $1/\Delta_m^*$ practically vanishes.

# Chapter 3

# Theory

In this chapter, we develop a theory on the optimal bin size of a histogram constructed from the data sampled from a Poisson point process. If the data obtained by an experimental trial is mutually independent, the sampling procedure is modeled by the inhomogeneous Poisson point process, which is the simplest point process in that the event occurrence depends only on the instantaneous rate. Even if the data obtained by an experimental trial is mutually dependent, as long as the samples are obtained by repeating an identical trial, the accumulated data are virtually independent, and obeys the inhomogeneous Poisson point process due to a general limit theorem (Cox, 1962; Snyder, 1975; Daley & Vere-Jones, 1988).

## 3.1 Construction of a bar-graph histogram

A bar-graph histogram is constructed simply by counting the number of events that belong to each bin of width $\Delta$. We divide the data range $T$ into $N$ intervals with each length given by $\Delta = T/N$. The number of events accumulated from all $n$ sequences in the $i$th interval is counted as $k_i$. The bar height at the $i$th bin is given as $k_i/n\Delta$. Figure 3.1 shows the schematic diagram for the construction of a bar-graph histogram.

Given a bin of width $\Delta$, the expected height of a bar graph for

$t \in [0, \Delta]$ is the time-averaged rate,

$$\theta = \frac{1}{\Delta} \int_0^{\Delta} \lambda_t \, dt. \tag{3.1}$$

The total number of events $k$ from $n$ sequences that enter a bin of width $\Delta$ obeys a Poisson distribution with the expected number $n\Delta\theta$,

$$p(k \,|\, n\Delta\theta) = \frac{(n\Delta\theta)^k}{k!} e^{-n\Delta\theta} \,. \tag{3.2}$$

The unbiased estimator for $\theta$ is given as $\hat{\theta} = k/(n\Delta)$, which is the empirical height of the bar graph for $t \in [0, \Delta]$.

## 3.2 Selection of the bin size

In this thesis, we assess the goodness of the fit of the estimator $\hat{\lambda}_t$ to the underlying rate $\lambda_t$ over the data range $T$ by the mean integrated squared error (MISE),

$$\mathrm{MISE} \equiv \frac{1}{T} \int_0^{T} E\,(\hat{\lambda}_t - \lambda_t)^2 \, dt, \tag{3.3}$$

where $E$ refers to the expectation over different realization of point events, given $\lambda_t$.

By segmenting the range $T$ into $N$ intervals of size $\Delta$, the MISE defined in Eq.(3.3) can be rewritten as

$$\mathrm{MISE} = \frac{1}{\Delta} \int_0^{\Delta} \frac{1}{N} \sum_{i=1}^{N} E \left\{ \hat{\theta}\,(i) - \lambda_{t+(i-1)\Delta} \right\}^2 dt, \tag{3.4}$$

where $\hat{\theta}\,(i)$ is an empirical height of a bar-graph histogram at $i$th bin. Hereafter we denote the average over those segmented rate $\lambda_{t+(i-1)\Delta}$ as an average over an ensemble of (segmented) rate functions $\{\lambda_t\}$ defined in an interval of $t \in [0, \Delta]$:

$$\mathrm{MISE} = \frac{1}{\Delta} \int_0^{\Delta} \left\langle E\,(\,\hat{\theta} - \lambda_t\,)^2 \right\rangle dt. \tag{3.5}$$

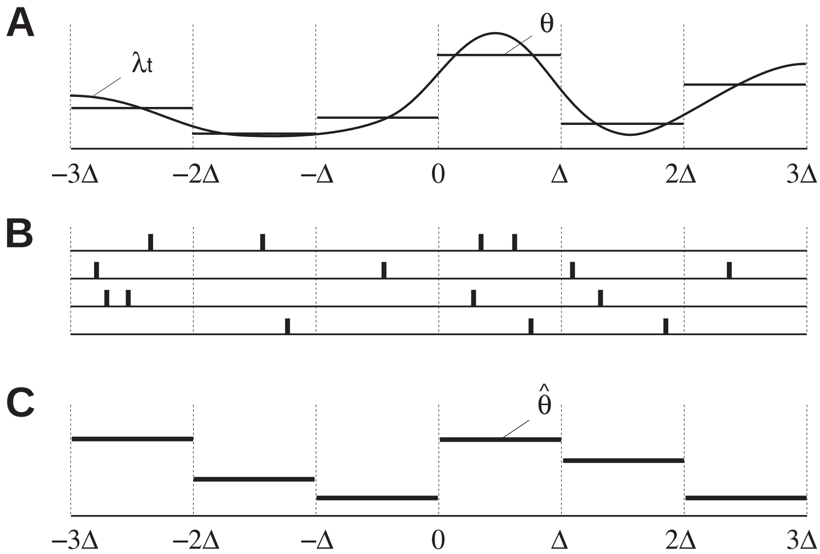


Figure 3.1: The bar–graph histogram. A: an underlying spike rate, $\lambda_t$. The horizontal bars indicate the time averaged rates $\theta$ for individual bins of width $\Delta$. B: sequences of spikes derived from the underlying rate. C: a time histogram for the sample sequences of spikes. The estimated rate $\hat{\theta}$ is the total number of spikes $k$ that entered each bin, divided by the number of sequences $n$ and the bin size $\Delta$.

The expectation $E$ now refers to the average over the event count, or $\hat{\theta} = k/(n\Delta)$, given a rate function $\lambda_t$, or its mean value, $\theta$. The MISE can be decomposed into two parts,

$$\begin{aligned} \text{MISE} &= \frac{1}{\Delta}\int_0^{\Delta} \left\langle E\left(\hat{\theta} - \theta + \theta - \lambda_t\right)^2 \right\rangle dt \\ &= \left\langle E(\hat{\theta} - \theta)^2 \right\rangle + \frac{1}{\Delta}\int_0^{\Delta} \left\langle (\lambda_t - \theta)^2 \right\rangle dt. \end{aligned} \tag{3.6}$$

The first and second terms are respectively the stochastic fluctuation of the estimator $\hat{\theta}$ around the expected mean rate $\theta$, and the temporal fluctuation of $\lambda_t$ around its mean $\theta$ over an interval of length $\Delta$, averaged over the segments.

The second term of Eq.(3.5) can further be decomposed into two parts,

$$\frac{1}{\Delta}\int_0^{\Delta} \left\langle (\lambda_t - \langle\theta\rangle)^2 \right\rangle dt - \left\langle (\theta - \langle\theta\rangle)^2 \right\rangle . \tag{3.7}$$

The first term in the rhs of Eq.(3.7) represents a mean squared fluctuation of the underlying rate $\lambda_t$ from the mean rate $\langle\theta\rangle$, and is independent of the bin size $\Delta$, because

$$\frac{1}{\Delta}\int_0^{\Delta} \left\langle (\lambda_t - \langle\theta\rangle)^2 \right\rangle dt = \frac{1}{T}\int_0^{T} (\lambda_t - \langle\theta\rangle)^2 \, dt. \tag{3.8}$$

We define a cost function by subtracting this term from the original MISE,

$$\begin{aligned} C_n'(\Delta) &\equiv \text{MISE} - \frac{1}{T}\int_0^{T} (\lambda_t - \langle\theta\rangle)^2 \, dt \\ &= \left\langle E(\hat{\theta} - \theta)^2 \right\rangle - \left\langle (\theta - \langle\theta\rangle)^2 \right\rangle . \end{aligned} \tag{3.9}$$

The second term in Eq.(3.9) represents the temporal fluctuation of the expected mean rate $\theta$ for individual intervals of width $\Delta$. As the expected mean rate is not an observable quantity, we must replace the fluctuation of the expected mean rate with that of the observable estimator $\hat{\theta}$. Using the decomposition rule for an unbiased estimator ($E\hat{\theta} = \theta$),

$$\begin{aligned} \left\langle E(\hat{\theta} - \langle E\hat{\theta}\rangle)^2 \right\rangle &= \left\langle E(\hat{\theta} - \theta + \theta - \langle\theta\rangle)^2 \right\rangle \\ &= \left\langle E(\hat{\theta} - \theta)^2 \right\rangle + \left\langle (\theta - \langle\theta\rangle)^2 \right\rangle , \end{aligned} \tag{3.10}$$

the cost function is transformed into

$$C_n(\Delta) = 2\left\langle E(\hat{\theta} - \theta)^2 \right\rangle - E\left\langle (\hat{\theta} - \langle E\hat{\theta}\rangle)^2 \right\rangle . \tag{3.11}$$

The second term can be taken apart into two parts,

$$\begin{aligned} \left\langle E(\hat{\theta} - \langle E\hat{\theta}\rangle)^2 \right\rangle &= E\left\langle (\hat{\theta} - \langle\hat{\theta}\rangle + \langle\hat{\theta}\rangle - \langle E\hat{\theta}\rangle)^2 \right\rangle \\ &= E\left\langle (\hat{\theta} - \langle\hat{\theta}\rangle)^2 \right\rangle + E(\langle\hat{\theta}\rangle - \langle E\hat{\theta}\rangle)^2 \end{aligned} \tag{3.12}$$

The second term of this equation is not dependent on the choice of the bin size. Therefore, we add this term to the cost function $C_n^{'}(\Delta)$, and redefine the cost function as

$$\begin{aligned} C_n(\Delta) &= C_n^{'}(\Delta) + E(\langle\hat{\theta}\rangle - \langle E\hat{\theta}\rangle)^2 \\ &= 2\left\langle E(\hat{\theta}-\theta)^2\right\rangle - E\left\langle(\hat{\theta}-\langle\hat{\theta}\rangle)^2\right\rangle. \end{aligned} \tag{3.13}$$

Due to the assumed Poisson nature of the point process, the number of events $k$ counted in each bin obeys a Poisson distribution: the variance of $k$ is equal to the mean. For the estimated rate defined as $\hat{\theta} = k/(n\Delta)$, this variance-mean relation corresponds to

$$E(\hat{\theta}-\theta)^2 = \frac{1}{n\Delta}E\hat{\theta}. \tag{3.14}$$

By incorporating Eq.(3.14) into Eq.(3.13), the cost function is given as a function of the estimator $\hat{\theta}$,

$$C_n(\Delta) = \frac{2}{n\Delta}E\left\langle\hat{\theta}\right\rangle - E\left\langle(\hat{\theta}-\langle\hat{\theta}\rangle)^2\right\rangle. \tag{3.15}$$

Therefore, the cost function can be estimated by using the mean and variance of a sample histogram,

$$C_n(\Delta) = \frac{2}{n\Delta}\left\langle\hat{\theta}\right\rangle - \left\langle(\hat{\theta}-\langle\hat{\theta}\rangle)^2\right\rangle. \tag{3.16}$$

The optimal bin size is obtained by minimizing the cost function $C_n(\Delta)$:

$$\Delta^* \equiv \arg\min_{\Delta} C_n(\Delta). \tag{3.17}$$

By replacing $\hat{\theta}$ in Eq.(3.16) with the sample event counts, the method is converted into a user-friendly recipe summarized in Algorithm 1.

