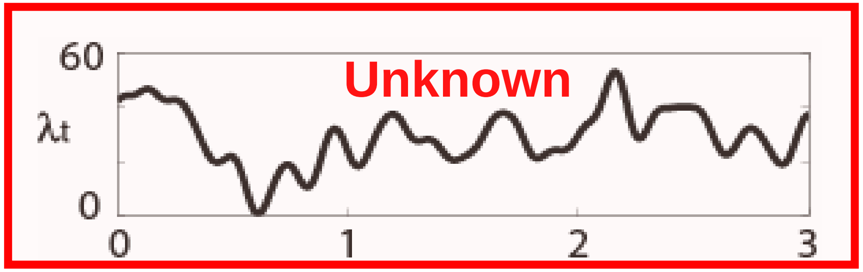
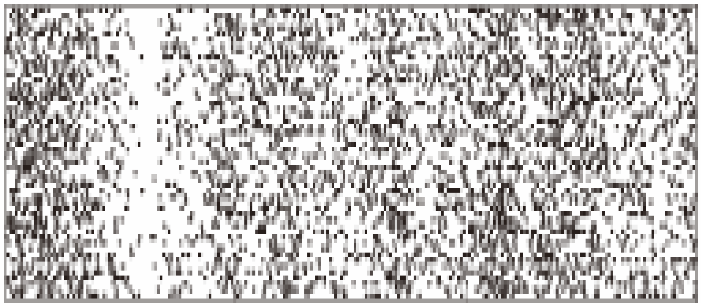
Abstract

The time-histogram method is a handy tool for capturing the instantaneous rate of spike occurrence. In most of the neurophysiological literature, the bin size that critically determines the goodness of the fit of the time-histogram to the underlying rate has been selected by individual researchers in an unsystematic manner. We propose an objective method for selecting the bin size of a time-histogram from the spike data, so that the resulting time-histogram best approximates the unknown underlying rate.

The resolution of the histogram increases, or the optimal bin size decreases, with the number of spike sequences sampled. It is notable that the optimal bin size diverges if only a small number of experimental trials are available from a moderately fluctuating rate process. In this case, any attempt at characterizing the underlying spike rate will lead to spurious results. Given a paucity of data, our method can also suggest how many more trials are needed until the set of data can be analyzed with the required resolution.

