Optimization of a Histogram of Spike Data

Hideaki Shimazaki and Shigeru Shinomoto
Department of Physics, Kyoto University, Kyoto 606-8502, Japan

Overview

A classical tool for estimating the neuronal spike rate is a peri-stimulus time histogram (PSTH) constructed from spike sequences aligned at the onset of a stimulus repeatedly applied to an animal. Generally in the neurophysiological literature, the bin size that critically determines the goodness of the fit of the time histogram to the underlying spike rate has been subjectively selected by individual researchers.

We have recently established a method for selecting the bin size, so that the PSTH best represents the unknown underlying rate (1; 2). The goodness of the fit we adopted as the optimization principle is minimizing the mean integrated squared error (MISE) between the underlying rate λ_t and the PSTH $\hat{\lambda}_t$,

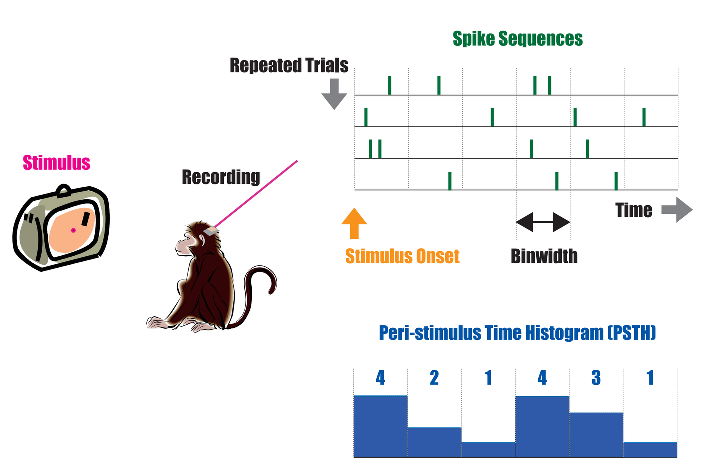
$$\mathrm{MISE} = \int_a^b E(\lambda_t - \hat{\lambda}_t)^2 dt, \qquad (1)$$

where E refers to the expectation with respect to the spike generation process under a given time-dependent rate λ_t. The method allows us to minimize the MISE from spike count statistics alone, without knowing the underlying rate. Generally, the cross-validation method is applicable to the least squares minimization (5; 6). Here, we estimate the MISE fully utilizing the Poissonian nature of spikes.

For a small number of spike sequences generated from a modestly fluctuating rate, the optimal bin size may diverge, indicating that any time histogram is likely to capture a spurious rate. Given a paucity of data, the present method can nevertheless suggest how many experimental trials should be added in order to obtain a meaningful time-dependent histogram with the required accuracy.

Rate Estimation of Neuronal Spikes

Neurophysiologists construct a peristimulus time histogram.



In neurophysiological studies, many researchers are interested in estimating the time-dependent rate of spike occurrence with relation to an external stimulus. For this purpose, a peristimulus time histogram (PSTH) is frequently constructed from the data obtained by repeatedly applying an identical stimulus to an animal.

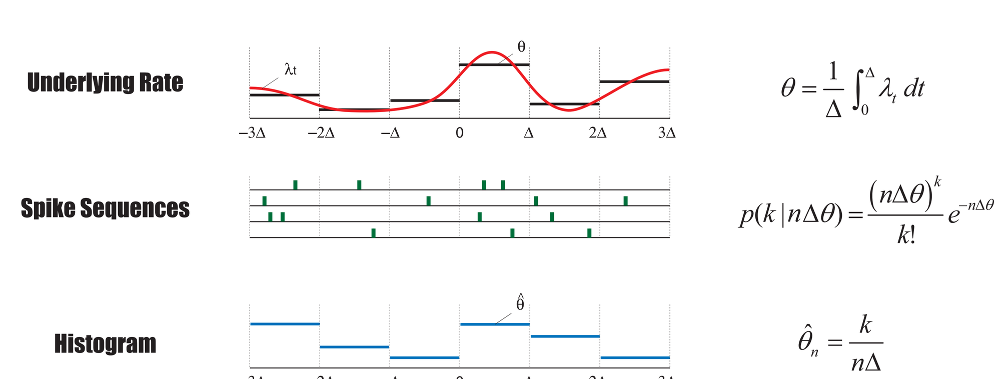
A PSTH is constructed as follows: Align spike sequences to the onset of stimuli, divide time into discrete bins, count the number of spikes that enter each bin, and divide the counts by the bin size and the number of sequences.