## 3.3 Extrapolation of the bin size

With the method developed in the preceding subsection, we can determine the optimal bin size for a given set of experimental data. In this section, we develop a method to estimate how the optimal bin size decreases when more experimental trials are added to the data set.

Assume that we are in possession of $n$ event sequences. The fluctuation of the expected mean rate $\langle(\theta - \langle\theta\rangle)^2\rangle$ in Eq.(3.9) is replaced with the empirical fluctuation of the histogram $\hat{\theta}_n$ using the decomposition rule for the unbiased estimator $\hat{\theta}_n$ satisfying $E\hat{\theta}_n = \theta$,

$$\left\langle E(\hat{\theta}_n - \langle E\hat{\theta}_n\rangle)^2 \right\rangle = \left\langle E(\hat{\theta}_n - \theta + \theta - \langle\theta\rangle)^2 \right\rangle$$
$$= \left\langle E(\hat{\theta}_n - \theta)^2 \right\rangle + \left\langle(\theta - \langle\theta\rangle)^2\right\rangle .$$

The expected cost function $C'_m(\Delta)$ for $m$ sequences can be obtained by substituting the above equation into Eq.(3.9), yielding

$$C'_m(\Delta) = \left\langle E(\hat{\theta}_m - \theta)^2 \right\rangle + \left\langle E(\hat{\theta}_n - \theta)^2 \right\rangle - \left\langle E(\hat{\theta}_n - \langle E\hat{\theta}_n\rangle)^2 \right\rangle$$
$$= \left\langle E(\hat{\theta}_m - \theta)^2 \right\rangle - \left\langle E(\hat{\theta}_n - \theta)^2 \right\rangle + C'_n(\Delta) . \tag{3.18}$$

where $C_n(\Delta)$ is the original cost function, Eq.(3.16), computed using the estimators $\hat{\theta}_n$. We then obtain the extrapolated cost function $C_m(\Delta)$ given $n$ sequences as

$$C_m(\Delta|n) = \left\langle E(\hat{\theta}_m - \theta)^2 \right\rangle - \left\langle E(\hat{\theta}_n - \theta)^2 \right\rangle + C_n(\Delta) . \tag{3.19}$$

Here the bias term in Eq. 3.12, which is not related to the choice of bin size, was eliminated. Using the variance-mean relation for the Poisson distribution, Eq.(3.14), and

$$E(\hat{\theta}_m - \theta)^2 = \frac{1}{m\Delta} E\hat{\theta}_m = \frac{1}{m\Delta} E\hat{\theta}_n, \tag{3.20}$$

we obtain

$$C_m(\Delta|n) = \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{\Delta} \left\langle E\hat{\theta}_n \right\rangle + C_n(\Delta) . \tag{3.21}$$

By replacing the expectation with sample event count averages, the cost function for $m$ sequences can be extrapolated as $C_m(\Delta|n)$ with this formula, using the sample mean $k$ and variance $v$ of the numbers of events, given $n$ sequences and the bin size $\Delta$. The extrapolation method is summarized in Table 2.

It may come to pass that the original cost function $C_n(\Delta)$ computed for $n$ event sequences does not have a minimum, or have a minimum at a bin size comparable to the observed data range $T$. In such a case, with the method summarized in Table 2, one may estimate the critical number of sequences $n_c$ above which the cost function has a finite bin size $\Delta^*$, and consider carrying out more experiments to obtain a reasonable rate estimation. In the case that the optimal bin size exhibits continuous divergence, the cost function can be expanded as

$$C_n(\Delta) \sim \mu \left( \frac{1}{n} - \frac{1}{n_c} \right) \frac{1}{\Delta} + u \frac{1}{\Delta^2}, \tag{3.22}$$

where we have introduced $n_c$ and $u$, which are independent of $n$. The optimal bin size undergoes a phase transition from the vanishing $1/\Delta^*$ for $n < n_c$ to a finite $1/\Delta^*$ for $n > n_c$. In this case, the inverse optimal bin size is expanded in the vicinity of $n_c$ as $1/\Delta^* \propto (1/n - 1/n_c)$. We can estimate the critical value $\hat{n}_c$ by applying this asymptotic relation to the set of $\hat{\Delta}^*_m$ estimated from $C_m(\Delta|n)$ for various values of $m$:

$$\frac{1}{\Delta^*_m} \propto \left( \frac{1}{m} - \frac{1}{\hat{n}_c} \right). \tag{3.23}$$

It should be noted that there are cases that the optimal bin size exhibits discontinuous divergence from a finite value. Even in such cases, the plot of $\{1/m, 1/\Delta^*\}$ could be useful in exploring a discontinuous transition from nonvanishing values of $1/\Delta^*$ to practically vanishing values.

## 3.4 Theoretical cost function

In this section, we obtain a "theoretical" cost function directly from a process with a known underlying rate, $\lambda_t$. Note that this theoretical cost function is not available in real experimental conditions in which the underlying rate is not known.

The present estimator $\hat{\theta} \equiv k/(n\Delta)$ is a uniformly minimum variance unbiased estimator (UMVUE) of $\theta$, which achieves the lower bound of the Cramér-Rao inequality (Blahut, 1987; Cover & Thomas, 1991),

$$E(\hat{\theta} - \theta)^2 = \left[ -\sum_{k=0}^{\infty} p\left(k \,|\, n\Delta\theta\right) \frac{\partial^2 \log p\left(k \,|\, n\Delta\theta\right)}{\partial\theta^2} \right]^{-1} = \frac{\theta}{n\Delta}. \qquad (3.24)$$

Inserting this into Eq. 3.9, the cost function is represented as

$$\begin{aligned} C_n\left(\Delta\right) &= \frac{\langle\theta\rangle}{n\Delta} - \left\langle (\theta - \langle\theta\rangle)^2 \right\rangle \\ &= \frac{\mu}{n\Delta} - \frac{1}{\Delta^2} \int_0^{\Delta} \int_0^{\Delta} \phi\left(t_1 - t_2\right)\, dt_1 dt_2, \end{aligned} \qquad (3.25)$$

where $\mu = \langle\theta\rangle$ is the mean rate, and $\phi(t_1 - t_2) = \langle(\lambda_{t_1} - \mu)(\lambda_{t_2} - \mu)\rangle$ is the autocorrelation function of the rate fluctuation, $\lambda_t - \mu$.

### 3.4.1 Divergence of the optimal bin size

We obtain analytical solution of the critical number $n_c$, below which the optimal bin size that minimizes the cost function diverges. The cost function with a large bin size can be rewritten as

$$\begin{aligned} C_n\left(\Delta\right) &= \frac{\mu}{n\Delta} - \frac{1}{\Delta^2} \int_{-\Delta}^{\Delta} (\Delta - |t|)\phi(t)\, dt \\ &\sim \frac{\mu}{n\Delta} - \frac{1}{\Delta} \int_{-\infty}^{\infty} \phi(t)\, dt + \frac{1}{\Delta^2} \int_{-\infty}^{\infty} |t|\phi(t)\, dt, \end{aligned} \qquad (3.26)$$

based on the symmetry $\phi(t) = \phi(-t)$ for a stationary process. Eq. 3.26 can be identified with Eq. 3.22 with parameters

$$n_c = \mu \bigg/ \int_{-\infty}^{\infty} \phi(t)\, dt\ , \qquad (3.27)$$

$$u = \int_{-\infty}^{\infty} |t|\phi(t)\, dt. \qquad (3.28)$$

The minimum number of repeated trials required for the construction of a histogram is given by $\lceil n_c \rceil$. For the process giving the correlation of the form $\phi\,(t) = \sigma^2 e^{-|t|/\tau}$ and the process giving a Gaussian correlation $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$ , the critical numbers are obtained as $n_c = \mu/2\sigma^2\tau$ and $n_c = \mu/\sigma^2\tau\sqrt{\pi}$, respectively.

Note that the expansion of the cost function with respect to $1/\Delta$ assumes that the optimal bin size takes an infinitely large value at the onset of the transition. However, the finite optimal bin size may appear suddenly as we increase the number of trials. The transitions of the optimal bin width from infinite to finite length can be systematically examined by considering Fourier domain of the cost function.

The Fourier transform of the window function, whose height is $1/\Delta$ within the window of width $\Delta$ and zero outside the window, is a sinc function given by

$$H_\Delta(\omega) = \int_{-\Delta/2}^{\Delta/2} \frac{1}{\Delta} e^{-j\omega t}\, dt = \frac{\sin(\omega\Delta/2)}{\omega\Delta/2}. \tag{3.29}$$

Let the Fourier transform of the correlation $\phi(t)$ be $\Phi(\omega)$ (the power spectrum of the rate process). Using $H_\Delta(\omega)$ and $\Phi(\omega)$, the cost function Eq. 3.25 is written as

$$C_n(\Delta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H_\Delta(\omega)^2 \left\{ \frac{\mu}{n} - \Phi(\omega) \right\} d\omega. \tag{3.30}$$

The cost function vanishes for the large bin size: $\lim_{\Delta\to\infty} C_n\,(\Delta) = 0$. Therefore, if the finite optimal bin size $\Delta^*$ exists, it should satisfy $C_n\,(\Delta^*) < 0$.