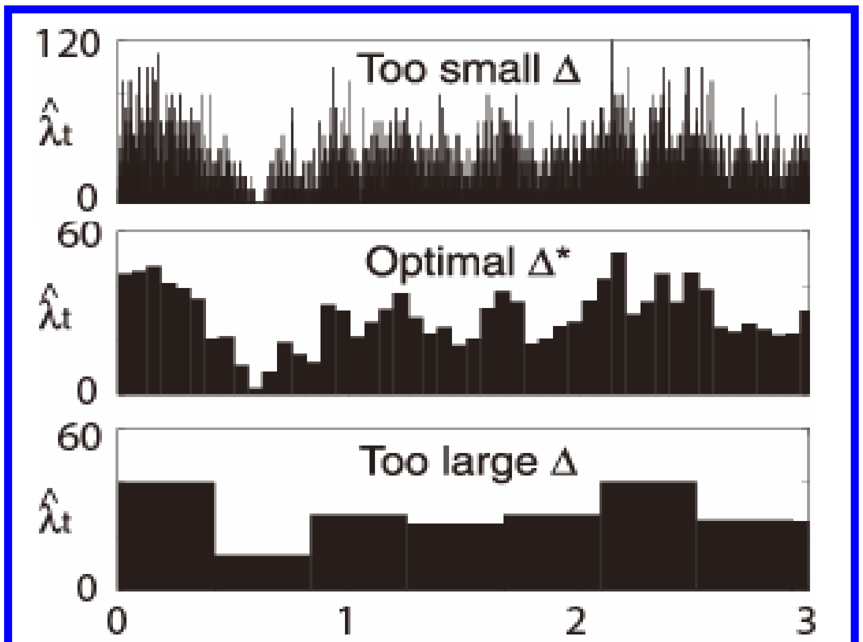
Optimizing a time-histogram

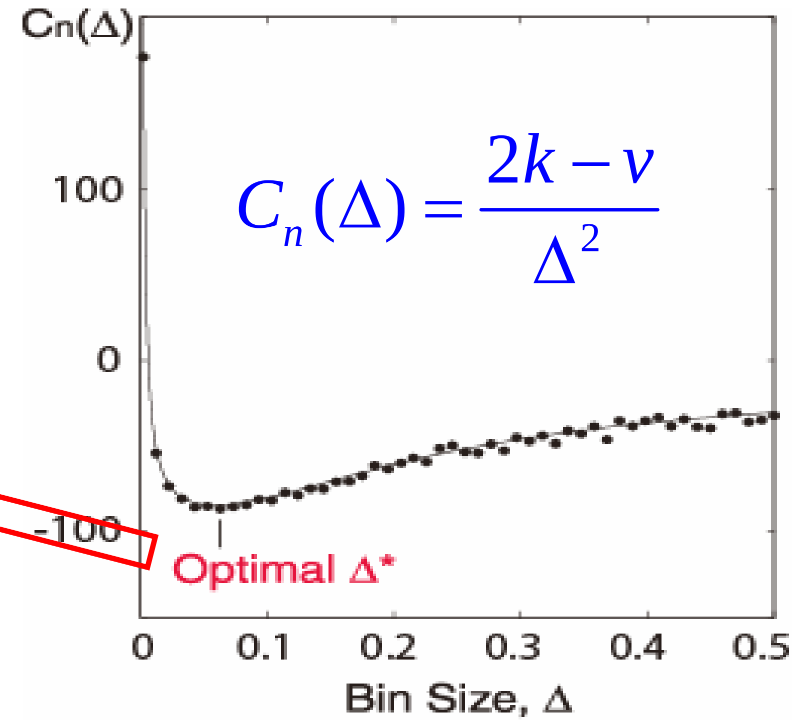




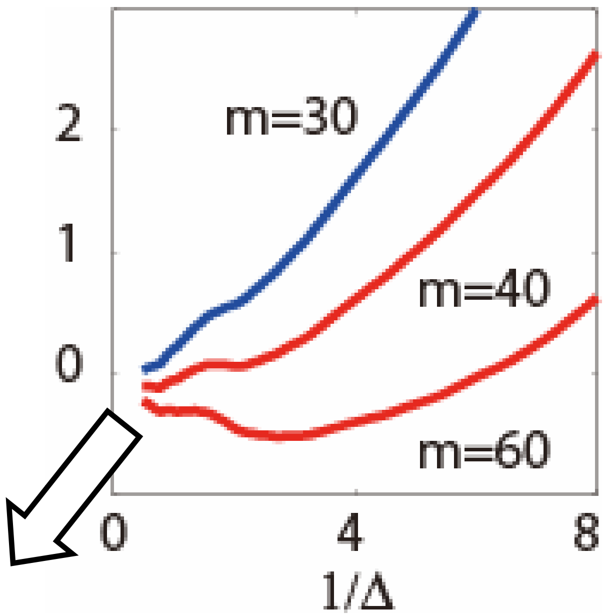
$$\mathrm{MISE} = \int \left(\boxed{\lambda_t} - \boxed{\hat{\lambda}_t} \right)^2 dt$$

Underlying Rate **Histogram**





How many trials are required to make a Histogram?



