The empirical Bayes estimation of an instantaneous spike rate with a Gaussian process prior

NIPS 2006 Workshop in Whistler
Dynamical systems, stochastic processes and Bayesian inference

Shinsuke Koyama[1,2], Takeaki Shinokawa[1] and Shigeru Shinomoto[1]

[1] Department of Physics, Kyoto University

[2] Department of Statistics and CNBC, Carnegie Mellon University

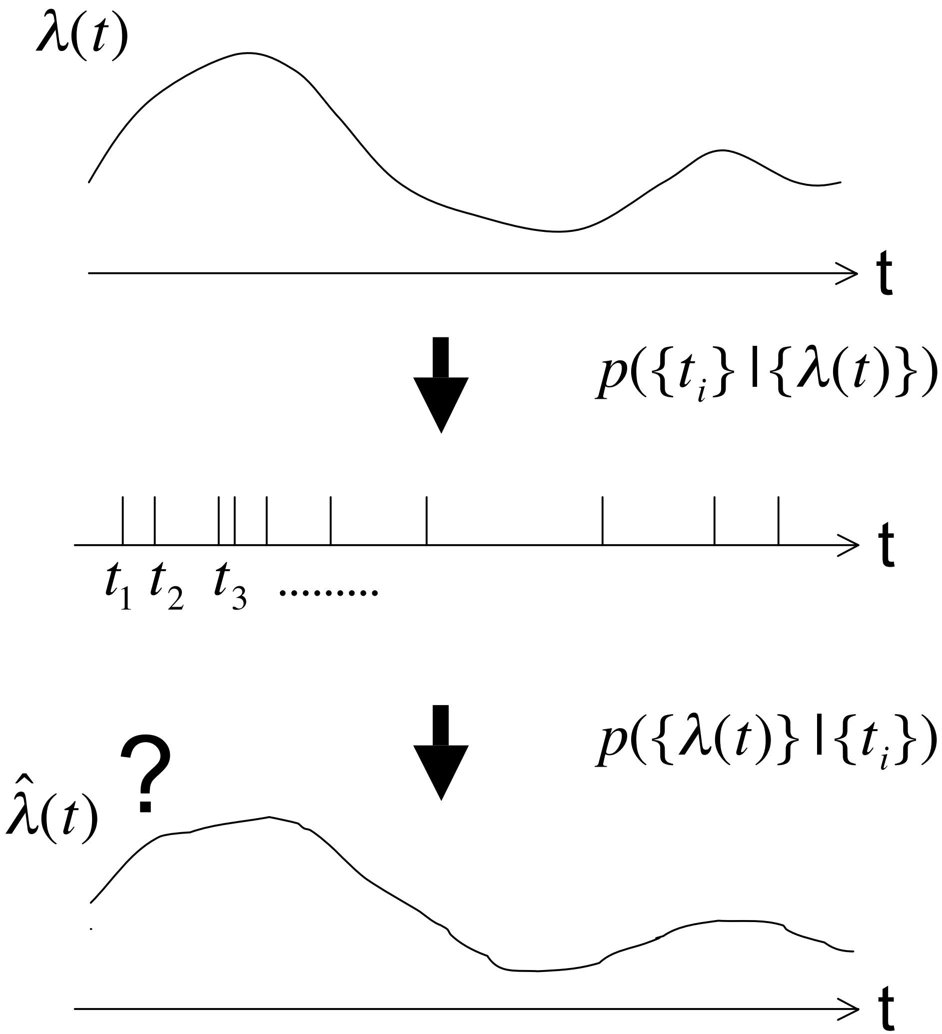
Contents

- Motivation and formulation of problem
- Theoretical analysis using a path integral method
- Practical algorithm (if I have enough time)
- Summary

Motivation

- Estimation of a time-dependent rate of point process: continuous time, non-Gaussian likelihood
- Application to analysis of neuronal spike data (my interest)

Inference problem



Bayesian inference

Likelihood function: time-dependent Poisson process

$$p(\{t_i\} \mid \{\lambda(t)\}) = \left[\prod_{i=1}^{n} \lambda(t_i)\right] \exp\left(-\int_0^T \lambda(t)dt\right)$$

