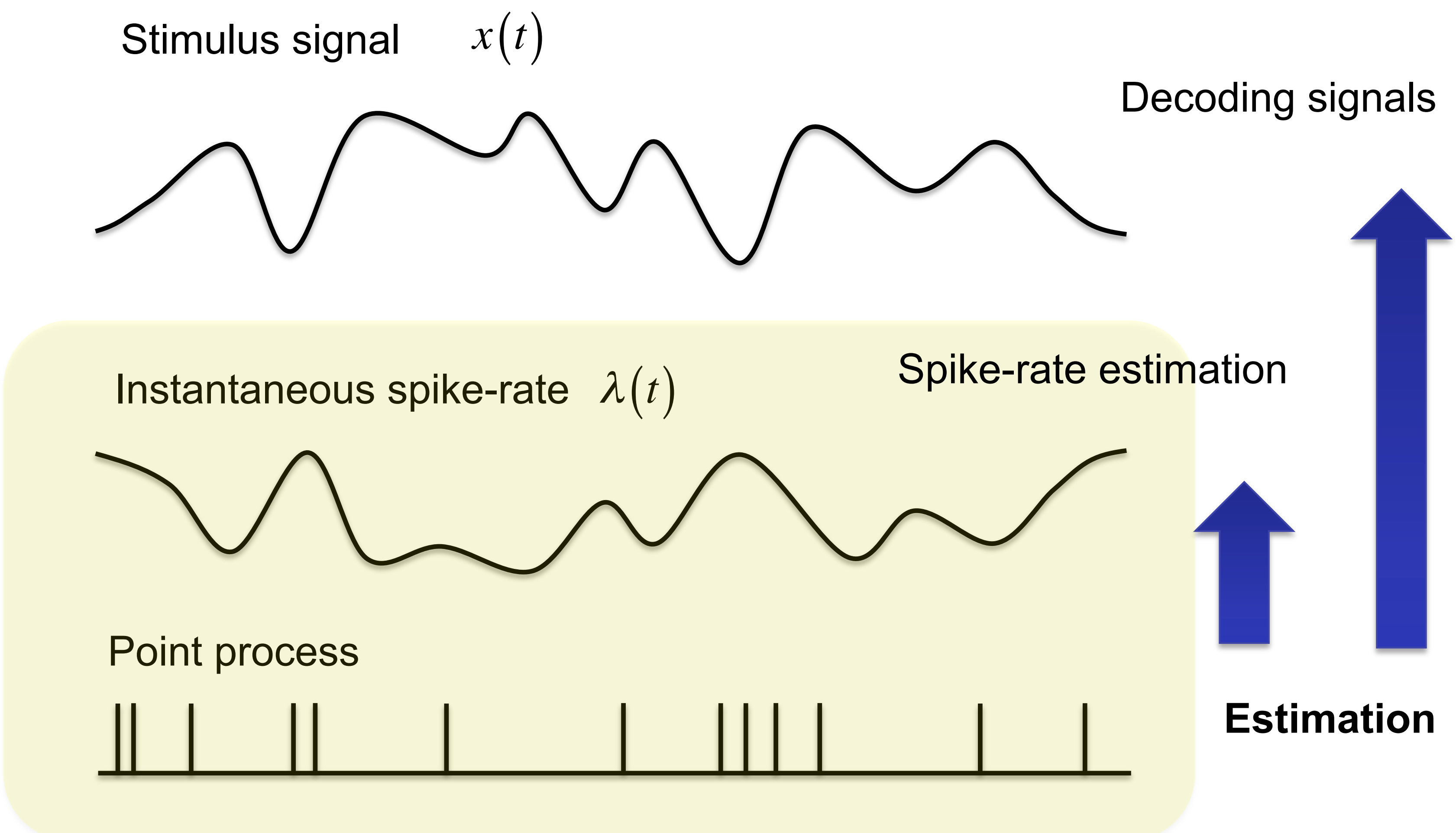
Inference for an inhomogeneous Poisson process

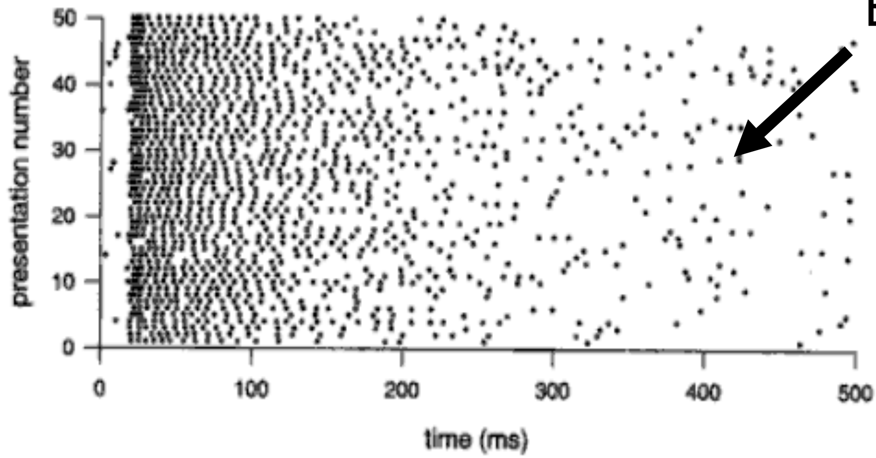
SPIKE-RATE ESTIMATION

Inference problems



Peri-stimulus time histogram (PSTH)

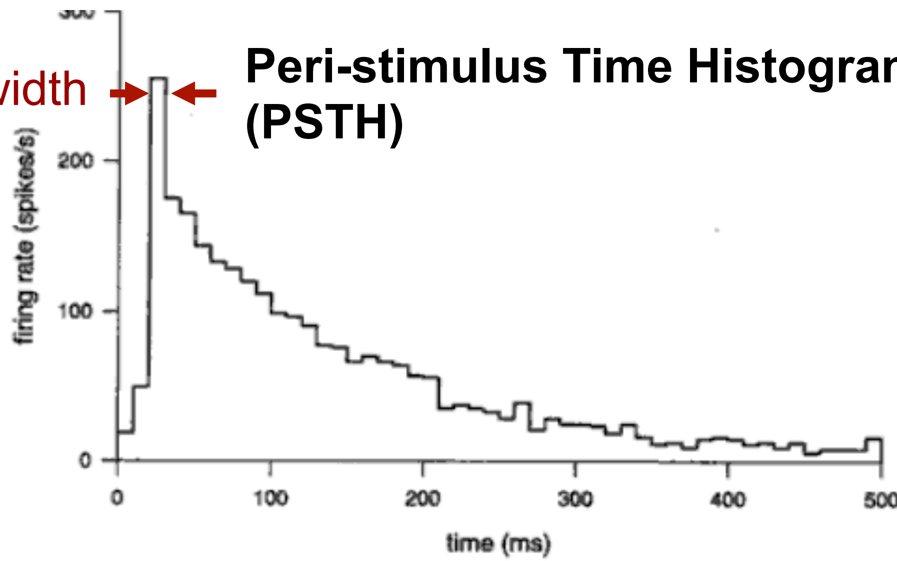
Event (Spike)



Repeated trials

Bin-width

Peri-stimulus Time Histogram (PSTH)



From *Spikes: Exploring the Neural Code*, Rieke et al. 1997

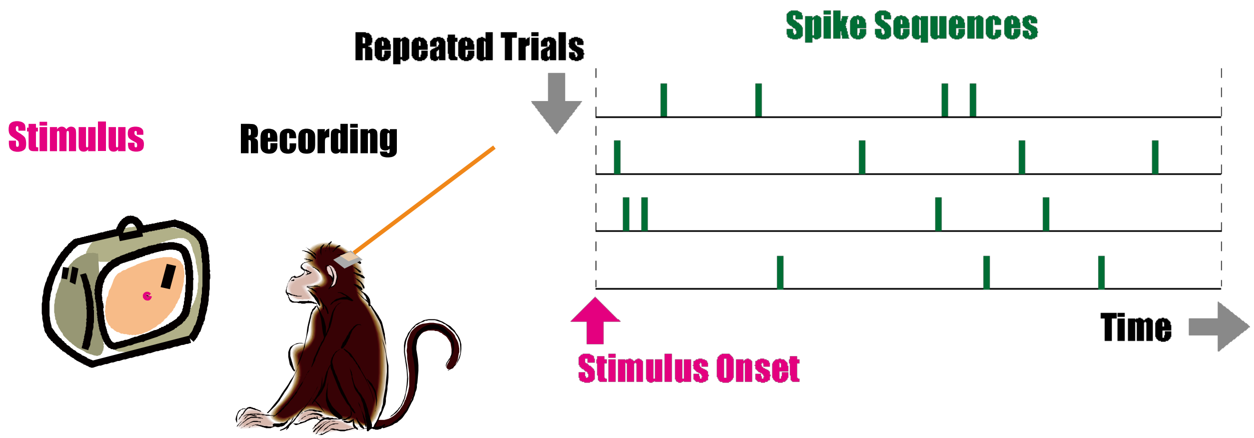
Inventors of the PSTH

Adrian, E. (1928). *The basis of sensation: The action of the sense organs*.

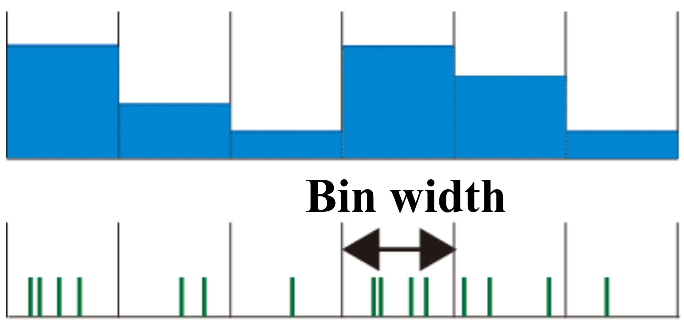
George Gerstein and Nelson Kiang (1960) An Approach to the Quantitative Analysis of Electrophysiological Data from Single Neurons, Biophys J. 1(1): 15–28.