Statistical Description of a PSTH

The Poissonian assumption holds for a PSTH.

The Poissonian assumption holds for the spikes accumulated from a large number of trials because spikes repeatedly recorded from a single neuron under identical experimental conditions are in the majority mutually independent.

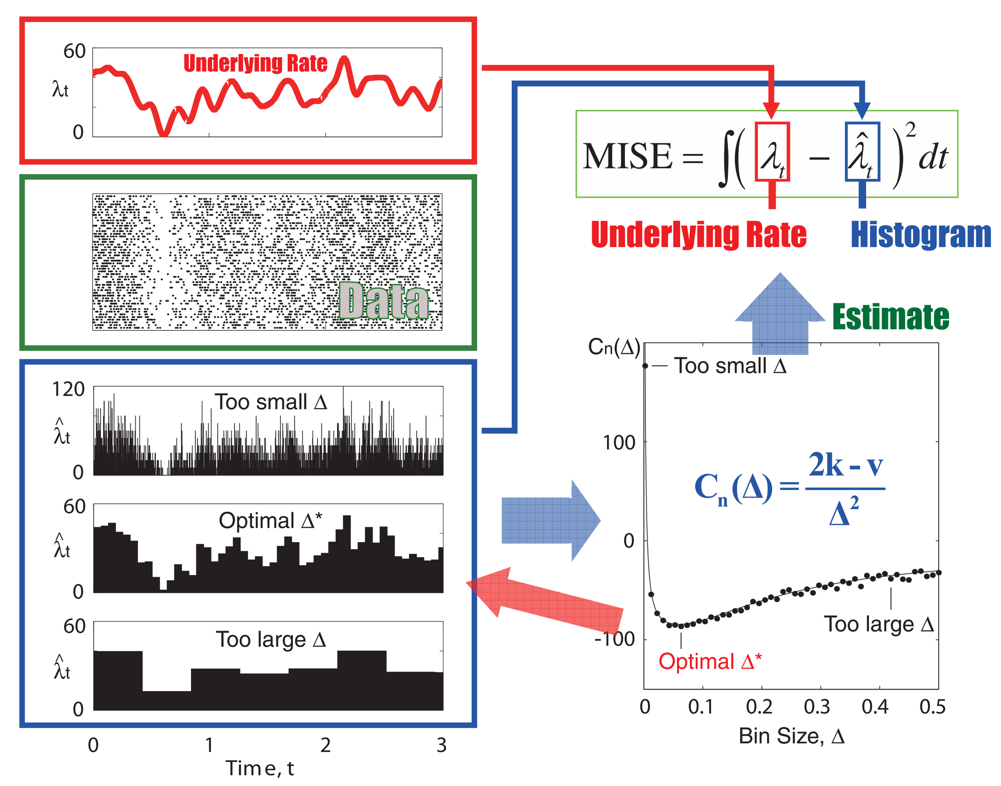
A PSTH is an estimate for the piecewise average of the time-dependent rate of a Poisson point process.



1. The average of a time-dependent rate over the bin size Δ is denoted as θ.
2. The number of spikes within a bin obeys the Poisson statistics with the mean $n\Delta\theta$.
3. A histogram bar height $\hat{\theta}$ is an unbiased estimator of the ideal height θ.

Optimization of a Histogram

We use the MISE criterion to define optimality of a PSTH.



1. An unknown underlying rate that generates observable events.
2. Spike sequences generated from the unknown rate.
3. Histograms of the events are constructed. It is not obvious which bin size should be used.
4. By introducing a **Mean Integrated Squared Error (MISE)**, a measure of the goodness-of-the-fit of a histogram to the unknown rate, the optimal bin size is defined as the one that minimizes the MISE.
5. Note that we can not directly compute the MISE because we do not know the underlying rate. However, the MISE can be estimated with the formula provided in the Algorithm 1.
6. The optimal bin size can be estimated as the one that minimizes the estimated MISE.