Prior distribution of λ(t): Brownian motion

$$p(\{\lambda(t)\} \mid \Delta) = \frac{1}{Z(\Delta)} \exp\left[-\frac{\Delta^2}{2} \int_0^T \left(\frac{d\lambda(t)}{dt}\right)^2 dt\right]$$

Δ: smoothness hyperparameter

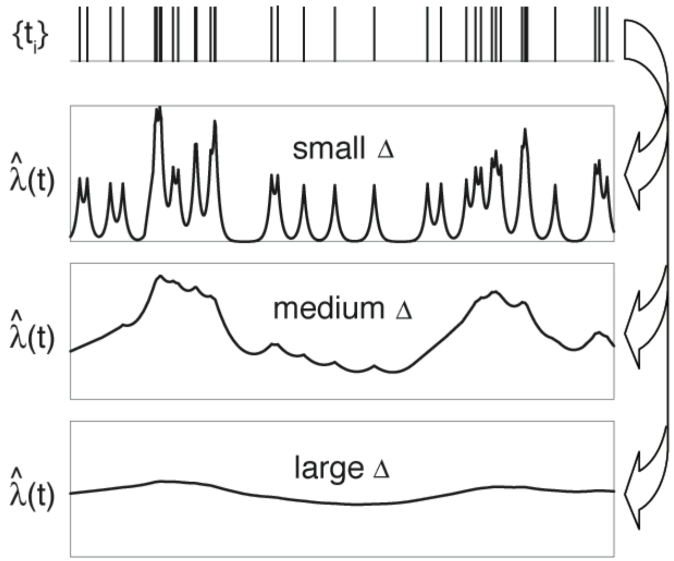
Posterior distribution

$$p(\{\lambda(t)\} \mid \{t_i\};\Delta) = \frac{p(\{t_i\} \mid \{\lambda(t)\})\, p(\{\lambda(t)\} \mid \Delta)}{p(\{t_i\} \mid \Delta)}$$

Hyperparameter selection

MAP estimation of λ(t):

$$\Delta^2 \frac{d^2\hat{\lambda}(t)}{dt^2} = 1 - \frac{1}{\hat{\lambda}(t)} \sum_{i=1}^{n} \delta(t - t_i)$$



Hyperparameter selection: maximizing the marginal likelihood

$$p(\{t_i\} \mid \Delta) = \int p(\{t_i\} \mid \{\lambda(t)\})\, p(\{\lambda(t)\} \mid \Delta)\, d\{\lambda(t)\}$$

$$\hat{\Delta} = \underset{\Delta}{\arg\max}\, p(\{t_i\} \mid \Delta)$$

Theoretical analysis: Path integral method

Marginal likelihood function:

$$p(\{t_i\} \mid \Delta) = \int p(\{t_i\} \mid \{\lambda(t)\}) p(\{\lambda(t)\} \mid \Delta) d\{\lambda(t)\}$$

$$= \int \exp\left[-\int L(\dot{\lambda}, \lambda, t) dt\right] d\{\lambda(t)\}$$

where

$$L(\dot{\lambda}, \lambda, t) = \frac{\Delta^2}{2} \dot{\lambda}^2 + \lambda - \sum_{i=1}^{n} \delta(t - t_i) \log \lambda$$ "Lagrangian"

The MAP estimate is obtained from:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\hat{\lambda}}}\right) - \frac{\partial L}{\partial \hat{\lambda}} = 0$$ "Euler-Lagrange equation"