Prof. Gerstein

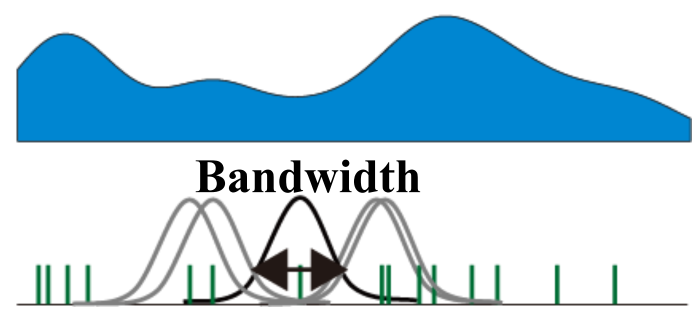
Spike-rate estimation



Histogram Method



Kernel Method



Choice of the bin/band width is critical in firing rate estimation.

Bin-width optimization

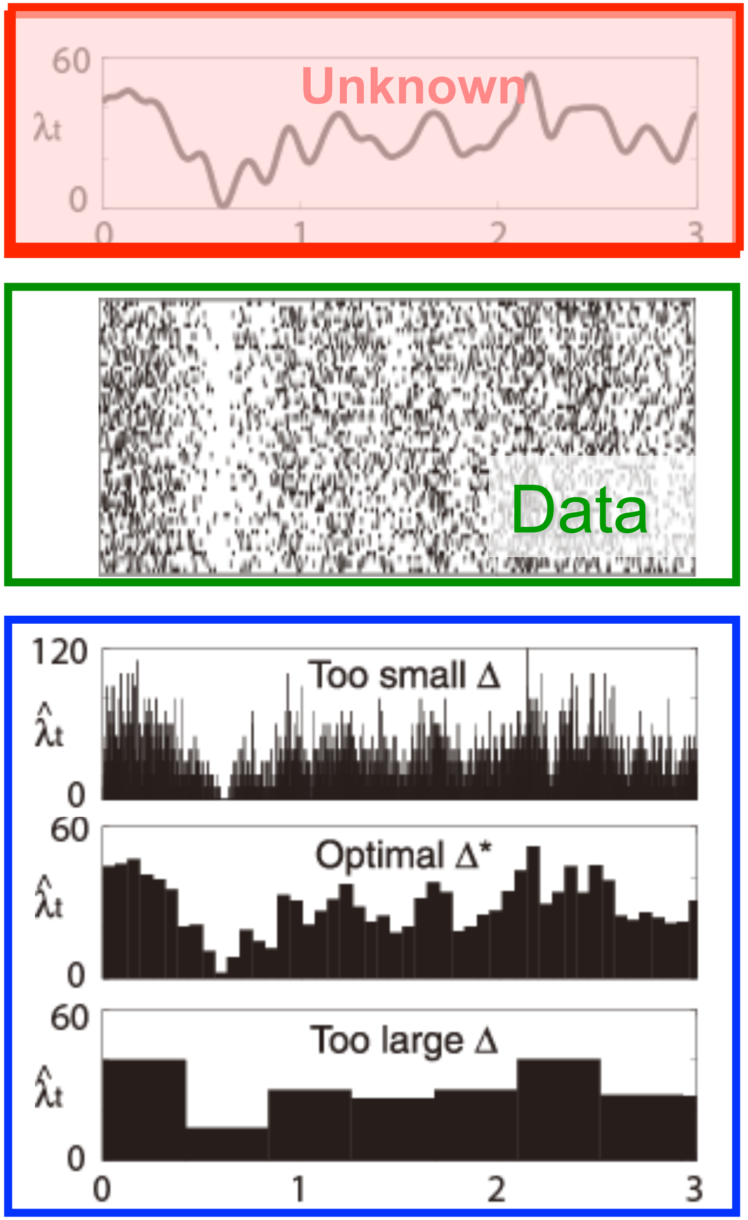
Shimazaki & Shinomoto
Neural Computation, 2007

Bandwidth optimization

Shimazaki & Shinomoto
J. Comput. Neurosci, 2009

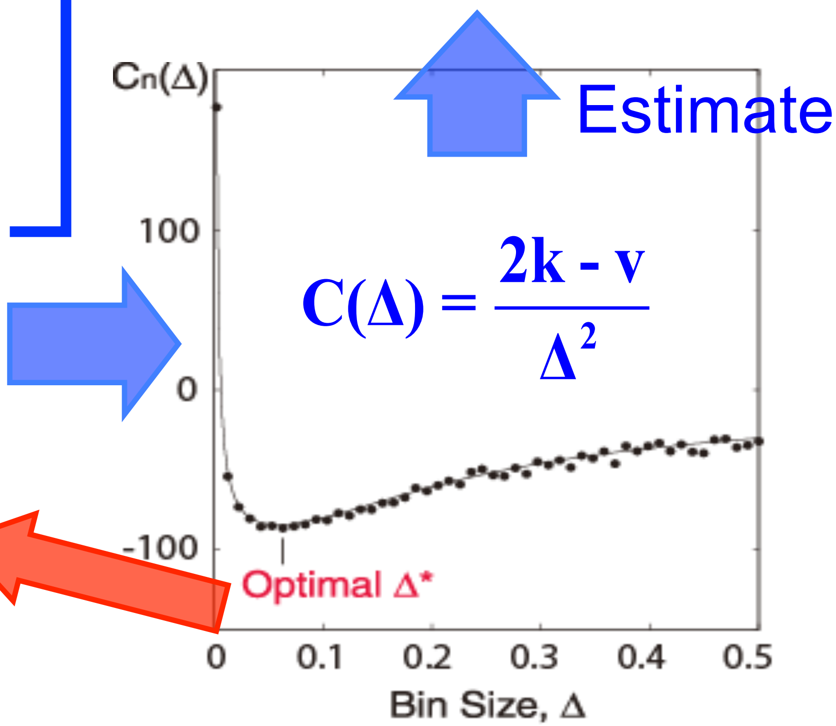
HISTOGRAM OPTIMIZATION

Histogram optimization



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

Underlying Rate Histogram



$$C(\Delta) = \frac{2k - v}{\Delta^2}$$

Method for Selecting the Bin Size

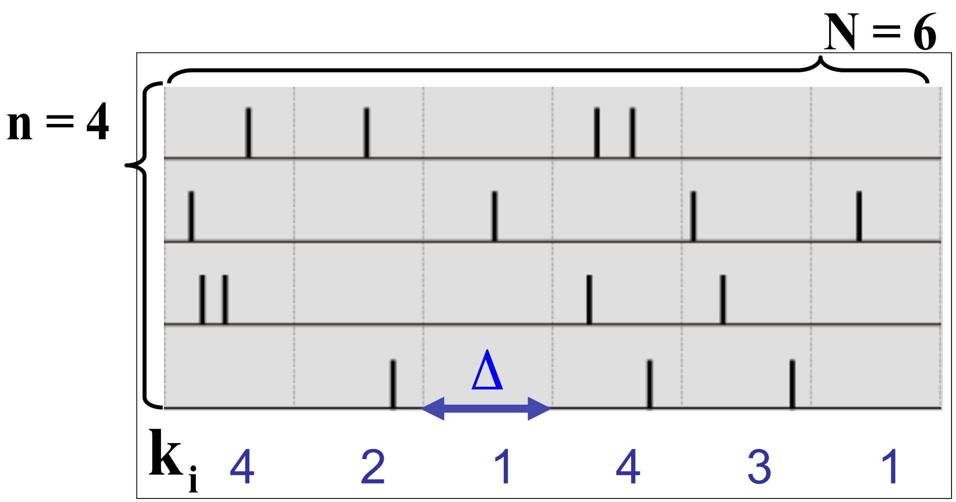
- Divide the data range into *N* bins of width *Δ*.
  Count the number of events k_i in the *i* th bin.

- Compute the cost function

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

  while changing the bin size *Δ*.

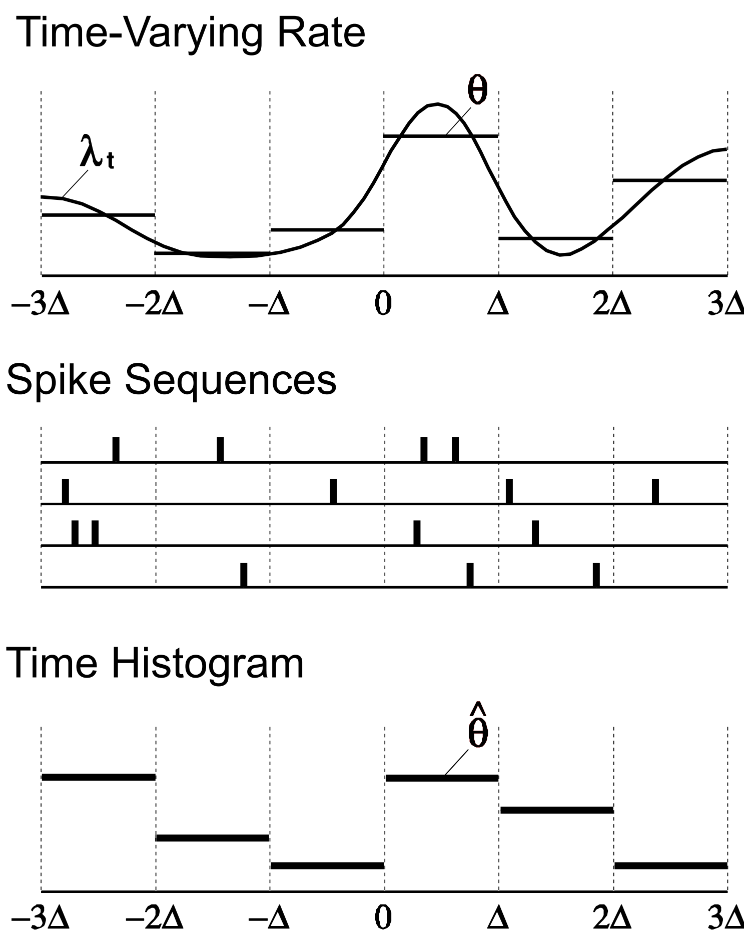
- Find *Δ** that minimize the cost function.



Mean $$k = \frac{1}{N}\sum_{i=1}^{N} k_i,$$

Variance $$v = \frac{1}{N}\sum_{i=1}^{N} (k_i - k)^2$$

Histogram construction



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

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

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

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

A histogram bar-height is an estimator of θ:

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

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

Mean integrated squared error

MISE can be written as average of bin-by-bin MISEs.