We first examine the cost function of the stochastic processes whose power spectrum has a maximum at origin ($\Phi(0) > \Phi(\omega)$ for $\omega \neq 0$). For a small number of sequences that satisfies $\mu/n > \Phi(0)$, the cost function is positive for all $\Delta$. Therefore the optimal bin size diverges if $n < \mu\,/\Phi(0)$. The critical number of trial is then given by $n_c = \mu\,/\Phi(0)$, which is identical to Eq.3.27. To examine the condition near the transition point, suppose that $n$ is slightly larger than $n_c$. The term $\mu/n - \Phi(\omega)$ in the integral of Eq.3.30 is negative in the vicinity of the

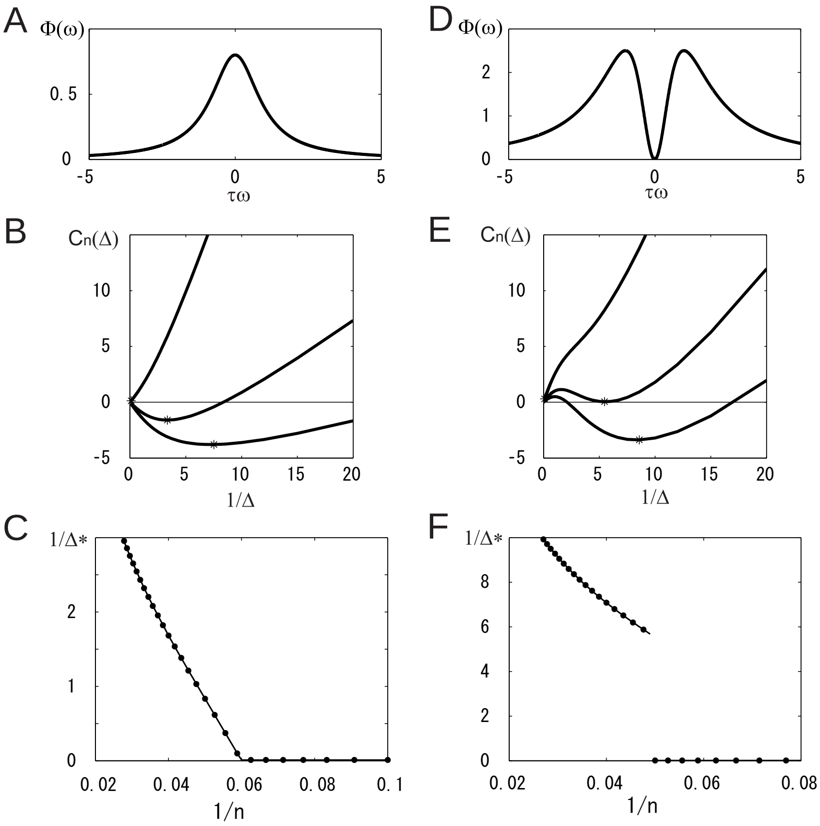


Figure 3.2: The first and the second order phase transitions observed for optimal bin sizes with two different rate processes. (Left; A, B, C) a process whose mean is $\mu = 30$ and correlation function is $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$ ($\sigma^2 = 9, \tau = 0.1$) (Right; D, E, F) a process whose mean is $\mu = 30$ and correlation function is $\phi(t) = \sigma^2(1 - |t|/\tau)e^{-|t|/\tau}$ ($\sigma^2 = 25, \tau = 0.1$). (Top; A, D) Power spectra of the rate process. (Middle; B, E) Theoretical cost functions derived from Eq.3.25 (B $n = 10, 40,$and 100; E $n = 10, 20,$and 30). The asterisks indicate optimal bin sizes (Bottom, CF) The inverse of optimal bin sizes as a function of the inverse of $n$. The dots represent the optimal bin sizes for integer $n$.

origin and positive elsewhere. As the bin size $\Delta$ increases, the spectra of $H_\Delta(\omega)^2$ vanish except for the origin whose height is kept constant as 1. Therefore the narrow negative spectrum range of $\mu/n - \Phi(\omega)$ can be exploited by $H_\Delta(\omega)^2$ with a large bin size: An appropriately large bin size yields the negative cost function. This indicates that the optimal bin size appears as an infinitely large $\Delta$. This type of phase transitions is called the second order (continuous) phase transitions. For example, a process with a Gaussian correlation function ($\phi(t) = \sigma^2 e^{-t^2/\tau^2}$) displayed in Fig. 3.4.1A exhibit the second order phase transition.

For the stochastic processes whose power spectrum has a maximum other than origin, it is possible that the first order (discontinuous) phase transitions take place. Figure 3.4.1A displays the power spectrum of the process with a correlation function $\phi(t) = \sigma^2(1 - |t|/\tau)e^{-|t|/\tau}$. This process has a negative correlation for $|t| > \tau$, and the power spectrum at origin is not a maximum. The cost function analytically obtained from Eq. 3.30 (Figure 3.4.1E) reveals that, when the number of sequences is increased, two local minima appear (one at origin and the other at a finite bin size). The optimal bin size undergoes a discontinuos phase transition as shown in Figure 3.4.1F. The example is an extreme case, where the first order phase transition always occur because $\Phi(0) = 0$. It should be noted that even if the power spectrum has a maximum other than origin, the second order phase transition could take place, too.

### 3.4.2 Scaling of the optimal bin size

Based on the same theoretical cost function Eq. 3.25, we can also derive scaling relations of the optimal bin size, which is achievable with a large number of event sequences. With a small $\Delta$, the correlation of rate fluctuation $\phi(t)$ can be expanded as

$$\phi(t) = \phi(0) + \phi'(0_+)|t| + \frac{1}{2}\phi''(0)t^2 + O\left(|t|^3\right),$$

and we obtain an expansion of Eq. 3.25 with respect to $\Delta$,

$$C_n(\Delta) = \frac{\mu}{n\Delta} - \phi(0) - \frac{1}{3}\phi'(0_+)\Delta - \frac{1}{12}\phi''(0)\Delta^2 + O(\Delta^3). \quad (3.31)$$

The optimal bin size $\Delta^*$ is obtained from $dC_n(\Delta)/d\Delta = 0$. For a rate that fluctuates smoothly in time, the correlation function is a smooth function of $t$, resulting in $\phi'(0) = 0$ due to the symmetry $\phi(t) = \phi(-t)$. In this case, we obtain the scaling relation

$$\Delta^* \sim \left(-\frac{6\mu}{\phi''(0)\,n}\right)^{1/3}. \tag{3.32}$$

For a rate that fluctuates in a zigzag pattern, in which the correlation of rate fluctuation has a cusp at $t = 0$ ($\phi'(0_+) < 0$), we obtain the scaling relation

$$\Delta^* \sim \left(-\frac{3\mu}{\phi'(0_+)n}\right)^{1/2}, \tag{3.33}$$

by ignoring the second order term in Eq.3.31. Examples that exhibit this type of scaling are the Ornstein-Uhlenbeck process and the random telegraph process (Kubo & Hashitsume, 1985; Gardiner, 1985).

# Chapter 4

# Results

In this chapter, we apply the presently developed methods for the bar-graph histogram and the line-graph histogram to sequences of events generated by the simulations of inhomogeneous Poisson point processes.

## 4.1 Selection of the bin size

We applied the method for estimating the optimal bin size of a bar-graph histogram to a set of event sequences derived from an inhomogeneous Poisson process. The rate of the Poisson point process used in the simulation is a stationary stochastic process. Figure 4.1A shows the 'empirical' cost function $C_n(\Delta)$ computed from a set of $n$ event sequences according to Algorithm 2.1. The empirical cost function approximates the 'theoretical' cost function for the rate process, according to Eq. 3.25 in Section 3. In Figure 4.1D, a histogram with the optimal bin size is compared with those with non-optimal bin sizes, demonstrating the effectiveness of optimizing the bin size. We also constructed a method for selecting bin size of a line-graph histogram, which is summarized as 'Algorithm 3' in Appendix B.2. A line graph can be constructed by simply connecting top-centers of adjacent bar graphs. Figure 4.2 compares the optimal bar-graph histogram and the optimal line-graph histogram, obtained from the same set of event sequences,

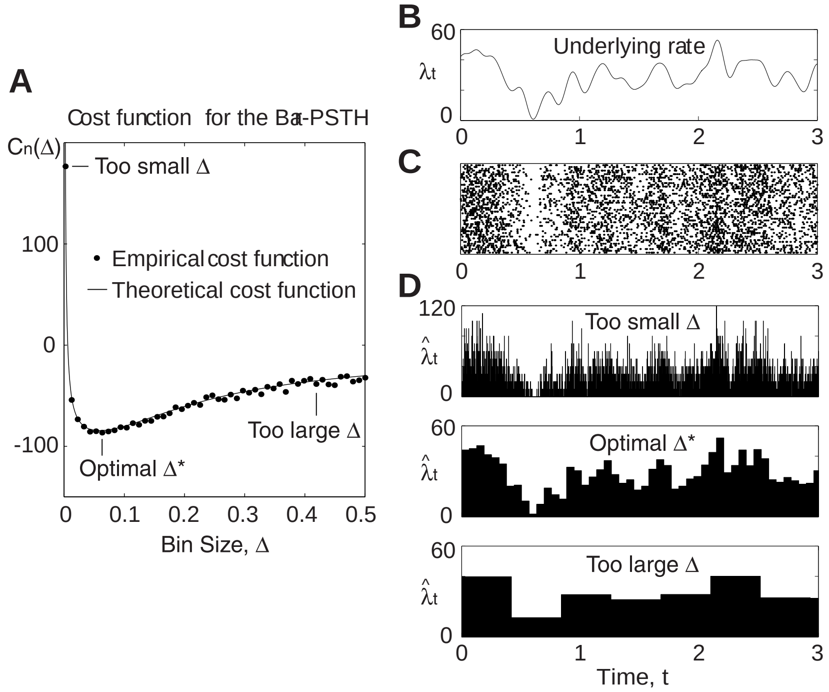


Figure 4.1: Construction of the optimal bar-graph histogram from point events. A: (Dots): an 'empirical' cost function, $C_n(\Delta)$, computed from the data according to the Algorithm 1. Here, 50 sequences of events for an interval of 20 are derived from an inhomogeneous Poisson process characterized by the mean rate $\mu$ and the correlation of the rate fluctuation $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$, with $\mu = 30$, $\sigma = 10$, and $\tau = 0.1$. (Solid line): the 'theoretical' cost function computed directly from the underlying fluctuating rate using Eq. 3.25. B: the underlying fluctuating rate $\lambda_t$. C: event sequences derived from the rate. D: bar-graph histograms made using three types of bin sizes: too small, optimal, and too large.

demonstrating the superiority of the line-graph histogram to the bar-graph histogram in the sense of the MISE. In addition, the line-graph histogram is suitable for the comparison of multiple rates, as is the case for filtering methods.