Bin Size Selection

We provide the method for selecting a bin size.

Segmentation of the MISE:

$$\mathrm{MISE} = \frac{1}{\Delta}\int_0^\Delta \frac{1}{N}\sum_{i=1}^N \left\{ E\left(\hat{\theta}_i - \lambda_{t+(i-1)\Delta}\right)^2 \right\} dt \;\equiv\; \frac{1}{\Delta}\int_0^\Delta \left\langle E\left(\hat{\theta} - \lambda_t\right)^2 \right\rangle dt.$$

Decomposition of the MISE:

$$\mathrm{MISE} \;=\; \underbrace{\left\langle E(\hat{\theta} - \theta))^2 \right\rangle}_{\textbf{Sampling Error}} \;+\; \underbrace{\frac{1}{\Delta}\int_0^\Delta \left\langle (\lambda_t - \theta)^2 \right\rangle dt.}_{\textbf{Systematic Error}} \qquad (2)$$

Decomposition of the Systematic Error:

$$\frac{1}{\Delta}\int_0^\Delta \left\langle (\lambda_t - \langle\theta\rangle + \langle\theta\rangle - \theta)^2 \right\rangle dt \;=\; \underbrace{\frac{1}{\Delta}\int_0^\Delta \left\langle (\lambda_t - \langle\theta\rangle)^2 \right\rangle dt}_{\textbf{Independent of }\Delta} \;-\; \underbrace{\left\langle (\theta - \langle\theta\rangle)^2 \right\rangle.}_{\textbf{Variance of an Ideal Histogram}}$$

Definition of a cost function:

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

The decomposition rule for an unbiased estimator ($E\hat{\theta} = \theta$):

$$\underbrace{\left\langle E(\hat{\theta} - \langle E\hat{\theta}\rangle)^2 \right\rangle}_{\textbf{Variance of a Histogram}} \;=\; \underbrace{\left\langle E(\hat{\theta} - \theta)^2 \right\rangle}_{\textbf{Sampling Error}} \;+\; \underbrace{\left\langle (\theta - \langle\theta\rangle)^2 \right\rangle.}_{\textbf{Variance of an Ideal Histogram}}$$

The cost function written with the observable:

$$\begin{aligned} C_n(\Delta) \;&=\; 2\left\langle E(\hat{\theta} - \theta)^2 \right\rangle \;-\; \left\langle E(\hat{\theta} - \langle E\hat{\theta}\rangle)^2 \right\rangle. \\ &=\; \underbrace{\frac{2}{n\Delta}\left\langle E\hat{\theta} \right\rangle}_{\textbf{Mean of a Histogram}} \;-\; \underbrace{\left\langle E(\hat{\theta} - \langle E\hat{\theta}\rangle)^2 \right\rangle.}_{\textbf{Variance of a Histogram}} \end{aligned}$$

Algorithm 1: A Method for Optimizing a Time Histogram.

(i) Divide the observation period T into N bins of width Δ, and count the number of spikes k_i from all n sequences that enter the ith bin.

(ii) Construct the mean and variance of the number of spikes $\{k_i\}$ as,
$$\bar{k} \equiv \frac{1}{N}\sum_{i=1}^N k_i, \text{ and } v \equiv \frac{1}{N}\sum_{i=1}^N (k_i - \bar{k})^2.$$