Expectation by the Poisson distribution, given the rate.

$$\mathrm{MISE} = \frac{1}{T}\int_0^T E\left(\hat{\lambda}_t - \lambda_t\right)^2 dt = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{\Delta}\int_{\Delta(j-1)}^{\Delta j} E\left(\hat{\theta}_n^j - \lambda_t\right)^2 dt$$

$$\equiv \left\langle \frac{1}{\Delta}\int_0^{\Delta} E\left(\hat{\theta}_n^j - \lambda_t^j\right)^2 dt \right\rangle = \left\langle \frac{1}{\Delta}\int_0^{\Delta} E\left(\hat{\theta}_n - \lambda_t\right)^2 dt \right\rangle$$

Average over segmented bins.

Decomposition of MISE (1)

Bias-Variance decomposition of MISE

$$E\hat{\theta}_n = \theta$$

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

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

Sampling Error (Variance)

Systematic Error (Bias)

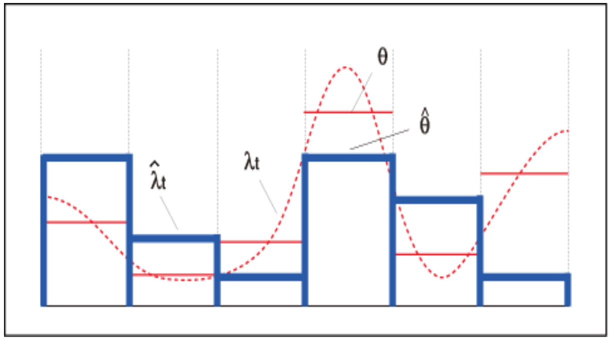
Decomposition of MISE (2)

Decomposition of the systematic error

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

Variance of the rate
Independent of Δ

Variance of an ideal histogram



The variance of an ideal histogram

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

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

Variance of a histogram — Mean fluctuation — Sampling error

Independent of Δ

Decomposition of MISE (3)

Hence, the MISE can be decomposed into the following parts.

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

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

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

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

Independent of Δ **Independent of Δ**

Cost function

We define a cost function by subtracting the terms independent from Δ.

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

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

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

Poisson:

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

The Δ that minimizes the cost function is an optimal bin size that minimizes the MISE.

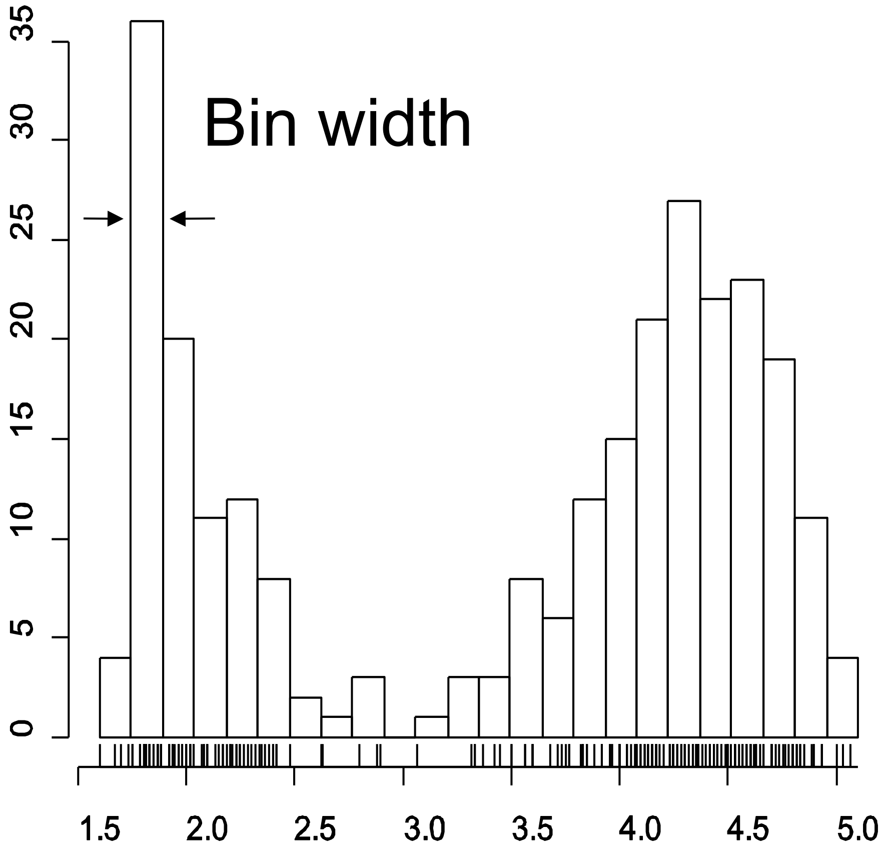
Finally, estimation of the cost function is given as

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

$$= \frac{2}{n\Delta}\frac{\bar{k}}{n\Delta} - \frac{1}{(n\Delta)^2}\left\langle\left(k_i - \bar{k}\right)^2\right\rangle = \frac{2\bar{k} - v}{(n\Delta)^2}$$

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

Background and significance



The duration for eruptions of the Old Faithful geyser in Yellowstone National Park (in minutes)

Sturges (1926) $$\Delta^* = \frac{\text{range of data}}{1+\log_2 n}$$

Scott (1979) $$\Delta^* = 3.49\sigma n^{-1/3}$$

Freedman and Diaconis (1981)

$$\Delta^* = 2\mathrm{IQR}\cdot n^{-1/3}$$

Rudemo (1982) Cross-validation

$$\hat{Q}(\Delta) = \frac{2}{(n-1)\Delta} - \frac{n+1}{n^2(n-1)}\sum_{i=1}^{N} k_i$$

Wand (1997) Plug-in method

Further topics on the optimal bin size

When the data size is large.

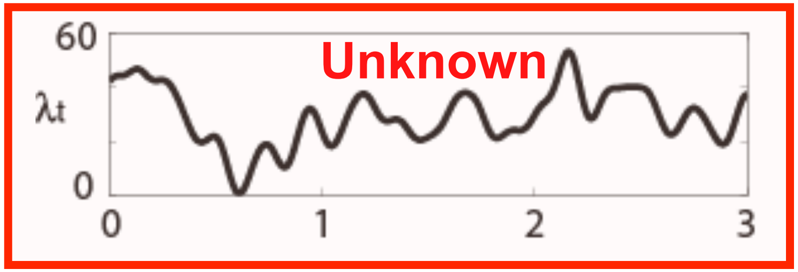
Asymptotic theory of an optimal bin size.

When the data size is small.

The divergence of an optimal bin size.
(The minimum number of trial to construct a histogram.)

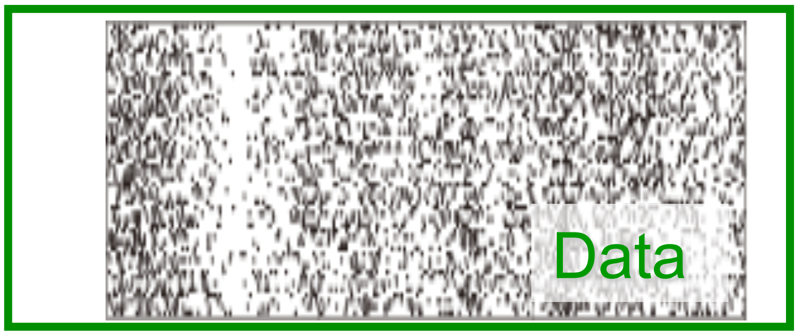
Solution:
A method to estimate the number of trials required to construct a histogram.

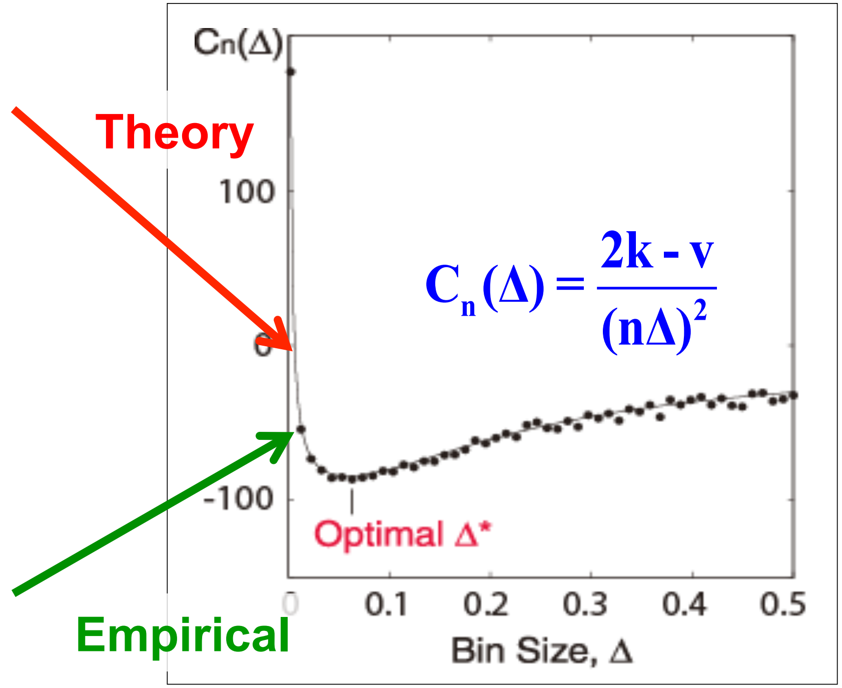
Theoretical cost function



Mean μ

Correlation function $\phi(t_1 - t_2)$





Theoretical cost function for a stationary underlying process

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

Scaling of the optimal bin size

Theoretical cost function: $C_n(\Delta) = \frac{\mu}{n\Delta} - \frac{1}{\Delta^2}\int_0^\Delta\int_0^\Delta \phi(t_1 - t_2)\,dt_1 dt_2.$

When the number of sequences is large, the optimal bin size becomes very small

Expansion of the cost function by Δ:

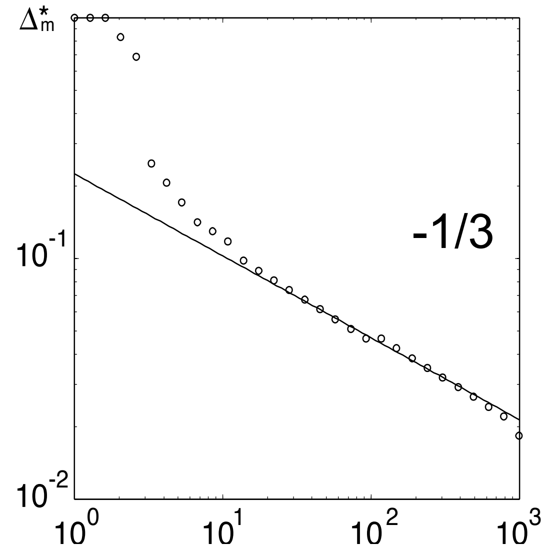
$$C_n(\Delta) = \frac{\mu}{n\Delta} - \phi(0) - \frac{1}{3}\phi'(0_+)\Delta - \frac{1}{12}\phi''(0)\Delta^2 + O(\Delta^3).$$

Scaling of the optimal bin size:

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

Ref. Scott (1979) $\Delta^* = 3.49\hat{\sigma}n^{-1/3}$

Scaling of the optimal bin size



Minimum number of trials for a histogram

When the number of sequences is small, the optimal bin size may become very large.