## 4.2 Extrapolation of the bin size

The optimal bin size for the number of experimental trials $m$ can be estimated from the currently available data of size $n$ by minimizing the extrapolated cost function in the Algorithm 2. In this method, the cost function of $m$ sequences is obtained by modifying the original cost function, $C_n(\Delta)$, computed from the event count statistics of $n$ sequences of events.

### 4.2.1 Divergence of the optimal bin size

For a small number of sampled data, the estimated optimal bin size may diverge, implying that by constructing a histogram, it is likely that one obtains spurious results for the estimation (Koyama & Shinomoto, 2004). Because a shortage of data underlies this divergence, one can carry out more experiments to obtain a reliable estimation. With the Algorithm 2, one can estimate how many more experimental trials should be performed in order to construct a meaningful bar-graph histogram.

We applied this extrapolation method for a bar-graph histogram to a set of event sequences derived from the smoothly regulated Poisson process. Figure 4.3A depicts the extrapolated cost function $C_m(\Delta|n)$ for several values of $m$, computed from a given set of $n(=30)$ event sequences. Figure 4.3B represents the dependence of an inverse optimal bin size $1/\Delta_m^*$ on $1/m$. The inverse optimal bin size, $1/\Delta_m^*$, stays near 0 for $1/m > 1/\hat{n}_c$, and departs from 0 linearly with $m$ for $1/m \leq 1/\hat{n}_c$. By fitting the linear function, we estimated the critical number of sequences $\hat{n}_c$. In Figure 4.3C, the critical number of sequences $\hat{n}_c$ estimated from a smaller or larger $n$ is compared to the theoretical value

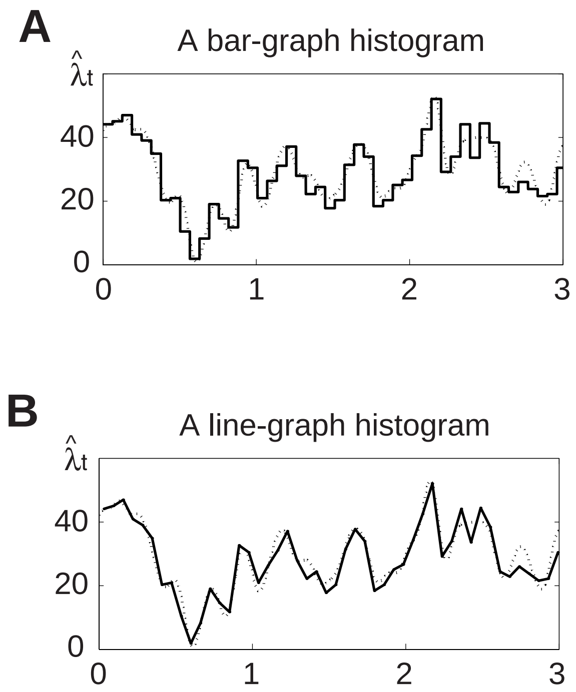


Figure 4.2: The comparison of the optimal bar-graph histogram and the optimal line-graph histogram. The data of the point events are the same as the data used in Fig. 4.1. A: the optimal bar-graph histogram constructed from the event sequences derived from an underlying fluctuating rate process (dashed line). B: the optimal line-graph histogram constructed from the same set of event sequences.

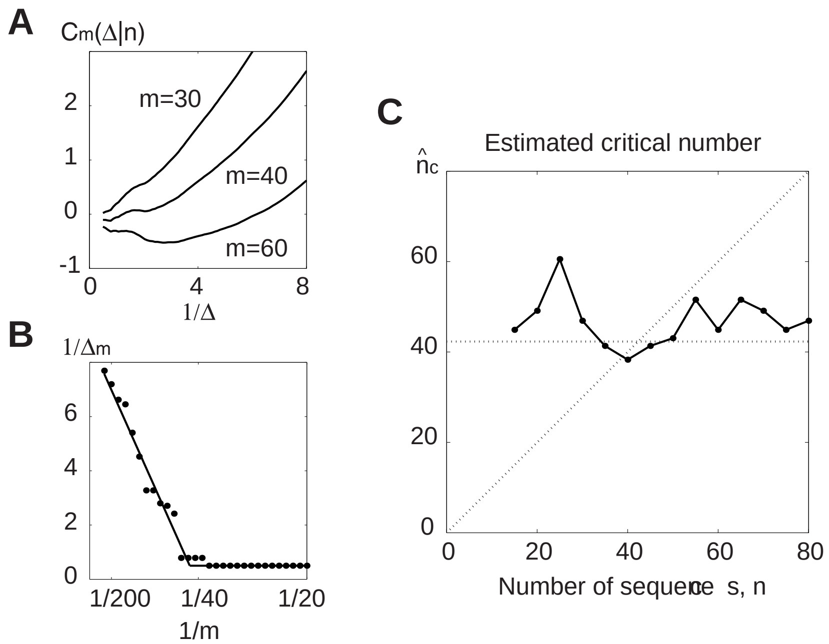


Figure 4.3: The extrapolation method. A: extrapolated cost functions $C_m(\Delta|n)$ of a bar-graph histogram for several values of $m$, computed from 30 sequences of events derived from a Poisson process characterized by the mean rate $\mu$ and the correlation of the rate fluctuation $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$, with $\mu = 30$, $\sigma = 2$, and $\tau = 0.1$. The modestly fluctuating rate with $\sigma = 2$ is used as compared with the rate with $\sigma = 10$ used in Fig. 1. B: the inverse optimal bin size $1/\Delta_m^*$ plotted against $1/m$. A solid line is a linear function fitted to the data. C: the critical number of sequences $\hat{n}_c$ extrapolated from a smaller or larger number of event sequences. The horizontal axis represents the number of event sequences $n$ used to obtain the extrapolated cost function $C_m(\Delta|n)$. The vertical axis represents an extrapolated critical number $\hat{n}_c$. The dashed lines represent a theoretical value $n_c$ computed using Eq. 3.27, and a diagonal.

of $n_c$ computed from Eq. 3.27. It is noteworthy that the estimated $\hat{n}_c$ approximates the theoretical $n_c$ well even from a fairly small number of event sequences, with which the estimated optimal bin size diverges.

### 4.2.2 Scaling of the optimal bin size

With the extrapolation method for a bar-graph histogram (Algorithm 2), one can also estimate how much the optimal bin size decreases (*i.e.*, the resolution increases) with the number of event sequences.

Figure 4.2.2A shows log-log plots of the optimal bin sizes $\Delta_m^*$ versus $m$ with respect to two rate processes that fluctuate either smoothly or in a zigzag pattern. These plots exhibit power-law scaling relations with distinctly different scaling exponents. The estimated exponents ($-0.34 \pm 0.04$ for the smooth rate process and $-0.56 \pm 0.04$ for the zigzag rate process) are close to the exponents of $m^{-1/3}$ and $m^{-1/2}$ that were obtained analytically as in Eqs. 3.32 and 3.33. In this way, we can estimate the degree of smoothness of the underlying rate from a reasonable amount of data.

With the extrapolation method for a line-graph histogram (Algorithm 4 in Appendix B.3), the scaling relations for a line-graph histogram can be examined in a similar manner. Figure 4.2.2B represents the optimal bin sizes computed for two rate processes that either fluctuate smoothly or in a zigzag pattern. The estimated exponents are $-0.24 \pm 0.04$ for the smooth rate process and $-0.50 \pm 0.05$ for the zigzag rate process. The exponents obtained by the extrapolation method are similar to the analytically obtained exponents, $m^{-1/5}$ and $m^{-1/2}$ respectively (see Eqs. B.15 and B.16). Note that for the smoothly fluctuating rate process, the scaling relation for the line-graph histogram is $m^{-1/5}$, whereas the scaling relation for a bar-graph histogram is $m^{-1/3}$. In contrast, for the rate process that fluctuates jaggedly, the exponents of the scaling relations for both a bar-graph histogram and a line-graph histogram are $m^{-1/2}$.

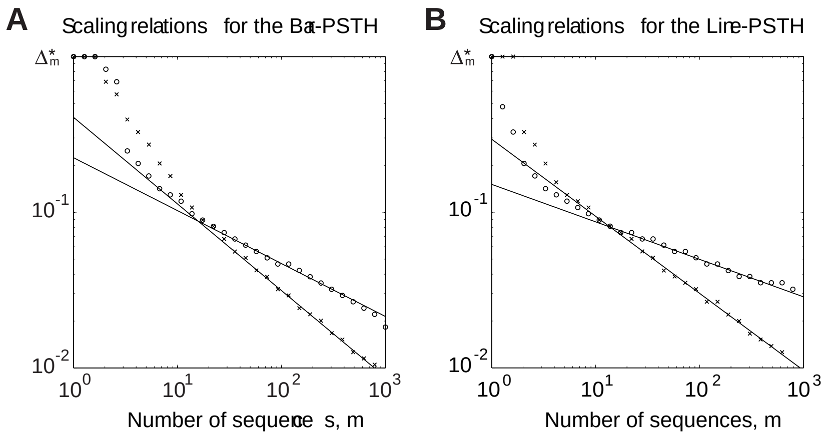


Figure 4.4: Scaling relations for the bar-graph histogram and the line-graph histogram. The optimal bin sizes $\Delta_m^*$ are estimated from the extrapolated cost function $C_m(\Delta|n)$, computed from 100 sequences of spikes. The optimal bin sizes $\Delta_m^*$ versus $m$ is shown in log-log scale. Two types of rate processes were examined; one that fluctuates smoothly, which is characterized by the correlation of the rate fluctuation $\phi(t) = \sigma^2 e^{-t^2/\tau^2}$, and the other that fluctuates jaggedly, which is characterized by the correlation of the rate fluctuation $\phi(t) = \sigma^2 e^{-|t|/\tau}$. The parameter values are the same as the ones used in Fig. 4.1. A: log-log plots of the optimal bin size with respect to $m$ for a bar-graph histogram. Lines are fitted to the data in an interval of $m \in [50, 500]$, whose regression coefficients, $-0.34 \pm 0.04$ and $-0.56 \pm 0.04$, correspond to the scaling exponents for a bar-graph histogram. B: log-log plots of the optimal bin size with respect to $m$ for a line-graph histogram. Lines are fitted to the data in an interval of $m \in [50, 500]$, whose regression coefficients, $-0.24 \pm 0.04$ and $-0.50 \pm 0.05$, correspond to the scaling exponents for a line-graph histogram.