Original: $C_n(\Delta)$

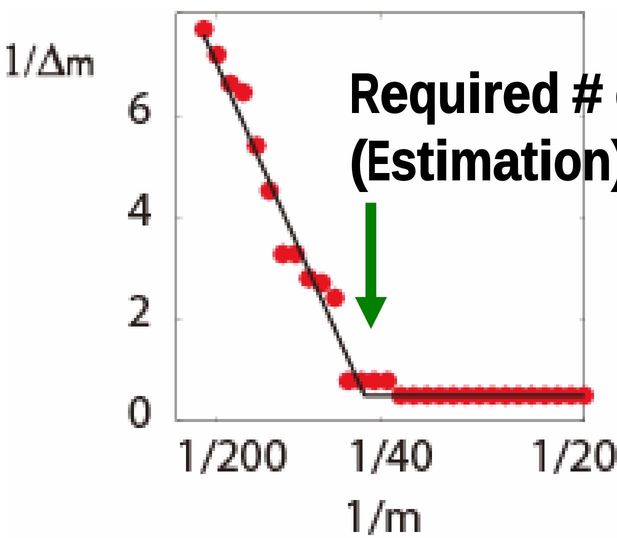
Optimal bin size diverges

Extrapolated:

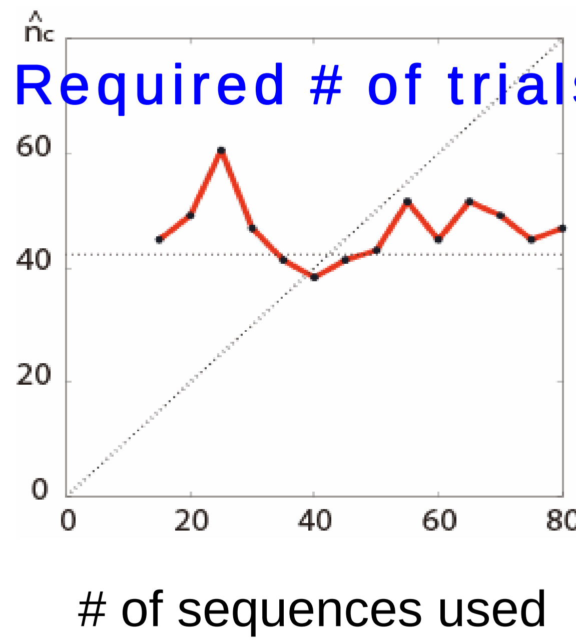
$$C_m(\Delta \mid n) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{k}{n\Delta^2} + C_n(\Delta)$$