The expansion of the cost function by 1/Δ:

$$C_n(\Delta) \sim \frac{\mu}{n\Delta} - \frac{1}{\Delta}\int_{-\infty}^{\infty}\phi(t)\,dt + \frac{1}{\Delta^2}\int_{-\infty}^{\infty}|t|\,\phi(t)\,dt$$

$$= \mu\left(\frac{1}{n} - \frac{1}{n_c}\right)\frac{1}{\Delta} + u\frac{1}{\Delta^2}$$

The second order phase transition.

Critical number of trials: $n_c = \mu \Big/ \int_{-\infty}^{\infty}\phi(t)\,dt$

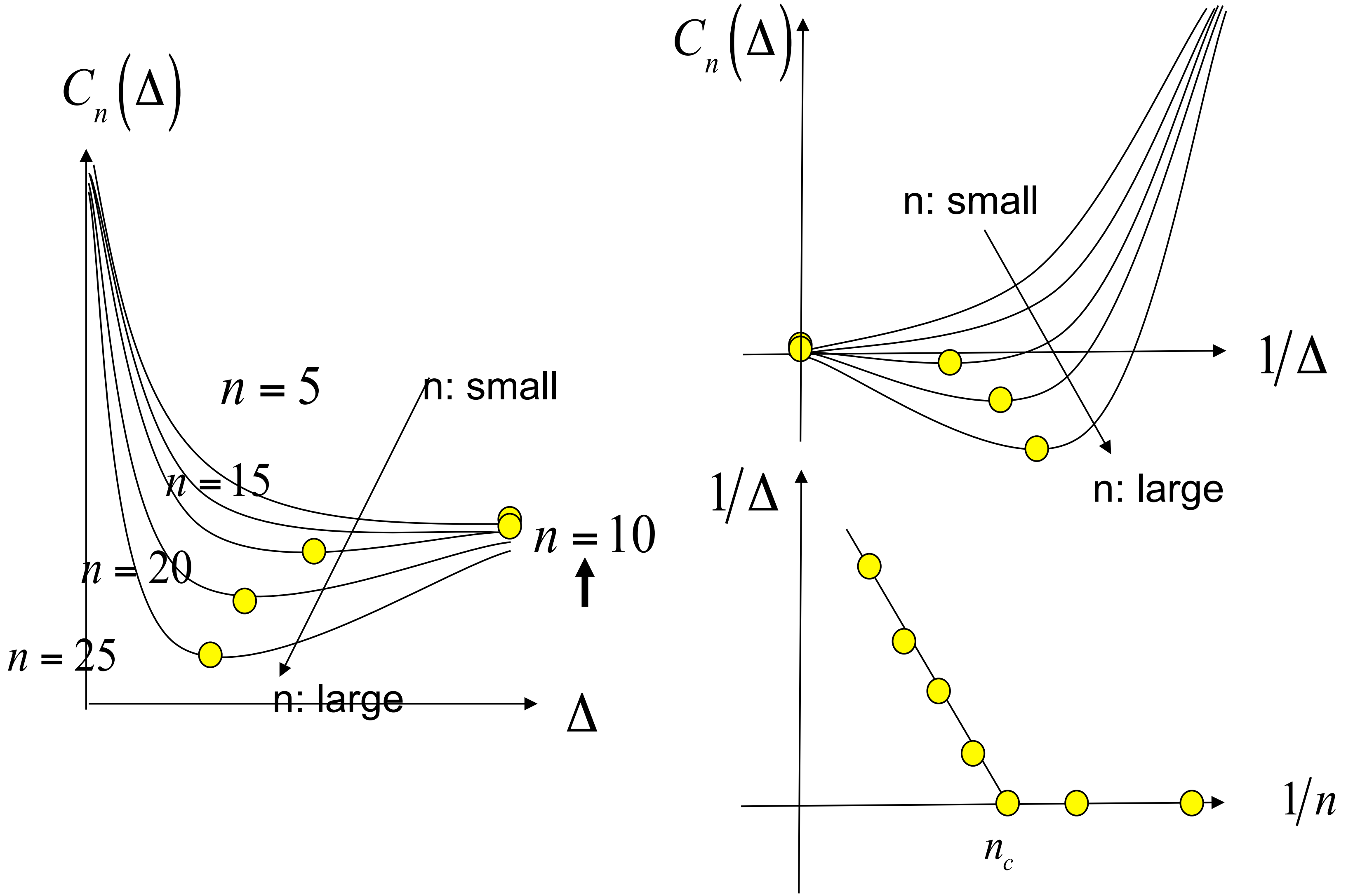
$n < n_c$ Optimal bin size diverges.

$n > n_c$ Finite optimal bin size.

cf. Koyama and Shinomoto *J. Phys. A*, 37(29):7255–7265. 2004

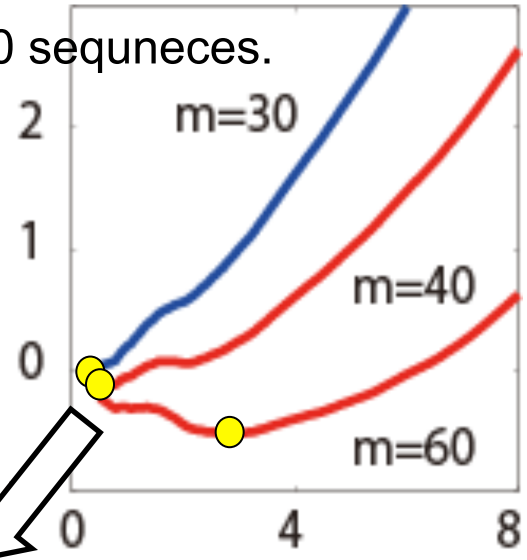
Not all the process undergoes the first order phase transition.
Others undergo the second order (discontinuous) phase transition.

Phase transitions of an optimal bin-width



Estimating the minimum number of trials

We have n=30 sequneces.



Optimal bin size v.s. m$^{1/\Delta}$

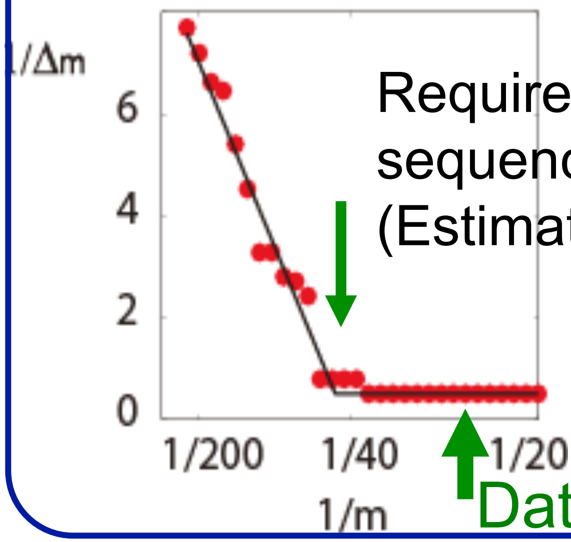
Original: $C_n(\Delta)$

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



Required # of sequences (Estimation)

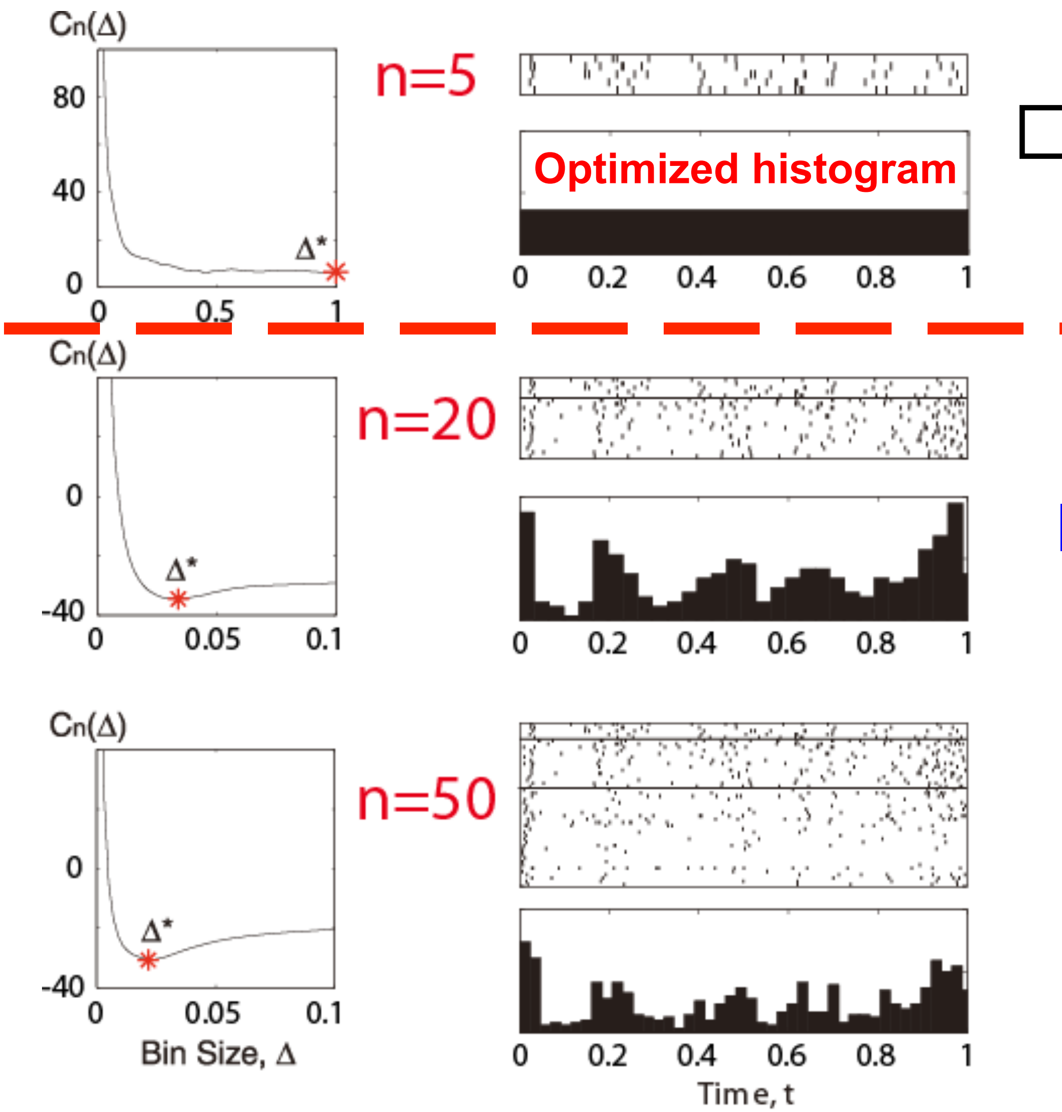
Data size used

Required # of trials

$$n_c = \frac{\mu}{\int_{-\infty}^{\infty}\phi(t)\,dt}$$

\# of sequences used

Application to MT neuron data



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

Too few to make a Histogram !