# Appendix A

# A d-Dimensional Histogram

In this Appendix, we generalize the method for selecting the bin size of a histogram developed in Chapter 3. In Appendix A.1, we re-derive the cost function so that the formula is applicable to selecting the bin size of a histogram of $d$-dimension, where $d$ is positive integer. In addition, the cost function derived in this appendix is generally applicable to a histogram that depicts the function related to point events. In Appendix A.2, we explicitly derive the cost function for a histogram for density estimation.

## A.1 Derivation of the cost function

For a histogram of a multivariate datasets $\mathbf{x}$, MISE is defined by multiple integral:

$$\text{MISE} = \frac{1}{|V|} \int_V E\left\{\hat{f}(\mathbf{x}) - f(\mathbf{x})\right\}^2 d\mathbf{x}. \tag{A.1}$$

where $V$ is the $d$-dimensional data space of interest and $|V|$ is its volume. The average value of a function $f(\mathbf{x})$ in the $i$th bin is given by

$$\bar{f}_i = \frac{1}{\Delta_i} \int_{V_i} f(\mathbf{x})\, d\mathbf{x}.$$

where $V_i$ represents the space of $i$th bin, and $\Delta_i$ is the volume of the $i$th bin ($\Delta_i = |V_i|$). As is in one dimensional histogram, we divided the multiple integral into the $N$ segmented bins:

$$\begin{aligned}\text{MISE} &= \frac{1}{|V|}\int_V E_f\left\{\hat{f}(\mathbf{x}) - f(\mathbf{x})\right\}^2 d\mathbf{x} \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\Delta_i}\int_{V_i} E_{f_i}\left\{\hat{f}_i - f(\mathbf{x})\right\}^2 d\mathbf{x}.\end{aligned} \tag{A.2}$$

By denoting the average over bins with a bracket, we rewrite MISE as

$$\text{MISE} = \left\langle \frac{1}{\Delta_i}\int_{V_i} E_{f_i}\left\{\hat{f}_i - f(\mathbf{x})\right\}^2 d\mathbf{x}\right\rangle. \tag{A.3}$$

By inserting the ideal model $\bar{f}_i$, the MISE is decomposed into two parts,

$$\begin{aligned}\text{MISE} &= \left\langle \frac{1}{\Delta_i}\int_{V_i} E_{f_i}\left\{\hat{f}_i - \bar{f}_i + \bar{f}_i - f(\mathbf{x})\right\}^2 d\mathbf{x}\right\rangle \\ &= \left\langle E_{f_i}\left(\hat{f}_i - \bar{f}_i\right)^2\right\rangle + \left\langle \frac{1}{\Delta_i}\int_{V_i}\left\{f(\mathbf{x}) - \bar{f}_i\right\}^2 d\mathbf{x}\right\rangle.\end{aligned} \tag{A.4}$$

By inserting the total average $\mu = \frac{1}{V}\int_V f(\mathbf{x})d\mathbf{x}$, the second term can be further decomposed into two,

$$\begin{aligned}\left\langle \frac{1}{\Delta_i}\int_{V_i}\left\{f(\mathbf{x}) - \bar{f}_i\right\}^2 d\mathbf{x}\right\rangle &= \left\langle \frac{1}{\Delta_i}\int_{V_i}\left\{f(\mathbf{x}) - \mu\right\}^2 d\mathbf{x}\right\rangle - \left\langle \frac{1}{\Delta_i}\int_{V_i}\left\{\bar{f}_i - \mu\right\}^2 d\mathbf{x}\right\rangle \\ &= \frac{1}{|V|}\int_V\left\{f(\mathbf{x}) - \mu\right\}^2 d\mathbf{x} - \left\langle\left(\bar{f}_i - \mu\right)^2\right\rangle.\end{aligned} \tag{A.5}$$

where the first term is irrespective of the choice of the bin size. We thus introduce the cost function,

$$\begin{aligned}C'(\{\Delta_i\}) &= \text{MISE} - \frac{1}{|V|}\int_V\left\{f(\mathbf{x}) - \mu\right\}^2 d\mathbf{x} \\ &= \left\langle E_{f_i}(\hat{f}_i - \bar{f}_i)^2\right\rangle - \left\langle\left(\bar{f}_i - \mu\right)^2\right\rangle.\end{aligned} \tag{A.6}$$

From the decomposition rule, we have

$$\left\langle E_{f_i}(\hat{f}_i - \left\langle E_{f_i}\hat{f}_i\right\rangle)^2\right\rangle = \left\langle E_{f_i}(\hat{f}_i - \bar{f}_i)^2\right\rangle + \left\langle E_{f_i}\left(\bar{f}_i - \langle\bar{f}_i\rangle\right)^2\right\rangle. \quad \text{(A.7)}$$

By substituting this equation, the cost function becomes

$$C\left(\{\Delta_i\}\right) = 2\left\langle E_{f_i}\left(\hat{f}_i - \bar{f}\right)^2\right\rangle - E_f\left\langle\left(\hat{f}_i - \left\langle E\hat{f}_i\right\rangle\right)^2\right\rangle. \quad \text{(A.8)}$$

The second term is written as

$$E_f\left\langle\left(\hat{f}_i - \left\langle E_{f_i}\hat{f}_i\right\rangle\right)^2\right\rangle = E_f\left\langle\left(\hat{f}_i - \left\langle\hat{f}_i\right\rangle\right)^2\right\rangle + E_f\left\langle\left(\left\langle\hat{f}_i\right\rangle - \left\langle E_{f_i}\hat{f}_i\right\rangle\right)^2\right\rangle. \quad \text{(A.9)}$$

The second term of the above equation does not depend on the choice of the bin size $\Delta$. We add this term to the cost function, and redefine the cost function as

$$\begin{aligned} C\left(\{\Delta_i\}\right) &= C'\left(\{\Delta_i\}\right) + E_f\left\langle\left(\left\langle\hat{f}_i\right\rangle - \left\langle E_{f_i}\hat{f}_i\right\rangle\right)^2\right\rangle \\ &= 2\left\langle E_{f_i}\left(\hat{f}_i - \bar{f}\right)^2\right\rangle - E_f\left\langle\left(\hat{f}_i - \left\langle\hat{f}_i\right\rangle\right)^2\right\rangle. \end{aligned} \quad \text{(A.10)}$$

Note that throughout the derivation of the cost function, we only assumed that each height of the histogram is unbiased estimator of the average of the underlying function in the bin. The cost function given by A.10 is applicable to not only obtaining the bin size of histogram for a density (or rate) estimation, but also obtaining the bin size of a histogram that depicts the function related to a point process, such as an auto-correlation function of the underlying rate. When the experimental trials are repeated to obtain the data, the first term can be evaluated as the variance of the histograms constructed from individual trials.

## A.2 Selection of the bin size for a Poisson process

For a rate estimation of a Poisson point process, the bar-height is defined as

$$\hat{f}_i = \hat{\theta}_i = \frac{k_i}{n\Delta_i}, \tag{A.11}$$

where $n$ is the number of repeated trials. The number of events $k_i$ in the volume $V_i$ follows a Poisson distribution (Diggle, 1983) with mean,

$$n\int_{V_i} f(\mathbf{x})\, d\mathbf{x} = n\Delta f_i. \tag{A.12}$$

Since the mean and variance of the Poisson point process equals, we have

$$\left\langle E\left(\hat{f}_i - \bar{f}_i\right)^2 \right\rangle = \left\langle \frac{1}{n\Delta_i} E\hat{f}_i \right\rangle. \tag{A.13}$$

The cost function is then written as

$$C_n(\{\Delta_i\}) = \frac{2}{n} E_f \left\langle \frac{1}{\Delta_i}\hat{f}_i \right\rangle - E_f \left\langle \left(\hat{f}_i - \left\langle \hat{f}_i \right\rangle\right)^2 \right\rangle. \tag{A.14}$$

By replacing the expectation by one sample histogram, we obtain

$$C_n(\{\Delta_i\}) = \frac{2}{n^2}\left\langle \frac{1}{\Delta_i^2} k_i \right\rangle - \frac{1}{n^2}\left\langle \left(\frac{k_i}{\Delta_i} - \left\langle \frac{k_i}{\Delta_i} \right\rangle\right)^2 \right\rangle.$$

For a regular histogram, whose volume is equal in all bins, we have

$$C_n(\Delta) = \frac{1}{n^2\Delta^2}\left\{2\langle k_i\rangle - \left\langle (k_i - \langle k_i\rangle)^2 \right\rangle\right\}.$$

# Appendix B

# A Line-Graph Histogram

In this Appendix, we provide a thoery and a recipe for i) selecting the bin size of a line-graph histogram and ii) estimating required number of experimental trials to construct a line-graph histogram.

## B.1 Construction of a line-graph histogram

A line graph can be constructed by simply connecting top-centers of adjacent bar graphs of the height $k_i/(n\Delta)$. Figure B.1 schematically shows the construction of a line-graph histogram. For the same set of event sequences, the optimal bin size (window size) of a line-graph histogram is, however, different from that of a bar-graph histogram. We develop here a method for selecting the bin size for a line-graph histogram.