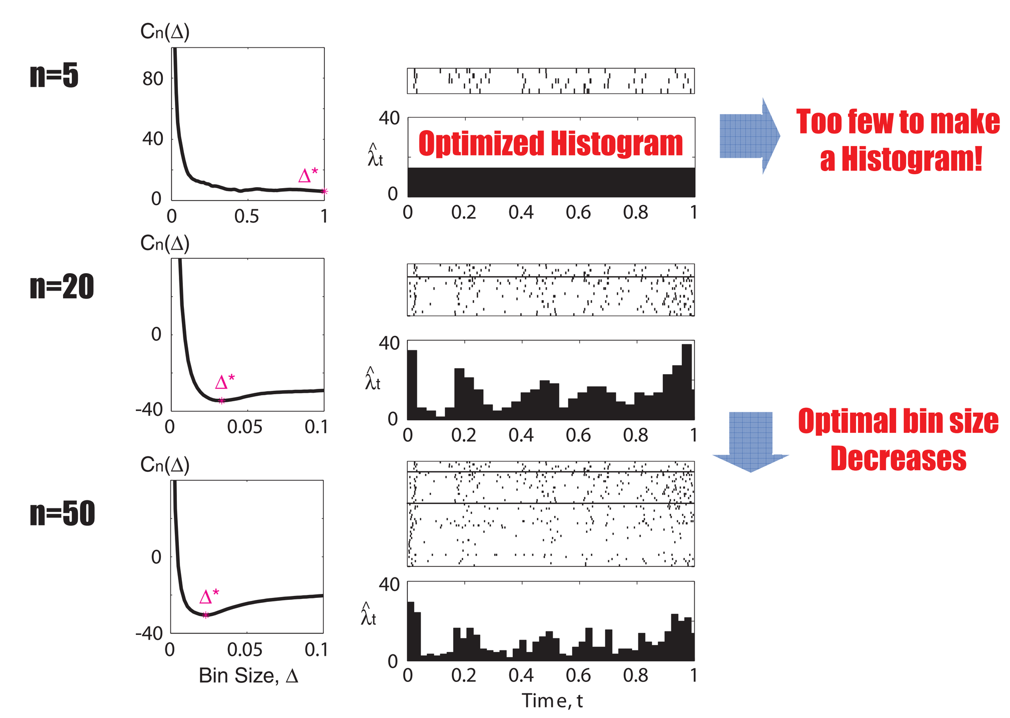
(iii) Compute the cost function,
$$C_n(\Delta) = \frac{2\bar{k} - v}{(n\Delta)^2}.$$

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

Application to Neuronal Spike Data

There is a minimum number of experimental trials required to construct a PSTH.

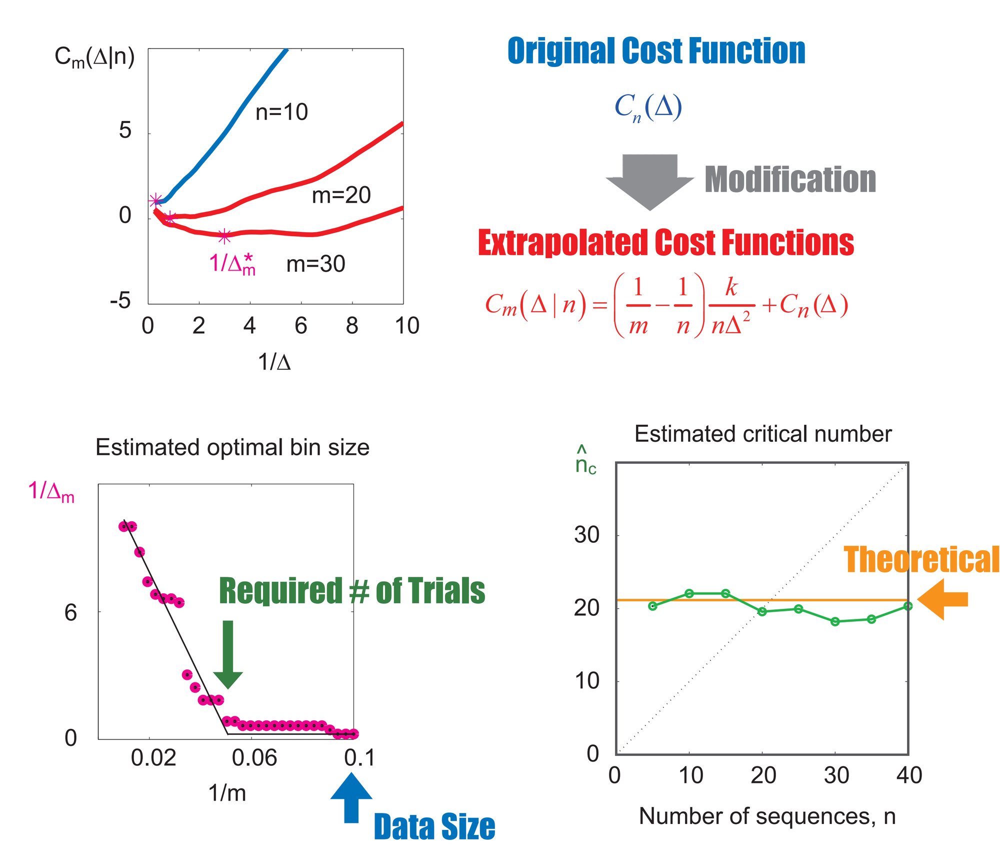
PSTHs for the spike sequences recorded from a MT neuron (w052 in nsa2004.1, Britten et al., 2004).