Extrapolation

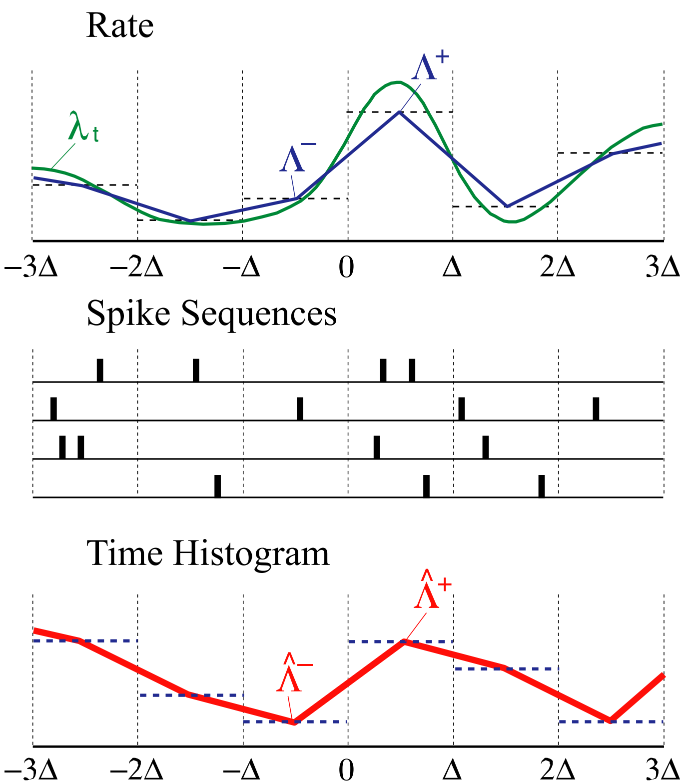
$$C_m(\Delta \mid n) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{k}{n\Delta^2} + C_n(\Delta)$$

Estimation:

At least **12** trials are required.

LINE-GRAPH HISTOGRAM

Line-graph time histogram



Line-Graph Model

A line-graph is constructed by connecting top-centers of adjacent bar-graphs.

$$L_t = \frac{\theta^+ + \theta^-}{2} + \frac{\theta^+ - \theta^-}{\Delta} t. \qquad \theta^- \equiv \frac{1}{\Delta}\int_{-\Delta}^{0} \lambda_t \, dt. \quad \Lambda^+ \equiv \frac{1}{\Delta}\int_{0}^{\Delta} \lambda_t \, dt.$$

The spike count obeys the Poisson distribution

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

An estimator of a line-graph

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

An algorithm for optimizing line-graph histogram

(i) Define the four spike counts,

$$k_i^{(+)}(j) \quad k_i^{(-)}(j) \quad k_i^{(0)}(j) \quad k_i^{(*)}(j) \qquad p = \{-,+,0,*\}$$

(ii) Summation of the spike count $\quad k_i^{(p)} \equiv \sum_{j=1}^{n} k_i^{(p)}(j)$

Covariations w.r.t. bins

$$s^{(p,q)} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(k_i^{(p)} - \bar{k}^{(p)}\right)\left(k_i^{(q)} - \bar{k}^{(q)}\right) \qquad \bar{k}^{(p)} \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{(p)}$$

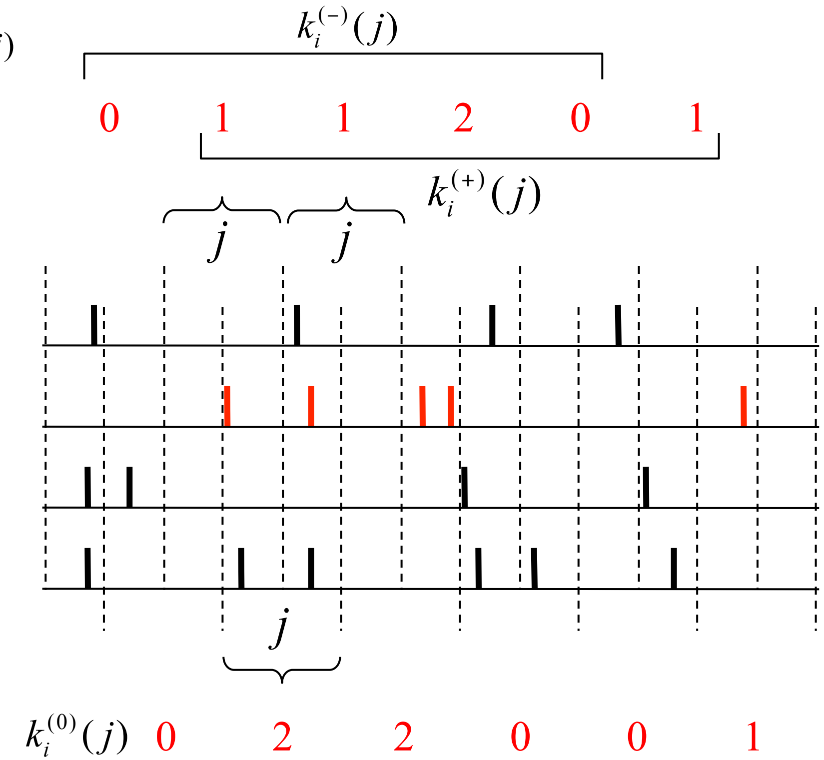
Binned-average

Bin-average of the covariation of spike count w.r.t. sequences,

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

(iii) The covariances of an ideal line-graph model is

$$\sigma^{(p,q)} \equiv \frac{s^{(p,q)}}{(n\Delta)^2} - \frac{\bar{s}^{(p,q)}}{n\Delta^2}$$

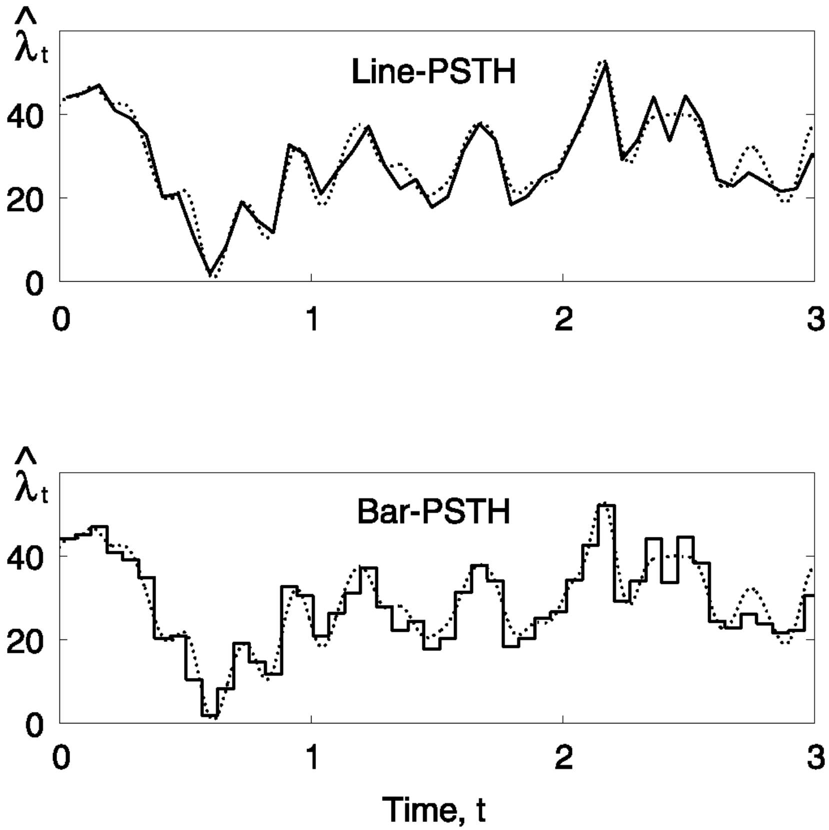


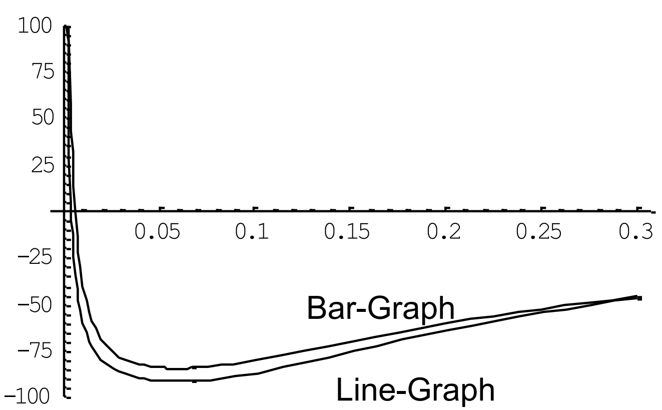
(iv) Cost function:

$$C_n(\Delta) = \frac{2}{3}\frac{\bar{k}^{(+)}}{(n\Delta)^2} + \frac{2}{3}\sigma^{(+,+)} + \frac{1}{3}\sigma^{(+,-)} - 2\sigma^{(+,0)} - 2\sigma^{(+,*)}. \qquad k_i^{(*)}(j) \equiv 2\sum_l t_i^l(j)/\Delta$$

(v) Repatn i through iv by changing Δ. Find the optimal Δ that minimizes the cost function.