Finite optimal bin size

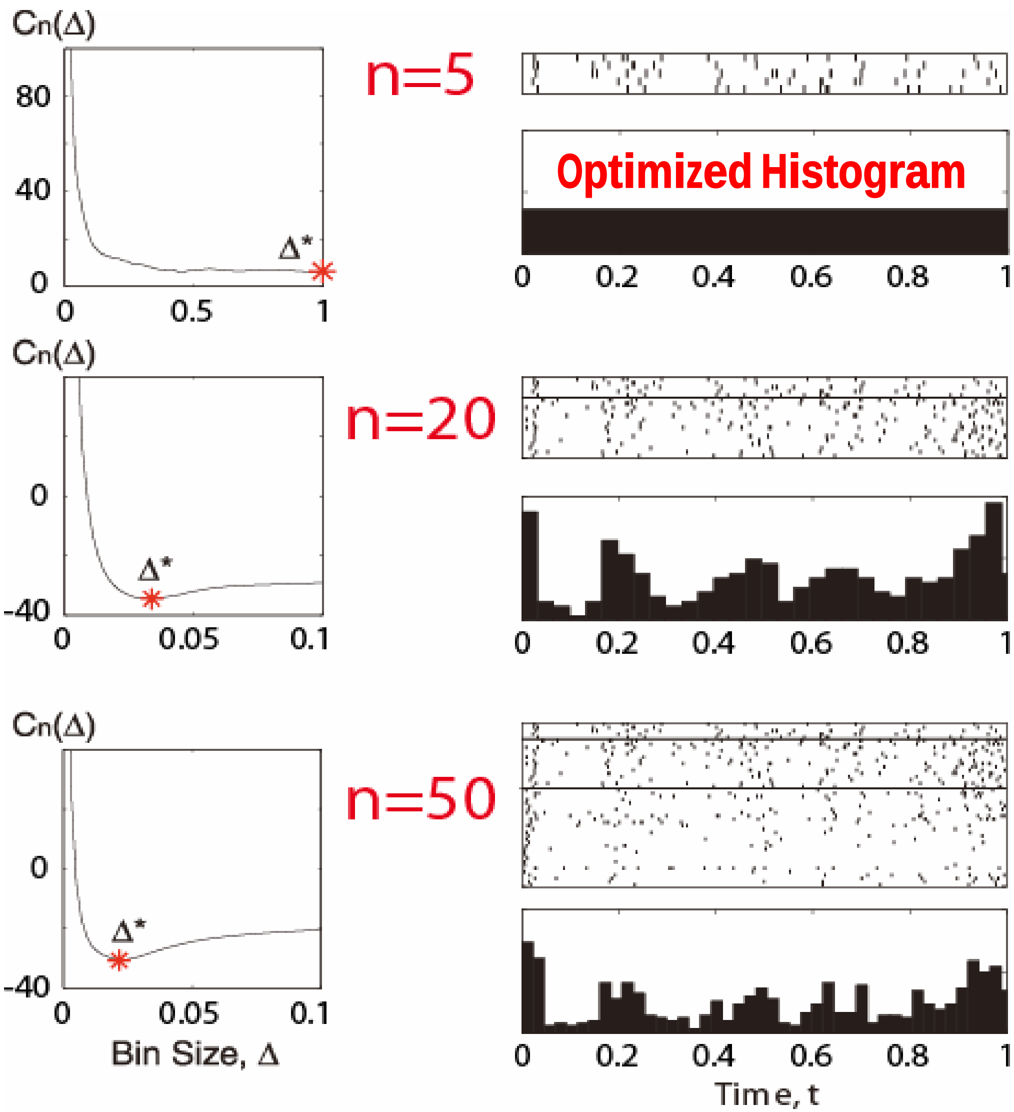
Optimal bin size v.s. m



Required # of trials



Rate modulation of an MT neuron



Too few to make a Histogram !

Extrapolation

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

Estimation:

At least 12 trials are required.

Data : Britten et al. (2004) neural signal archive

Methods

Method I. Selection of the Bin Size

(i) Divide spike sequences with length *T* [s] into *N* bins of width Δ.

(ii) Calculate the mean and variance of the number of spikes.

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

(iii) Compute the cost function

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

(iv) Repeat i through iii while changing the bin size Δ. Find Δ* that minimize the cost function.

Method II. Extrapolation of the Bin Size

(A) Construct the extrapolated cost function,

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

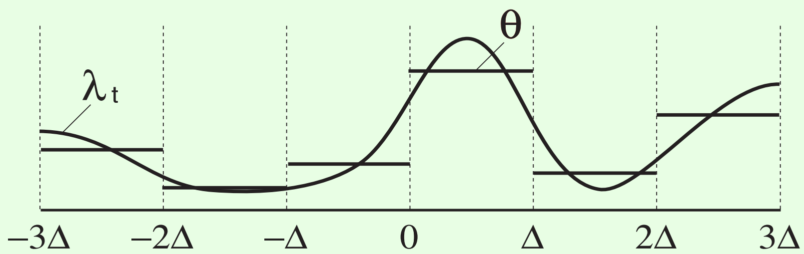
using the sample mean k of the number of spikes obtained from *n* sequences of spikes. $C_n(\Delta)$ is the cost function computed for *n* sequences of spikes.

(B) Search for Δ_m that minimizes the extrapolated cost function .

(C) Repeat A and B while changing , and plot 1=1/ Δ_m vs. 1/*m* to search for the critical value 1/*m*=1/n_c above which 1/ Δ_m practically vanishes.