The expected heights of adjacent bar graphs for intervals of $[-\Delta, 0]$ and $[0, \Delta]$ are

$$\theta^- \equiv \frac{1}{\Delta}\int_{-\Delta}^{0} \lambda_t \, dt,$$
$$\theta^+ \equiv \frac{1}{\Delta}\int_{0}^{\Delta} \lambda_t \, dt.$$

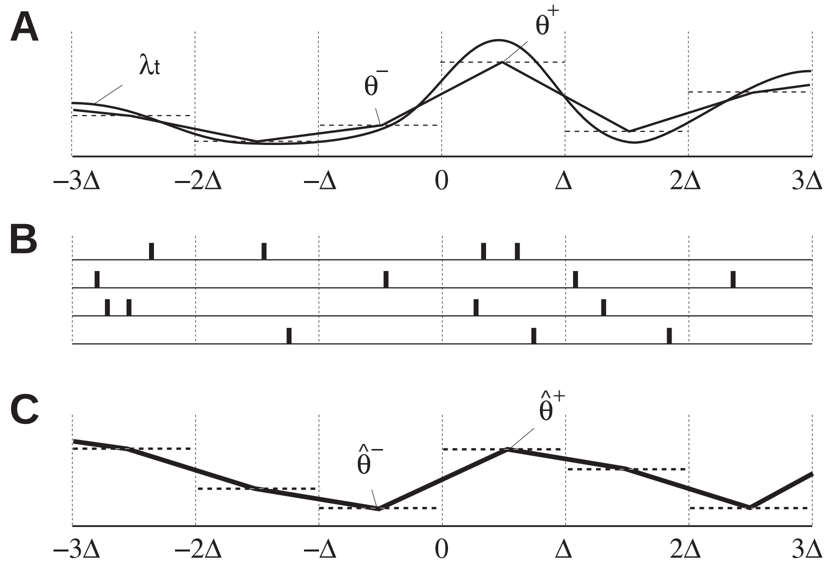


Figure B.1: The line-graph histogram. A: an underlying rate, $\lambda_t$. The horizontal bars indicate the original bar graphs $\theta^-$ and $\theta^+$ for adjacent intervals $[-\Delta, 0]$ and $[0, \Delta]$. The expected line graph $L_t$ is a line connecting the two top-centers of adjacent expected bar graphs $(-\frac{\Delta}{2}, \theta^-)$ and $(\frac{\Delta}{2}, \theta^+)$. B: sequences of events derived from the underlying rate. C: a line-graph histogram for the sample sequences of events. The empirical line graph $\hat{L}_t$ is a line connecting the two top-centers of adjacent empirical bar graphs $(-\frac{\Delta}{2}, \hat{\theta}^-)$ and $(\frac{\Delta}{2}, \hat{\theta}^+)$.

The expected line graph $L_t$ in an interval of $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$ is a line connecting top-centers of these bar graphs,

$$L_t = \frac{\theta^+ + \theta^-}{2} + \frac{\theta^+ - \theta^-}{\Delta} t. \tag{B.1}$$

The unbiased estimators of the original bar graphs are $\hat{\theta}^- = k^-/(n\Delta)$ and $\hat{\theta}^+ = k^+/(n\Delta)$, where $k^-$ and $k^+$ are the numbers of events from $n$ event sequences that enter the intervals $[-\Delta, 0]$ and $[0, \Delta]$, respectively. The empirical line graph $\hat{L}_t$ in an interval of $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$ is a line connecting two top-centers of adjacent empirical bar graphs,

$$\hat{L}_t = \frac{\hat{\theta}^+ + \hat{\theta}^-}{2} + \frac{\hat{\theta}^+ - \hat{\theta}^-}{\Delta} t. \tag{B.2}$$

## B.2 Selection of the bin size

The time average of the MISE (Eq. 3.3) is rewritten by the average over the $N$ segmented rates,

$$\mathrm{MISE} = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \left\langle E\,(\hat{L}_t - \lambda_t)^2 \right\rangle dt. \tag{B.3}$$

The MISE can be decomposed into two parts,

$$\mathrm{MISE} = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \left\langle E(\hat{L}_t - L_t)^2 \right\rangle dt + \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \left\langle (\lambda_t - L_t)^2 \right\rangle dt. \tag{B.4}$$

The first term of Eq. B.4 is the stochastic fluctuation of the empirical linear estimator $\hat{L}_t$ due to the stochastic point events, which can be computed as

$$\begin{aligned} &\frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \left\langle E(\hat{L}_t - L_t)^2 \right\rangle dt \\ &= \frac{1}{3} \left\langle E(\hat{\theta}^+ - \theta^+)^2 \right\rangle + \frac{1}{3} \left\langle E(\hat{\theta}^- - \theta^-)^2 \right\rangle. \end{aligned} \tag{B.5}$$

Here we have used the relations, $E(\hat{\theta}^+ - \theta^+)(\hat{\theta}^- - \theta^-) = 0$.

The second term of Eq. B.4 is the temporal fluctuation of $\lambda_t$ around the expected linear estimator $L_t$. We expands the second term by inserting $\mu = \langle \theta^- \rangle = \langle \theta^+ \rangle = \langle \theta^0 \rangle$, and obtain

$$\frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle (\lambda_t - L_t)^2 \right\rangle dt = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle (\lambda_t - \mu)^2 \right\rangle dt \tag{B.6}$$

$$- \frac{2}{\Delta}\int_{-\Delta/2}^{\Delta/2} \langle (\lambda_t - \mu)\,(L_t - \mu) \rangle \, dt \tag{B.7}$$

$$+ \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \left\langle (L_t - \mu)^2 \right\rangle dt \tag{B.8}$$

The second term B.7 is computed as

$$\begin{aligned}(B.7) = {} & \left\langle \left(\theta^+ - \mu\right)\left(\theta^0 - \mu\right) \right\rangle + \left\langle \left(\theta^- - \mu\right)\left(\theta^0 - \mu\right) \right\rangle \\ & + \left\langle \left(\theta^+ - \mu\right)\theta^* \right\rangle - \left\langle \left(\theta^- - \mu\right)\theta^* \right\rangle ,\end{aligned}$$

where

$$\begin{aligned}\theta^0 &\equiv \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \lambda_t \, dt, \\ \theta^* &\equiv \frac{2}{\Delta^2}\int_{-\Delta/2}^{\Delta/2} t\lambda_t \, dt.\end{aligned}$$

To obtain the above equation, we have used relations, $\langle \theta^- \rangle = \langle \theta^+ \rangle = \langle \theta^0 \rangle = \mu$. The third term B.8 can be computed as

$$(B.8) = \frac{1}{3}\left\langle \left(\theta^+ - \mu\right)^2 \right\rangle + \frac{1}{3}\left\langle \left(\theta^+ - \mu\right)\left(\theta^- - \mu\right) \right\rangle + \frac{1}{3}\left\langle \left(\theta^- - \mu\right)^2 \right\rangle .$$

The first term B.6 of the systematic error represents a mean squared fluctuation of the underlying rate, and is independent of the choice of the bin size $\Delta$ (See Eq. 3.8). We introduce the cost function by subtracting the variance of the underlying rate from the original MISE:

$$
\begin{aligned}
C_n'(\Delta) &\equiv \mathrm{MISE} - \frac{1}{T}\int (\lambda_t - \mu)^2\, dt \\
&= \frac{1}{3}\left\langle E(\hat\theta^+ - \theta^+)^2 \right\rangle + \frac{1}{3}\left\langle E(\hat\theta^- - \theta^-)^2 \right\rangle \\
&\quad - \left\langle \left(\theta^+ - \mu\right)\left(\theta^0 - \mu\right)\right\rangle - \left\langle \left(\theta^- - \mu\right)\left(\theta^0 - \mu\right)\right\rangle \\
&\quad - \left\langle \left(\theta^+ - \mu\right)\theta^* \right\rangle + \left\langle \left(\theta^- - \mu\right)\theta^* \right\rangle \\
&\quad + \frac{1}{3}\left\langle \left(\theta^+ - \mu\right)^2 \right\rangle + \frac{1}{3}\left\langle \left(\theta^- - \mu\right)^2 \right\rangle \\
&\quad + \frac{1}{3}\left\langle \left(\theta^+ - \mu\right)\left(\theta^- - \mu\right)\right\rangle . \qquad \text{(B.9)}
\end{aligned}
$$

By using the relations, $\left\langle E(\hat\theta^+ - \theta^+)^2 \right\rangle = \left\langle E(\hat\theta^- - \theta^-)^2 \right\rangle$, $\langle(\theta^+ - \mu)^2\rangle = \langle(\theta^- - \mu)^2\rangle$, $\langle(\theta^+ - \mu)\,(\theta^0 - \mu)\rangle = \langle(\theta^- - \mu)\,(\theta^0 - \mu)\rangle$, and $\langle(\theta^+ - \mu)\,\theta^*\rangle = -\langle(\theta^- - \mu)\,\theta^*\rangle$, the cost function becomes

$$
\begin{aligned}
C_n(\Delta) &= \frac{2}{3}\left\langle E(\hat\theta^+ - \theta^+)^2 \right\rangle \\
&\quad - 2\left\langle \left(\theta^+ - \mu\right)\left(\theta^0 - \mu\right)\right\rangle - 2\left\langle \left(\theta^+ - \mu\right)\theta^* \right\rangle \\
&\quad + \frac{2}{3}\left\langle \left(\theta^+ - \mu\right)^2 \right\rangle + \frac{1}{3}\left\langle \left(\theta^+ - \mu\right)\left(\theta^- - \mu\right)\right\rangle . \qquad \text{(B.10)}
\end{aligned}
$$

The first term of Eq. B.10 can be estimated from the data, using the variance-mean relation (Eq. 3.14). The last four terms of Eq. B.10 are the covariances of the expected rates $\theta^-, \theta^+, \theta^0$, and $\theta^*$, which are not observables. We can estimate them by using the covariance decomposition rule for unbiased estimators:

$$
\begin{aligned}
&\left\langle E\left(\hat\theta^+ - \left\langle E\hat\theta^+ \right\rangle\right)\left(\hat\theta^p - \left\langle E\hat\theta^p \right\rangle\right)\right\rangle \\
&= \left\langle E\left[\left(\hat\theta^+ - \theta^+\right)\left(\hat\theta^p - \theta^p\right)\right]\right\rangle + \left\langle (\theta^+ - \langle\theta^+\rangle)(\theta^p - \langle\theta^p\rangle)\right\rangle . \qquad \text{(B.11)}
\end{aligned}
$$

where $p$ denotes $-, +, 0$, or $*$. Unlike the bar-graph histogram, however, the mean-variance relation for the Poisson statistics is not directly applicable to the second term. We suggest estimating these covariance terms from multiple sample sequences. The r.h.s of Eq. B.11 is decomposed into

$$E\left\langle\left(\hat{\theta}^{+}-\left\langle E\hat{\theta}^{+}\right\rangle\right)\left(\hat{\theta}^{p}-\left\langle E\hat{\theta}^{p}\right\rangle\right)\right\rangle$$
$$=E\left\langle\left(\hat{\theta}^{+}-\left\langle\hat{\theta}^{+}\right\rangle\right)\left(\hat{\theta}^{p}-\left\langle\hat{\theta}^{p}\right\rangle\right)\right\rangle+E\Big(\left\langle\hat{\theta}^{+}\right\rangle-\left\langle\theta^{+}\right\rangle\Big)\Big(\left\langle\hat{\theta}^{p}\right\rangle-\left\langle\theta^{p}\right\rangle\Big).$$

The second term of the r.h.s. of this equation is not dependent on the bin size selection. Therefore, we redefine the cost function appropriately as

$$C_n(\Delta)=C_n'(\Delta)+2E\Big(\left\langle\hat{\theta}^{+}\right\rangle-\left\langle\theta^{+}\right\rangle\Big)\Big(\left\langle\hat{\theta}^{0}\right\rangle-\left\langle\theta^{0}\right\rangle\Big)$$
$$+2E\Big(\left\langle\hat{\theta}^{+}\right\rangle-\left\langle\theta^{+}\right\rangle\Big)\Big(\left\langle\hat{\theta}^{*}\right\rangle-\left\langle\theta^{*}\right\rangle\Big)-\frac{2}{3}E\Big(\left\langle\hat{\theta}^{+}\right\rangle-\left\langle\theta^{+}\right\rangle\Big)^{2}$$

This cost function can be estimated by following a recipe summarized as 'Algorithm 3'.