The optimal line-graph histogram





A line-graph histogram generally performs better than a bar-graph histogram.

Theoretical cost functions

(Bar-Graph) $$C_n(\Delta) = \frac{\langle\theta\rangle}{n\Delta} - \left\langle \left(\theta - \langle\theta\rangle\right)^2 \right\rangle$$
$$= \frac{\mu}{n\Delta} - \frac{1}{\Delta^2}\int_0^\Delta \int_0^\Delta \phi(t_1 - t_2)\, dt_1 dt_2 .$$

(Line-Graph) $$C_n(\Delta) = \frac{2\mu}{3n\Delta} - \frac{2}{\Delta^2}\int_0^\Delta \int_{-\Delta/2}^{\Delta/2} \left(1 + \frac{2t_2}{\Delta}\right) \phi(t_1 - t_2)\, dt_1 dt_2$$
$$+ \frac{2}{3\Delta^2}\int_0^\Delta \int_0^\Delta \phi(t_1 - t_2)\, dt_1 dt_2 + \frac{1}{3\Delta^2}\int_0^\Delta \int_{-\Delta}^0 \phi(t_1 - t_2)\, dt_1 dt_2 .$$

Scalings of the optimal bin size

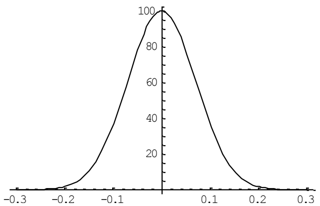
The expansion of the cost function by Δ:

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

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

the second order term vanishes.

A smooth process: A correlation function is smooth at origin.

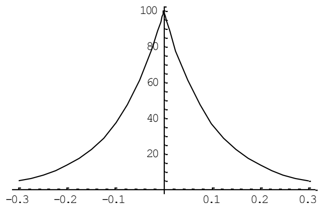


$\phi'(0_+) = 0$

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

(Line-Graph) $$\Delta^* \sim \left(\frac{1280\mu}{181\phi''''(0)n}\right)^{1/5}.$$

A jagged process: A correlation function has a cusp at origin.



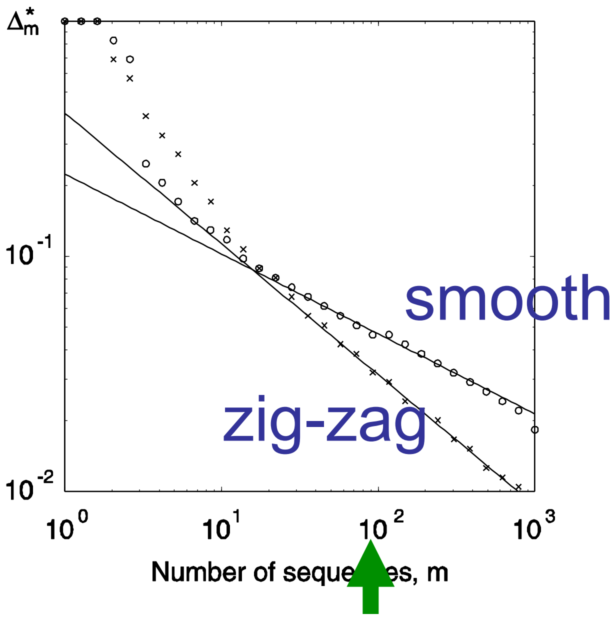
$\phi'(0_+) \neq 0$

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

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

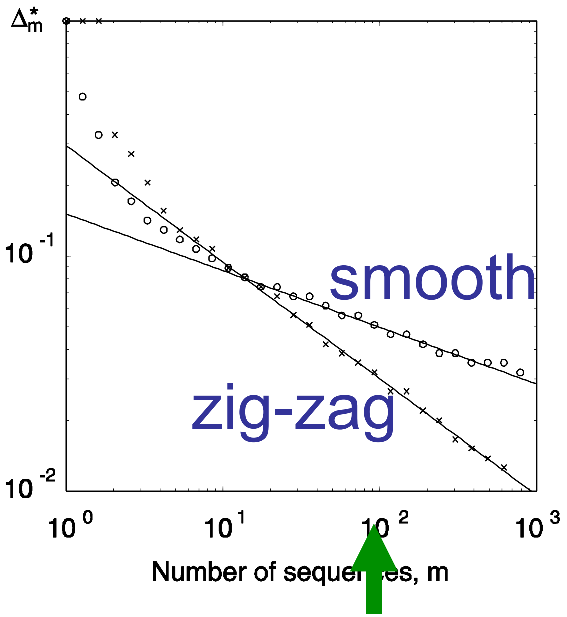
Identification of the scaling exponents

Bar-graph Histogram



Data size used.

Line-graph Histogram



Data size used.

	Bar-graph Histogram	Line-graph Histogram
smooth	$\Delta^* \sim n^{-1/3}$	$\Delta^* \sim n^{-1/5}$
zig-zag	$\Delta^* \sim n^{-1/2}$	$\Delta^* \sim n^{-1/2}$

KERNEL DENSITY OPTIMIZATION

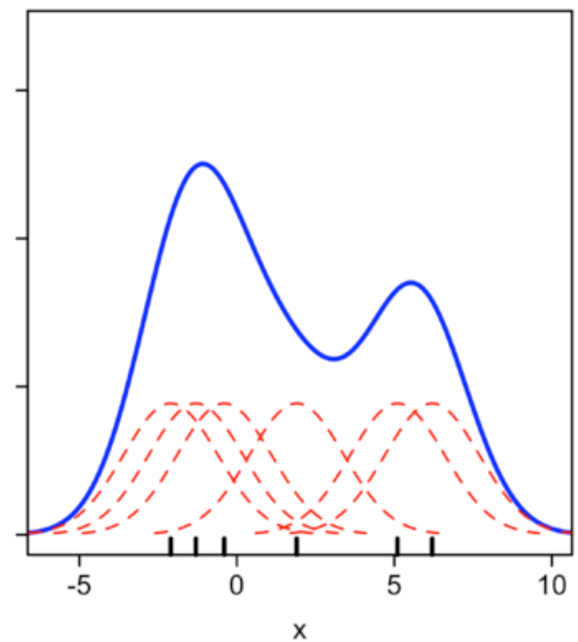
Rate estimation by kernel convolution

$$\lambda(t) = \sum_{i=1}^{n} k_w(t - t_i) = \int_0^T k_w(t-s)\,x(s)\,ds \qquad x(t) = \sum_{i=1}^{n} \delta(t - t_i)$$

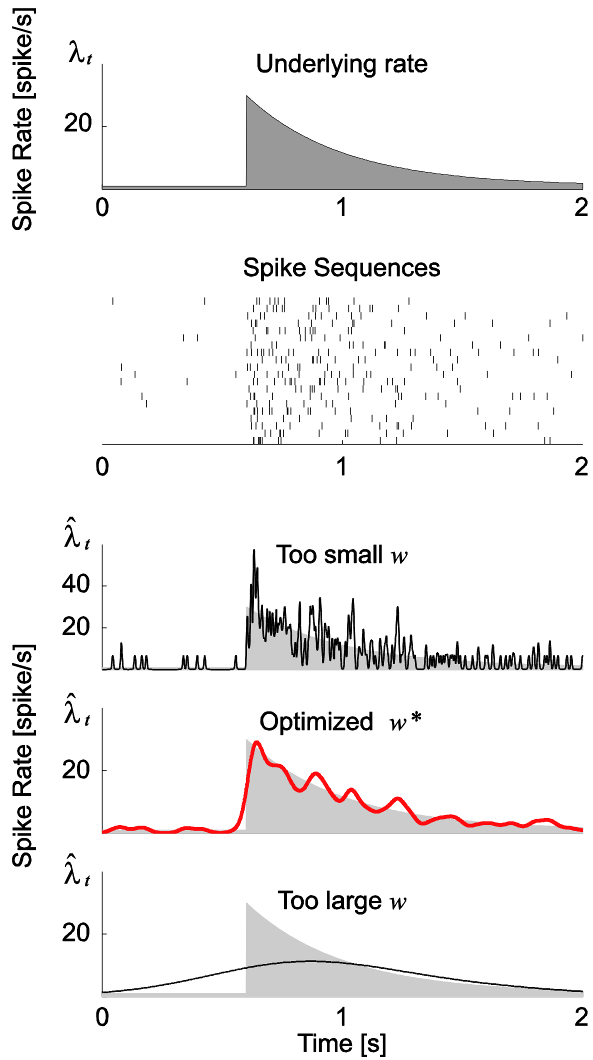
Kernel function with bandwidth w.

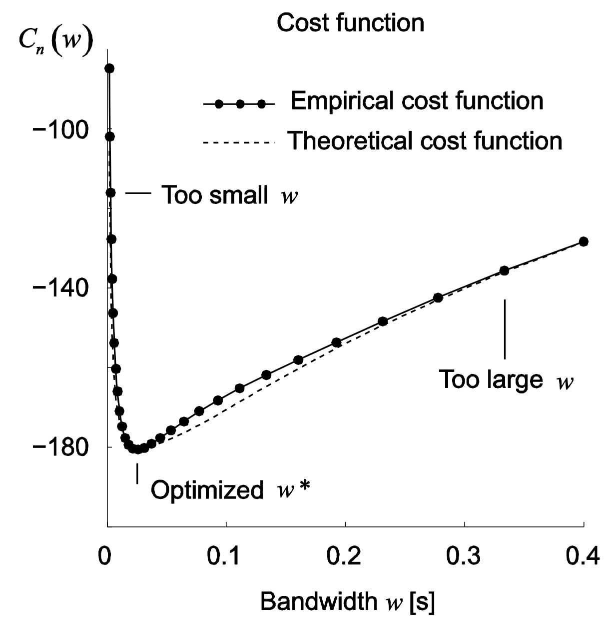
Example: A Gaussian function.