Theory

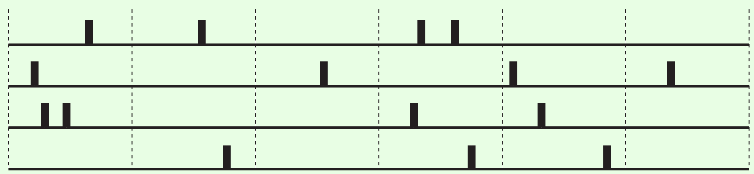
Time-Varying Rate



The mean underlying rate in an interval [0, Δ]:

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

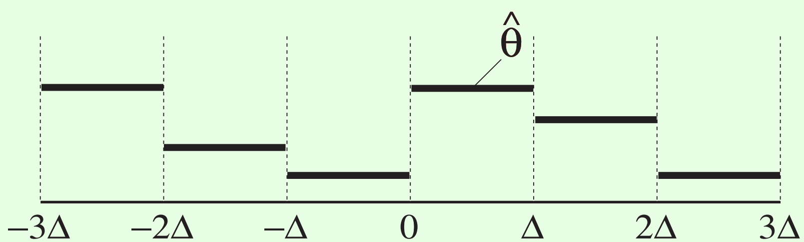
Spike Sequences



The spike count in the bin obeys the Poisson distribution*:

$$p(k \mid n\Delta\theta) = \frac{(n\Delta\theta)^k}{k!} e^{-n\Delta\theta}.$$

Time Histogram



A histogram bar-height is an estimator of θ:

$$\hat{\theta} = \frac{k}{n\Delta}$$

*When the spikes are obtained by repeating an independent trial, the accumulated data obeys the Poisson point process due to a general limit theorem.

Method I. Selection of the Bin Size

$$\mathrm{MISE} \equiv \frac{1}{T}\int_0^T E(\hat{\lambda}_t - \lambda_t)^2 dt = \left\langle \frac{1}{\Delta}\int_0^\Delta E(\underbrace{\hat{\theta}_n - \lambda_t}_{\theta - \theta})^2 dt \right\rangle.$$

Expectation by the Poisson statistics, given the rate.

Average over segmented bins.

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

Sampling Error   Systematic Error

Decomposition of the Systematic Error

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

Variance of the rate   Independent of Δ Variance of an ideal histogram

Introduction of the cost function:

$$C_n(\Delta) \equiv \mathrm{MISE} - \frac{1}{T}\int_0^T \left(\lambda_t - \langle\theta\rangle\right)^2 dt$$

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

Sampling error Unknown: Variance of ideal histogram

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

Variance of a histogram Sampling error

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

The Poisson statistics obeys: $E(\hat{\theta}_n - \theta)^2 = \frac{1}{n\Delta} E\hat{\theta}_n.$

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

Mean of a Histogram Variance of a Histogram

Method II. Extrapolation of the Bin Size

The cost function for m experimental trials:

$$C_m(\Delta) = \left\langle E(\hat{\theta}_m - \theta)^2 \right\rangle - \left\langle \left(\theta - \langle\theta\rangle\right)^2 \right\rangle$$

Unknown

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

A histogram constructed from n trials

The extrapolated cost function:

$$C_m(\Delta \mid n) = \left\langle E(\hat{\theta}_m - \theta)^2 \right\rangle + \left\langle E(\hat{\theta}_n - \theta)^2 \right\rangle - \left\langle E\left(\hat{\theta}_n - \left\langle E\hat{\theta}_n \right\rangle\right)^2 \right\rangle$$

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

The Poisson statistics obeys: $$\left\langle E(\hat{\theta}_m - \theta)^2 \right\rangle = \frac{1}{m\Delta} E\hat{\theta}_m = \frac{1}{m\Delta} E\hat{\theta}_n$$

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

Reference

A Method for Selecting the Bin Size of a Time Histogram

Hideaki Shimazaki and Shigeru Shinomoto

Neural Computation in Press

- **Web Application for the Bin Size Selection**
- **Matlab / Mathematica / R sample codes**

are available at our homepage

http://www.ton.scphys.kyoto-u.ac.jp/~shino//~hideaki/

See also

Koyama, S. and Shinomoto, S. Histogram bin width selection for time-dependent poisson processes. *J. Phys. A*, 37(29):7255–7265. 2004