1. For a small number of trials, the optimal bin size may become as large as the observation period.
   ⟶ Phasetransition theory on the optimal bin size. (See Ref. 1,3)
2. For a large number of trials, the optimal bin size decreases as the number of experimental trials increases.
   ⟶ Power-law scaling theory on the optimal bin size. (See Ref. 1,3)

The Extrapolation Method

It is possible to estimate the minimum number of trials required to construct a PSTH.



1. It is possible to construct a cost function for an arbitrary number spike sequences (an extrapolated cost function).
2. With the extrapolated cost functions, the experimentalist can estimate how many experimental trials is required to construct a histogram with a resolution they deem sufficient.
3. The minimum number of experimental trials needed to construct a histogram may be estimated with several initial trials.

Algorithm 2: A Method for Extrapolating the Cost Function.

(A) Construct the extrapolated cost function,
$$C_m(\Delta|n) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{\bar{k}}{n\Delta^2} + C_n(\Delta),$$
using the sample mean $\bar{k}$ and variance v of the number of spikes obtained from n sequences of spikes.

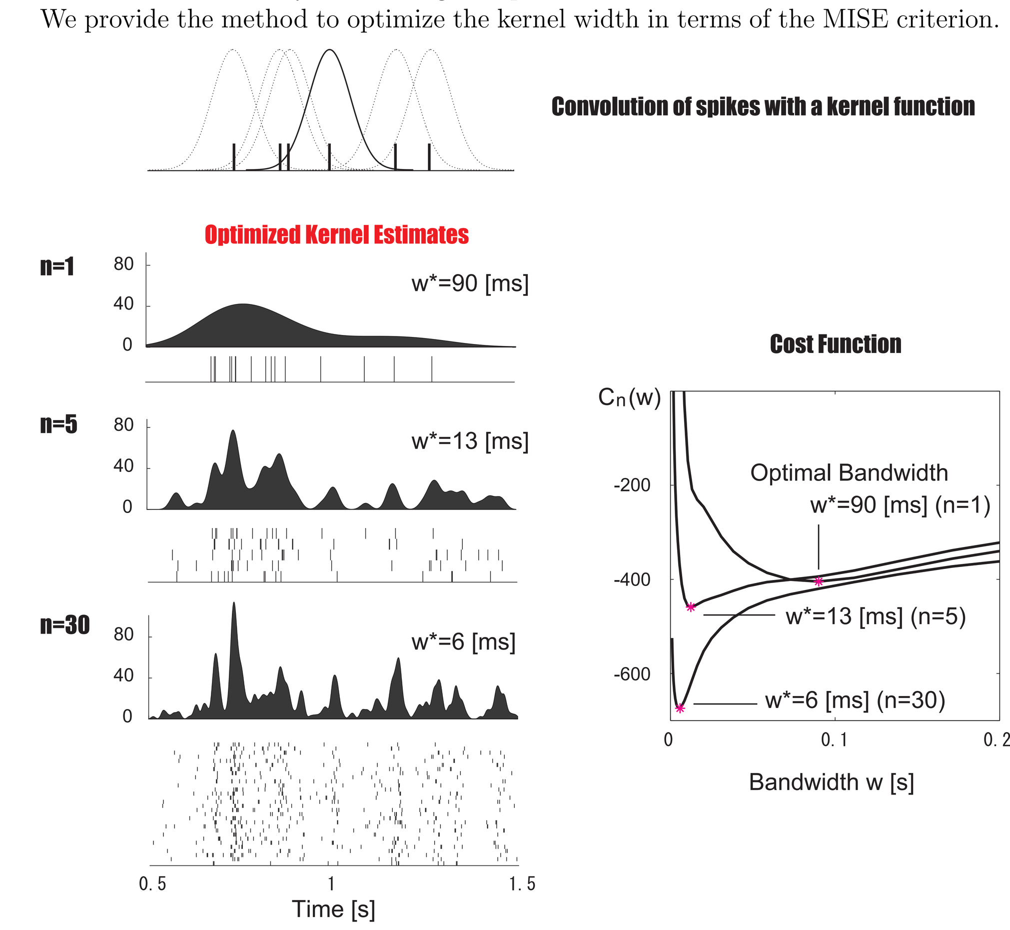
(B) Search for Δ_m^* that minimizes $C_m(\Delta|n)$.