## B.3 Extrapolation of the bin size

As in the case of the bar-graph histogram the cost function for any $m$ sequences of event sequences can be extrapolated using the variance-mean relation for the Poisson statistics,

$$C_m(\Delta|n)=\frac{2}{3\Delta}\left(\frac{1}{m}-\frac{1}{n}\right)\left\langle E\hat{\theta}_n^{+}\right\rangle+C_n(\Delta), \tag{B.12}$$

where $C_n(\Delta)$ is the original cost function (Eq. B.10). With the original cost function $C_n(\Delta)$, we can easily estimate a cost function for $m$ sequences with 'Algorithm 4'.

## B.4 Theoretical cost function

We can obtain a 'theoretical' cost function of a line-graph histogram directly from the mean rate $\mu$ and the correlation $\phi(t)$ of the rate fluctuation. According to the mean-variance relation based on the Cramér-Rao (in)equality (Eq. 3.24), the cost function (Eq. B.10) is given by

$$C_n(\Delta) = \frac{2\mu}{3n\Delta} - \frac{2}{\Delta^2}\int_0^{\Delta}\int_{-\Delta/2}^{\Delta/2}\left(1+\frac{2t_2}{\Delta}\right)\phi(t_1-t_2)\,dt_1dt_2 + \frac{2}{3\Delta^2}\int_0^{\Delta}\int_0^{\Delta}\phi(t_1-t_2)\,dt_1dt_2 + \frac{1}{3\Delta^2}\int_0^{\Delta}\int_{-\Delta}^{0}\phi(t_1-t_2)\,dt_1dt_2. \tag{B.13}$$

## Scaling relations of the optimal bin sizes

By expanding the correlation of the rate fluctuation,

$$\phi(t) = \phi(0)+\phi'(0_+)|t|+\frac{1}{2}\phi''(0)t^2+\frac{1}{6}\phi'''(0_+)|t|^3+\frac{1}{24}\phi''''(0)t^4+O\left(|t|^5\right),$$

we obtain

$$C_n(\Delta) = \frac{2\mu}{3n\Delta}-\phi(0)-\frac{37}{144}\phi'(0_+)\Delta+\frac{181}{5760}\phi'''(0_+)\Delta^3+\frac{49}{2880}\phi''''(0)\Delta^4+O(\Delta^5). \tag{B.14}$$

Unlike the bar-graph histogram, the line graph successfully approximates the original rate to first order in $\Delta$, and therefore the $O(\Delta^2)$ term in the cost function vanishes for a line-graph histogram.

The optimal bin size $\Delta^*$ is obtained from $dC_n(\Delta)/d\Delta = 0$. For a rate process that fluctuates smoothly in time, the correlation function is a smooth function, resulting in $\phi'(0) = 0$ and $\phi'''(0) = 0$ due to the symmetry $\phi(t) = \phi(-t)$, and we obtain the scaling relation

$$\Delta^* \sim \left(\frac{1280\mu}{181\phi''''(0)\,n}\right)^{1/5}. \tag{B.15}$$

The exponent $-1/5$ for a line-graph histogram is different than the exponent $-1/3$ for a bar-graph histogram (Eq. 3.32).

If the correlation of rate fluctuation has a cusp at $t = 0$ ($\phi'(0_+) < 0$), we obtain the scaling relation

$$\Delta^* \sim \left(-\frac{96\mu}{37\phi'(0_+)n}\right)^{1/2}, \tag{B.16}$$

by ignoring the higher order terms. The exponent $-1/2$ for a line-graph histogram is the same as the exponent for a bar-graph histogram (Eq. 3.33).

**Algorithm 3: Bin size selection for a line-graph histogram**

---

(i) Divide the observation period $T$ into $N+1$ bins of width $\Delta$.
Count the number of events $k_i(j)$ that enter $i$th bin from $j$th sequence.
Define $k_i^{(-)}(j) \equiv k_i(j)$ and $k_i^{(+)}(j) \equiv k_{i+1}(j)$ $(i = 1, 2, \cdots, N)$.
Divide the period $[\Delta/2, T - \Delta/2]$ into $N$ bins of width $\Delta$.
Count the number of events $k_i^{(0)}(j)$ that enter $i$th bin from $j$th sequence.
In each bin, compute $k_i^{(*)}(j) \equiv 2 \sum_\ell t_i^\ell(j)/\Delta$, where $t_i^\ell(j)$ is the time of the $\ell$th event that enters the $i$th bin, measured from the center of each bin.

(ii) Sum up $k_i^{(p)}(j)$ for all sequences: $k_i^{(p)} \equiv \sum_{j=1}^{n} k_i^{(p)}(j)$, where $p = \{-, +, 0, *\}$.
Average those event counts with respect to all the bins: $\bar{k}^{(p)} \equiv \frac{1}{N} \sum_{i=1}^{N} k_i^{(p)}$.
Compute covariances of $k_i^{(+)}$ and $k_i^{(p)}$:

$$c^{(+,p)} \equiv \frac{1}{N} \sum_{i=1}^{N} \left( k_i^{(+)} - \bar{k}^{(+)} \right) \left( k_i^{(p)} - \bar{k}^{(p)} \right).$$

Compute covariances of $k_i^{(+)}(j)$ and $k_i^{(p)}(j)$ with respect to sequences, and average over time:

$$\bar{c}^{(+,p)} \equiv \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n-1} \sum_{j=1}^{n} \left( k_i^{(+)}(j) - \frac{k_i^{(+)}}{n} \right) \left( k_i^{(p)}(j) - \frac{k_i^{(p)}}{n} \right).$$

**Algorithm 3: continued from page 46.**

Finally, compute

$$\sigma^{(+,p)} \equiv \frac{c^{(+,p)}}{(n\Delta)^2} - \frac{\bar{c}^{(+,p)}}{n\Delta^2}.$$

(iii) Compute the cost function,

$$C_n(\Delta) = \frac{2}{3}\frac{\bar{k}^{(+)}}{(n\Delta)^2} - 2\sigma^{(+,0)} - 2\sigma^{(+,*)} + \frac{2}{3}\sigma^{(+,+)} + \frac{1}{3}\sigma^{(+,-)}.$$

(iv) Repeat i through iii while changing $\Delta$. Find $\Delta^*$ that minimizes $C_n(\Delta)$.

---

**Algorithm 4: Extrapolation method for a line-graph histogram**

---

A Construct the extrapolated cost function,

$$C_m(\Delta|n) = \frac{2}{3}\left(\frac{1}{m} - \frac{1}{n}\right)\frac{\bar{k}^{(+)}}{n\Delta^2} + C_n(\Delta),$$

where $C_n(\Delta)$ is the cost function for the line-graph histogram computed for $n$ trials with the Algorithm 3.

B Search for $\Delta_m^*$ that minimizes $C_m(\Delta|n)$.

C A and B while changing $m$, and plot $1/\Delta_m^*$ vs $1/m$ to search for the critical value $1/m = 1/\hat{n}_c$ above which $1/\Delta_m^*$ practically vanishes.

---

# Appendix C

# A Density Estimation with Sample Size Constraint

In this Appendix, we derive cost functions of a bar- and a line-graph histogram when the data with predetermined size is sampled from a population. In Chapter 3, we investigated the histogram of the data that are derived from an inhomogeneous Poisson point process, where the sample size is not determined. Due to the assumption of iid random variables, histograms of the data obtained by the two different experimental designs can be equally seen as the estimation of the probability density function. However, the difference between the two experimental designs appears when one considers ensemble of histograms as in the MISE criterion.

The MISE is composed of the sampling and systematic errors as shown in Eq.3.6. A similar equation for a line-graph histogram was given by Eq.B.4. For the Poisson point process, the sampling error simply obeys the variance of the Poisson statistics. When a fixed number of samples are collected from a population, the sampling error becomes smaller than the variance of the Poisson statistics. As a result, the optimal bin size obtained under the sample size constraint is slightly smaller than that obtained under the Poisson point process framework.

## C.1 A bar-graph histogram

The bar height of a histogram for a density estimation is given by $\hat{\theta}_n = k_i/n\Delta$. In this context, $n$ should be read as the number of samples (it is not the number of experimental trials). We label the sequence of the samples by $j$ $(= 1, \cdots, n)$, and introduce $u_i(j)$. $u_i(j)$ is a unity if the $j$th sampled data is observed in the $i$th bin $(i = 1, \cdots, N)$ , and zero otherwise. Summation of $u_i(j)$ with respect to $j$ equals $k_i$:

$$k_i = \sum_{j=1}^{n} u_i(j). \tag{C.1}$$

The sampling error for the $i$th bin is given by

$$\frac{1}{n}\frac{1}{n-1}\sum_{j=1}^{n}\left\{\frac{u_i(j)}{\Delta} - \frac{1}{n}\sum_{j=1}^{n}\frac{u_i(j)}{\Delta}\right\}^2 \tag{C.2}$$

By using the notation, $\bar{u}_i = \frac{1}{n}\sum_{j=1}^{n} u_i(j) = k_i/n$, we obtain

$$\begin{aligned}\frac{1}{n\Delta^2}\frac{1}{n-1}\sum_{j=1}^{n}\left(u_i(j) - \bar{u}_i\right)^2 &= \frac{1}{n\Delta^2}\left\{\frac{1}{n-1}\sum_{j=1}^{n} u_i(j)^2 - \frac{n}{n-1}\bar{u}_i^2\right\} \\ &= \frac{1}{n\Delta^2}\left(\frac{1}{n-1}k_i - \frac{1}{n(n-1)}k_i^2\right)\end{aligned} \tag{C.3}$$

To obtain the last equality, we have used a relation,

$$k_i = \sum_{j=1}^{n} u_i(j) = \sum_{j=1}^{n} u_i(j)^2. \tag{C.4}$$

The last equality holds because $u_i(j) = \{0, 1\}$. This relation manifests the constraint on the total sample size.