$$k_w(s) = \frac{1}{\sqrt{2\pi w}} e^{-\frac{s}{2w^2}}$$



Kernel bandwidth optimization

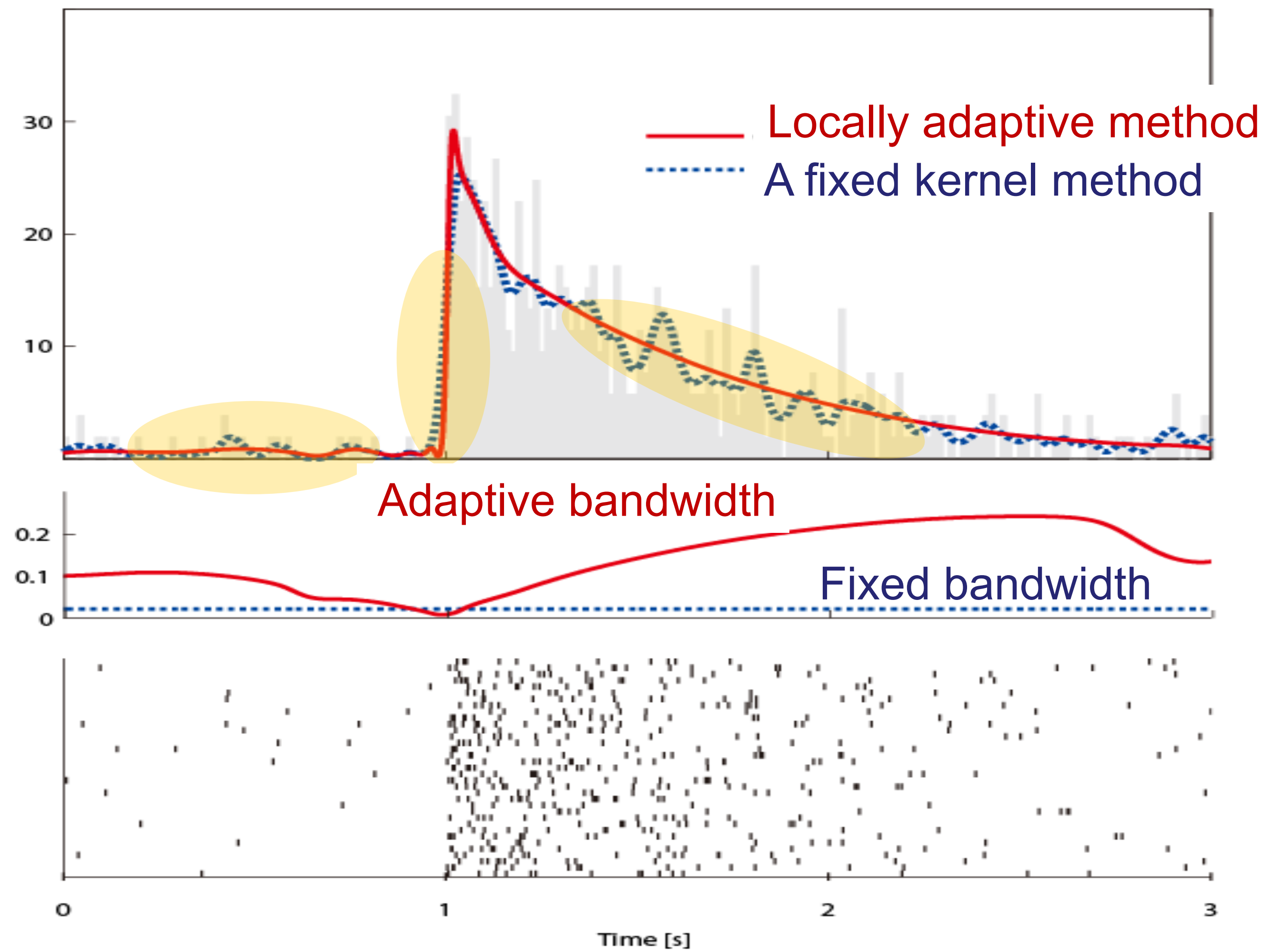




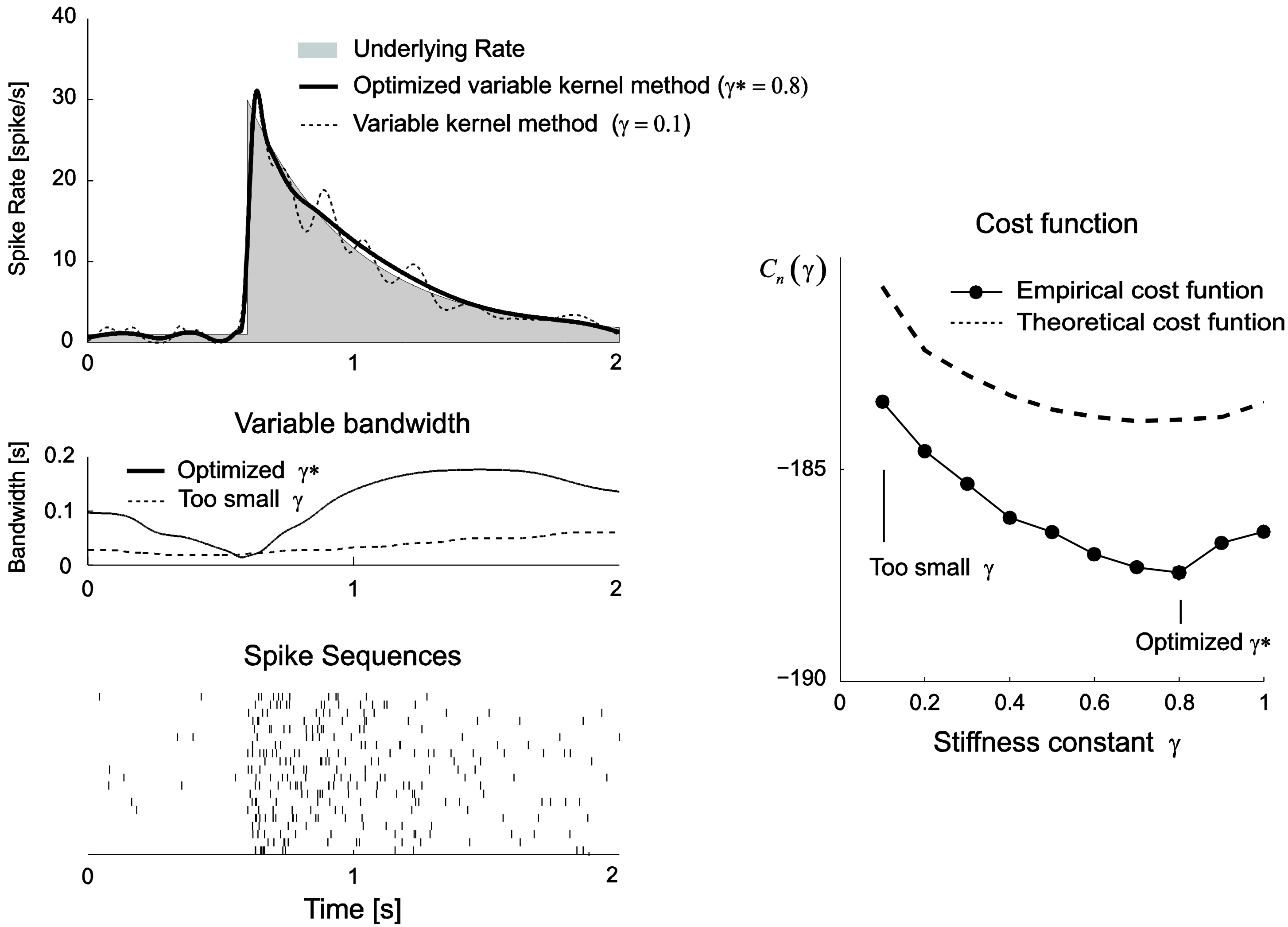
$$C(w) = \sum_{i,j} \int k_w(s - t_i) k_w(s - t_j) ds - 2 \sum_{i \neq j} k_w(t_i, t_j)$$

Shimazaki & Shinomoto, J. Comp. Neurosci 2010

Locally adaptive kernel density estimation



Optimization of locally adaptive kernel method

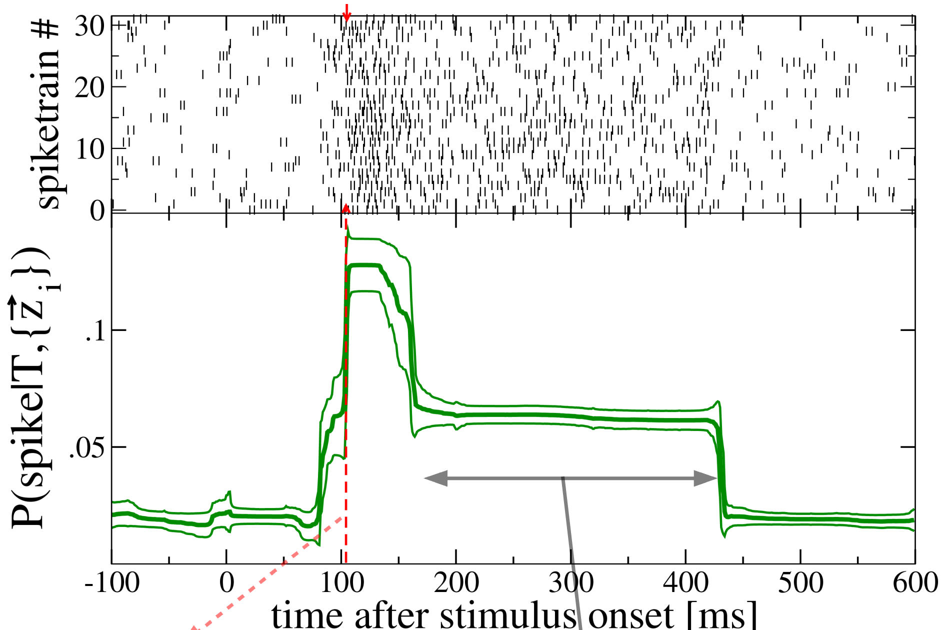


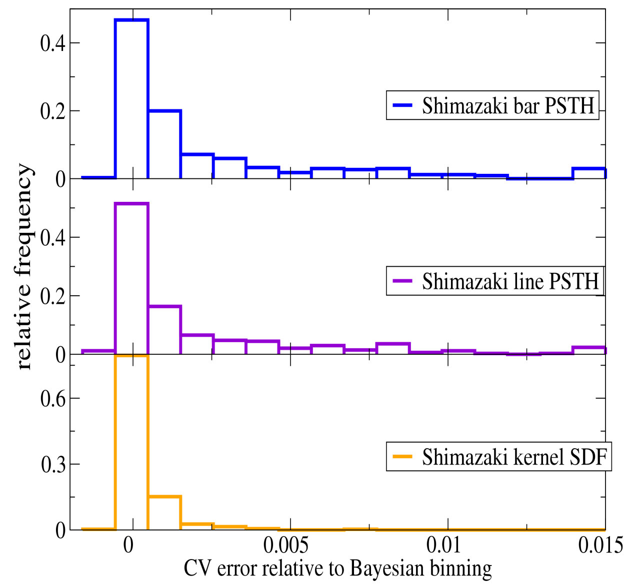
Automatically adjusts the stiffness of bandwidth variability.