(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.

Kernel Optimization

A method for kernel optimization is available.

Another frequently used tool for a rate estimation is a kernel method. In this method, the rate is estimated by convoluting the spikes with a kernel function with a width w. We provide the method to optimize the kernel width in terms of the MISE criterion.



Algorithm 3: Kernel Optimization Method. Ref. (2)

(i) Superimpose all the n spike sequences. Obtain a series of spike times $\{t_i\}_{i=1}^N$ in $[a, b]$. N is the total number of spikes.

(ii) Compute the cost function of a kernel $k_w(t)$ as
$$\hat{C}_n(w) = -\frac{4}{n^2}\sum_{i<j} k_w(t_i - t_j) + \frac{1}{n^2}\sum_{i,j} \psi_{w,a,b}(t_i - t_j),$$
where $\psi_{w,a,b}(t) \equiv \int_a^b k_w(s)\, k_w(s+t)\, ds$ is the correlation function $^{(*)}$.

(iii) Repeat ii while changing w to search for w^* that minimizes $\hat{C}_n(w)$.

$^{(*)}$ For a Gaussian kernel $k_w(t) = \frac{1}{\sqrt{2\pi}w}\exp\left(-\frac{t^2}{2w^2}\right)$,
$\psi_{w,a,b}(t) = \frac{1}{\sqrt{\pi}4w}\exp\left(-\frac{t^2}{4w^2}\right)\left\{\mathrm{erf}\left(\frac{2b+t}{2w}\right) - \mathrm{erf}\left(\frac{2a+t}{2w}\right)\right\}$.

Other Methods Available for a Rate Estimation

1. A method for optimizing a ling-graph histogram is available (Ref. 1).
2. Rate estimation in Bayesian framework has been developed by Koyama, Shimokawa, and Shinomoto (Ref. 4).
3. Post your data. We calculate the optimal bin size online. Please visit us at

http://www.ton.scphys.kyoto-u.ac.jp/~shino/toolbox/english.htm

Acknowledgements

This study is supported in part by Grants-in-Aid for Scientific Research to SS from the Ministry of Education, Culture, Sports, Science and Technology of Japan (16300068, 18020015) and the 21st Century COE "Center for Diversity and Universality in Physics". HS is supported by the Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists.

HS Current Address: shimazaki@brain.riken.jp Grün & Diesmann Research Unit at BSI, RIKEN.

REFERENCES

[1] Shimazaki, H. and Shinomoto, S. (2007) A method for selecting the bin size of a time histogram, *Neural Computation*, **19**: 1503-1527.

[2] Shimazaki, H. and Shinomoto, S. Kernel Width Optimization in the Spike-rate Estimation, *in Preparation*

[3] Koyama, S. and Shinomoto, S. (2004) Histogram bin width selection for time-dependent Poisson processes, *Journal of Physics A-Mathematical and General*, **37**: 7255-7265.

[4] Koyama, S. and Shinomoto, S. (2005) Empirical Bayes interpretations of random point events, *J. Phys.A*, **38**: L531-L537. Koyama, S., Shimokawa, T., and Shinomoto, S. Phase transitions in the estimation of event-rate: A path integral analysis (2007), *J. Phys. A* **40**: F383-F390.

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

[6] Bowman, A. W. (1984) An alternative method of cross-validation for the smoothing of density estimates, *Biometrika*, **71**: 353-360.

[7] Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (2004). Responses of single neurons in macaque MT/V5 as a function of motion coherence in stochastic dot stimuli. *Neural Signal Archive*, nsa2004.1. http://www.neuralsignal.org.

This poster was prepared with Brian Wolven's Poster LaTeX macros v2.1.