In the derivation of the cost function Eq.3.13 in Section 3.2 in Chapter 3, we have only assumed that a histogram is an unbiased estimator of the underlying function. Therefore, the cost function in the form of Eq.3.13,

$$C_n(\Delta) = 2\left\langle E(\hat{\theta}-\theta)^2\right\rangle - E\left\langle(\hat{\theta}-\langle\hat{\theta}\rangle)^2\right\rangle. \qquad (3.13)$$

can be used to obtain an optimal bin size of a regular histogram in general. Hence, by using Eq.C.3 to estimate the sampling error term in Eq.3.13, we obtain

$$\begin{aligned} C_n^{FIX}(\Delta) &= \frac{2}{n\Delta^2}\left\{\frac{1}{n-1}\frac{1}{N}\sum_{i=1}^{N}k_i - \frac{1}{n(n-1)}\frac{1}{N}\sum_{i=1}^{N}k_i^2\right\} \\ &\quad - \frac{1}{N}\sum k_i^2 - \left(\frac{1}{N}\sum k_i\right)^2 \\ &= \frac{2}{(n-1)\Delta T} - \frac{n+1}{n^2(n-1)\Delta T}\sum_{i=1}^{N}k_i^2 - \frac{1}{T^2}. \end{aligned} \qquad (C.5)$$

This cost function is identified with Rudemo's estimator for his risk function (Rudemo, 1982).

The relations between the cost function of the samples derived from the Poisson point process and $C_n^{FIX}(\Delta)$ is given by

$$C_n(\Delta) = \frac{n-1}{n}C_n^{FIX}(\Delta) + \frac{1}{n^3\Delta T}\sum k_i^2 + \frac{1}{nT^2}. \qquad (C.6)$$

Due to the second term, the expected optimal bin size of a Poisson point process is slightly larger than that obtained under fixed sample sizes. This result arises from the fact that the histograms constructed from the data derived from a Poisson point process contains an additional fluctuation in its total sample size compared to the histograms constructed from the data whose size is fixed. However, since the difference is on the order of $O(n^{-3})$, the two cost functions yield indistinguishable optimal bin size in practice.

## C.2 A line-graph histogram

From Eq.B.9, the cost function for a line-graph histogram becomes

$$\begin{aligned}C_n(\Delta) = {} & \frac{1}{(n\Delta)^2}\left\{\frac{1}{3}c^{(+,+)} + \frac{1}{3}c^{(-,-)} + \frac{1}{3}c^{(+,-)}\right\} \\ & - \frac{1}{(n\Delta)^2}\left\{\left(c^{(+,0)} + c^{(-,0)}\right) + \left(c^{(+,*)} + c^{(-,*)}\right)\right\} \\ & + \frac{1}{n\Delta^2}\left\{\left(\bar{c}^{(+,0)} + \bar{c}^{(-,0)}\right) - \left(\bar{c}^{(+,*)} + \bar{c}^{(-,*)}\right)\right\}. \end{aligned} \tag{C.7}$$

The covariance with respect to the bins can be easily calculated as

$$c^{(p,q)} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(k_i^{(p)} - \bar{k}^{(p)}\right)\left(k_i^{(q)} - \bar{k}^{(q)}\right), \tag{C.8}$$

where $p, q = \{+, -, 0, *\}$ and $\bar{k}^{(p)} = \frac{1}{N}\sum_{i=1}^{N} k_i^{(p)}$. The last term in the cost function is obtained by using

$$\bar{c}^{(+,0)} + \bar{c}^{(-,0)} = \frac{1}{n-1}\frac{1}{N}\sum_{i=1}^{N} k_i^{(0)} - \frac{1}{n(n-1)}\frac{1}{N}\sum_{i=1}^{N}\left\{k_i^{(+)} + k_i^{(-)}\right\}k_i^{(0)}, \tag{C.9}$$

$$\bar{c}^{(+,*)} + \bar{c}^{(-,*)} = \frac{1}{n-1}\frac{1}{N}\sum_{i=1}^{N} k_i^{(*)} - \frac{1}{n(n-1)}\frac{1}{N}\sum_{i=1}^{N}\left\{k_i^{(+)} + k_i^{(-)}\right\}k_i^{(*)}. \tag{C.10}$$

These equations are derived from the argument below under the constraint on the total sample size. We denote the covariance with respect to the sampled data in the$i$th bin by $\bar{c}_i^{(p,q)}$, which is defined as

$$\bar{c}_i^{(p,q)} \equiv \frac{1}{n-1}\sum_{j=1}^{n}\left(k_i^{(p)}(j) - \frac{k_i^{(p)}}{n}\right)\left(k_i^{(q)}(j) - \frac{k_i^{(q)}}{n}\right). \tag{C.11}$$

The covariance averaged over the bins is given by $\bar{c}^{(p,\,q)} = \frac{1}{N}\sum_{i=1}^{N}\bar{c}_i^{(p,\,q)}$, where $\bar{c}_i^{(p,\,q)}$ is given by

$$\begin{aligned}\bar{c}_i^{(p,q)} &= \frac{1}{n-1}\sum_{j=1}^{n}\left(u_i^{(p)}(j) - \bar{u}_i^{(p)}\right)\left(u_i^{(q)}(j) - \bar{u}_i^{(q)}\right) \\ &= \frac{1}{n-1}\sum_{j=1}^{n} u_i^{(p)}(j)u_i^{(q)}(j) - \frac{n}{n-1}\bar{u}_i^{(p)}\bar{u}_i^{(q)}. \end{aligned} \tag{C.12}$$

Since the $j$th sample that enters the $i$th bin locates in either a left side $[-\Delta/2, 0]$ or a right side $[0, \Delta/2]$ of the bin, we have

$$\sum_{j=1}^{n} u_i^{(+)}(j)u_i^{(0)}(j) + \sum_{j=1}^{n} u_i^{(-)}(j)u_i^{(0)}(j) = k_i^{(0)}, \tag{C.13}$$

$$\sum_{j=1}^{n} u_i^{(+)}(j)u_i^{(*)}(j) + \sum_{j=1}^{n} u_i^{(-)}(j)u_i^{(*)}(j) = k_i^{(*)}. \tag{C.14}$$

These equations hold because $u_i^{(p)}(j) = \{0, 1\}$. Note that

$$\bar{u}_i^{(p)}\bar{u}_i^{(q)} = k_i^{(p)}k_i^{(q)}/n^2. \tag{C.15}$$

By using these equations we obtain

$$\bar{c}_i^{(+,\,0)} + \bar{c}_i^{(-,\,0)} = \frac{1}{n-1}k_i^{(0)} - \frac{1}{n\,(n-1)}\left\{k_i^{(+)}k_i^{(0)} + k_i^{(-)}k_i^{(0)}\right\}, \tag{C.16}$$

$$\bar{c}_i^{(+,\,*)} + \bar{c}_i^{(-,\,*)} = \frac{1}{n-1}k_i^{(*)} - \frac{1}{n\,(n-1)}\left\{k_i^{(+)}k_i^{(*)} + k_i^{(-)}k_i^{(*)}\right\}. \tag{C.17}$$

# References

Blahut, R. E. (1987). *Principles and practice of information theory.* Reading, Mass: Addison-Wesley.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory.* New York: John Wiley & Sons, Inc.

Cox, R. D. (1962). *Renewal theory.* London: John Wiley & Sons Inc.

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

Devroye, L., & Györfi, L. (1985). *Nonparametric density estimation : the l[1] view.* New York: Wiley.

Diggle, P. J. (1983). *Statistical analysis of spatial point patterns.* London: Academic Press Inc.

Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator - l2 theory. *Zeitschrift Fur Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, *57*(4), 453-476.

Gardiner, C. (1985). *Handbook of stochastic methods: for physics, chemistry and the naural sciences.* Berlin: Springer-Verlag.

Hall, P. (1990). Akaikes information criterion and kullback-leibler loss for histogram density-estimation. *Probability Theory and Related Fields*, *85*(4), 449-467.

Izenman, A. J. (1991). Recent developments in nonparametric density-estimation. *Journal of the American Statistical Association*, *86*(413), 205-224.

Kanazawa, Y. (1993). Hellinger distance and akaike information criterion for the histogram. *Statistics & Probability Letters*, *17*(4), 293-298.

Koyama, S., & Shinomoto, S. (2004). Histogram bin width selection for time-dependent poisson processes. *Journal of Physics A-Mathematical and General*, *37*(29), 7255-7265.

Kubo, M., R. Toda, & Hashitsume, N. (1985). *Nonequilibrium statistical mechanics.* Berlin: Springer-Verlag.

Révész, P. (1968). *The laws of large numbers.* New York: Academic Press.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, *9*(2), 65-78.

Scott, D. W. (1979). Optimal and data-based histograms. *Biometrika*, *66*(3), 605-610.

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

Shimazaki, H., & Shinomoto, S. (2007b). A recipe for optimizing a time-histogram. In B. Schölkopf, J. C. Platt, & T. Hofmann (Eds.), *Advances in neural information processing systems 19.* Cambridge, MA: MIT Press. (in Press)

Snyder, D. (1975). *Random point processes.* New York: John Wiley & Sons, Inc.

Taylor, C. C. (1987). Akaike information criterion and the histogram. *Biometrika*, *74*(3), 636-639.

Wand, M. P. (1997). Data-based choice of histogram bin width. *American Statistician*, *51*(1), 59-64.