Shimazaki & Shinomoto, J. Comp. Neurosci 2010

OTHER ADAPTIVE ESTIMATION METHODS

Bayesian Binning



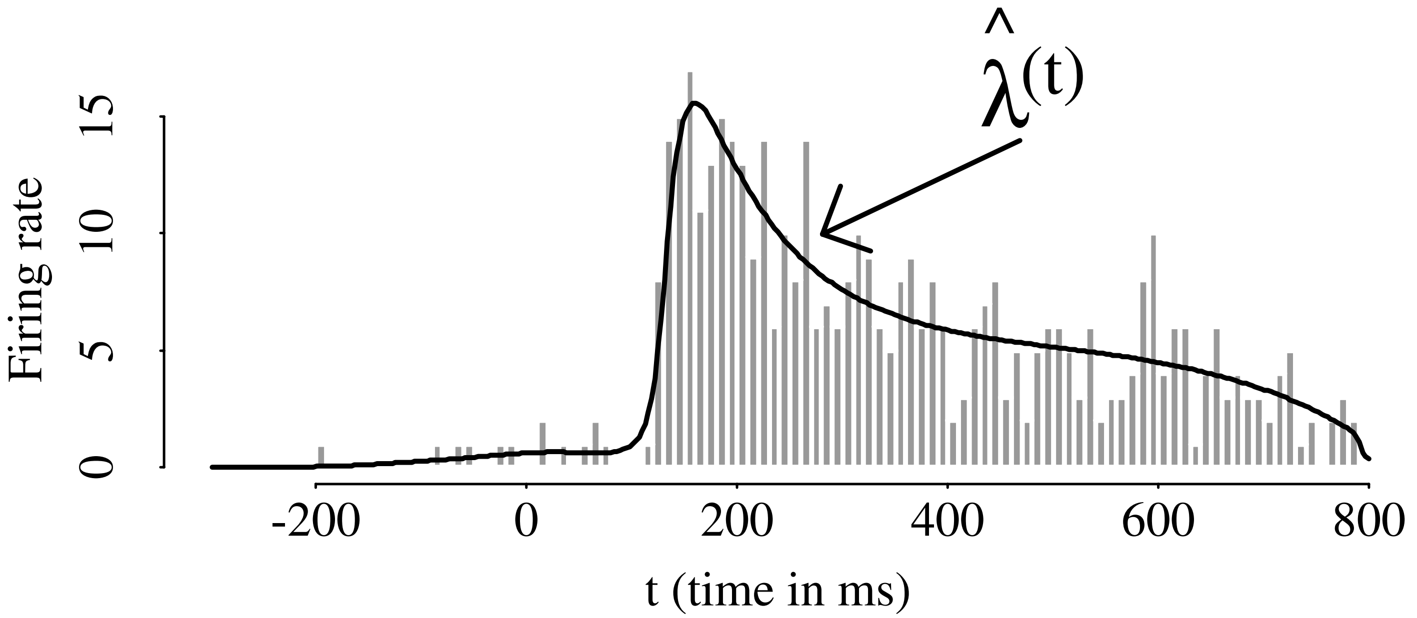


Latency is precisely captured!

noise is not overfitted!

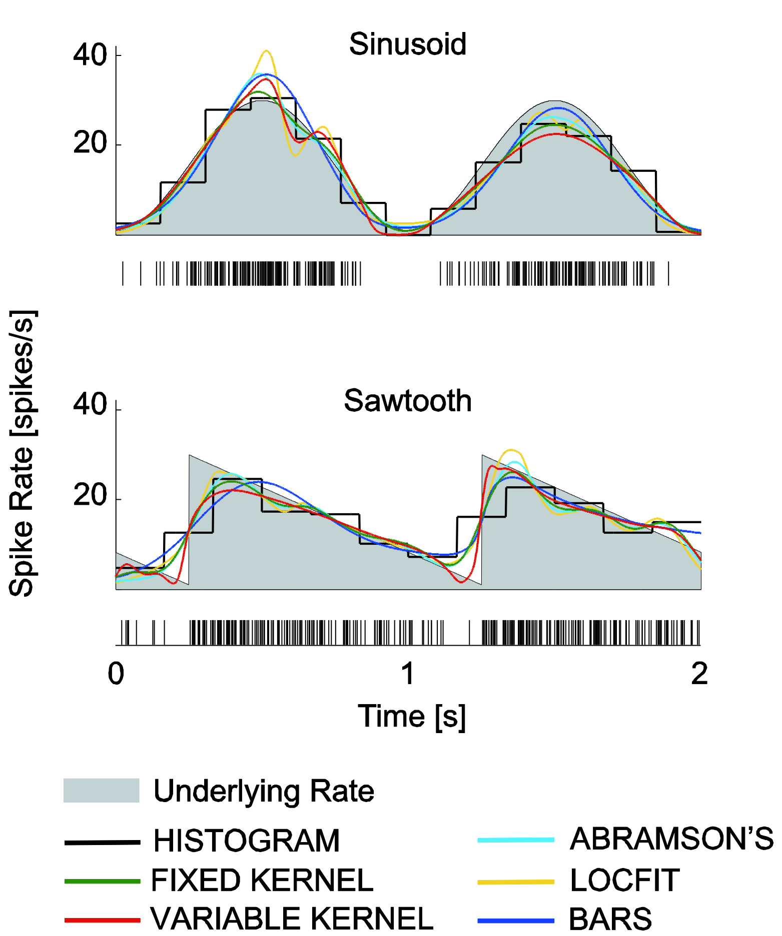
Endres, D., Oram, M., Schindelin, J., & F, P. Foldiak. Bayesian binning beats approximate alternatives : estimating peristimulus time histograms. NIPS 2008
Endres, D and Oram, M, Feature extraction from spike trains with Bayesian binning: 'Latency is where the signal starts' J Comput Neurosci 2009.

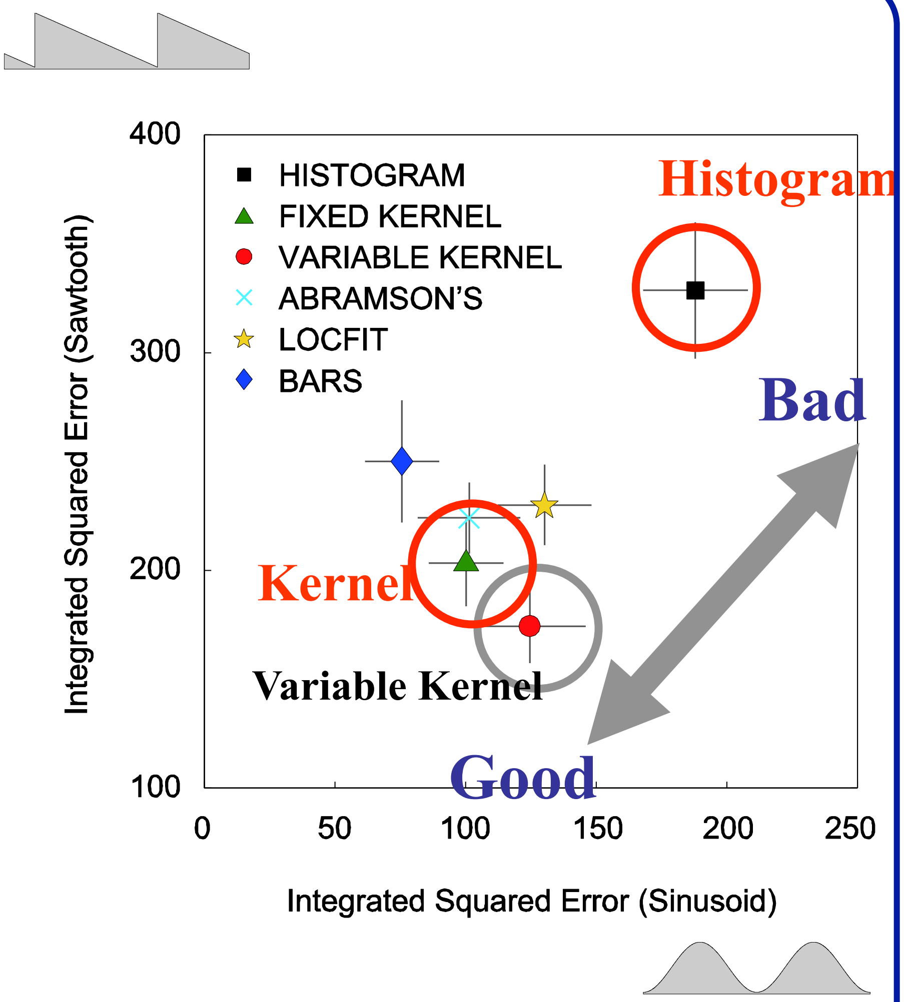
Bayesian adaptive regression splines



Kass RE, Ventura V, Cai C, Statistical smoothing of neuronal data. Network-Computation in Neural Systems 2003
DiMatteo I, Genovese C R, Kass RE, Bayesian curve-fitting with free-knot splines. Biometrika 2001.

Performance comparison





A simple kernel method is comparable to, or even better than modern methods.

Shimazaki & Shinomoto, J. Comp. Neurosci 2010

Conclusion

Single Neurons Spike-rate Estimation

⇩

Fixed Kernel Method

(Simple and Accurate enough = Practical)

What we learned

1. A simple formula for a **histogram optimization.**
2. **Scaling and phase-transition property of an optimal bin size.**
3. **Kernel density optimization method.**
4. **Adaptive methods** are still under active research area.

Tomorrow we will learn

1
- **Framework of a state-space model.**

2
- **Recursive Bayesian filter. Laplace's method, and Newton-raphson method.**

3
- **Simultaneous estimation of posterior and parameters (EM-algorithm).**

4
- **Model validation (Bayes factor, ABIC),**

5
- Applications to **neural decoding** and **plasticity**.