Then, the marginal likelihood function can be evaluated as:

$$p(\{t_i\} \mid \Delta) \approx \exp\left[-\int L(\dot{\hat{\lambda}}, \hat{\lambda}, t) dt\right] \left[\underline{\det\left(\tfrac{\partial^2 L}{\partial \dot{\lambda}^2} \partial_t^2 + \tfrac{\partial^2 L}{\partial \lambda \partial \dot{\lambda}} \partial_t^2 - \tfrac{\partial^2 L}{\partial \lambda^2}\right)}\right]^{-\frac{1}{2}}$$

quadratic term

Evaluation of the marginal likelihood

Original rate function: $\lambda_0(t) = \mu + \sigma f(t/\tau)$

Under the condition: $\sigma/\mu << 1 \qquad T >> 1$

- $$\sum_{i=1}^{n} \delta(t - t_i) = \lambda_0(t) + \sqrt{\lambda_0(t)}\xi(t)$$

  where $\langle \xi(t) \rangle = 0, \ \langle \xi(t)\xi(t') \rangle = \delta(t - t')$

- $$\hat{\lambda}(t) = \mu + \frac{1}{2\Delta\sqrt{\mu}} \int_0^T \exp\left(-\frac{|t - s|}{\Delta\sqrt{\mu}} \right) (f(s/\tau) + \sqrt{\mu}\xi(s)) ds$$

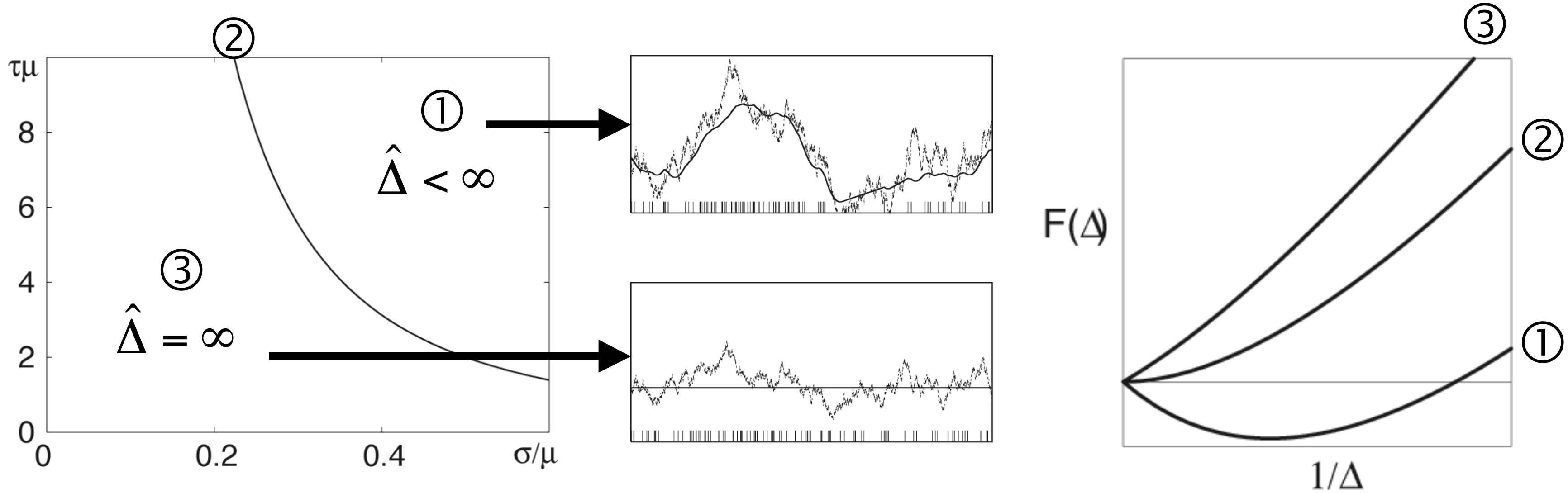
Free energy:

$$F(\Delta) = -\frac{1}{T} \log p(\{t_i\} \mid \Delta) = \frac{1}{4\Delta\sqrt{\mu}} \left(1 - \frac{2\tau\sigma^2}{\mu} \int_0^\infty \phi(u) e^{-\frac{\tau}{\Delta\sqrt{\mu}}u} du \right)$$

$\phi(u)$: correlation function of f(t)

Result 1: OU process

$$\lambda(t) = \mu + \sigma\eta(t) \qquad \langle \eta(t) \rangle = 0, \;\; \langle \eta(t)\eta(t') \rangle = e^{-|t-t'|/\tau}$$



Rate of convergence (MISE):

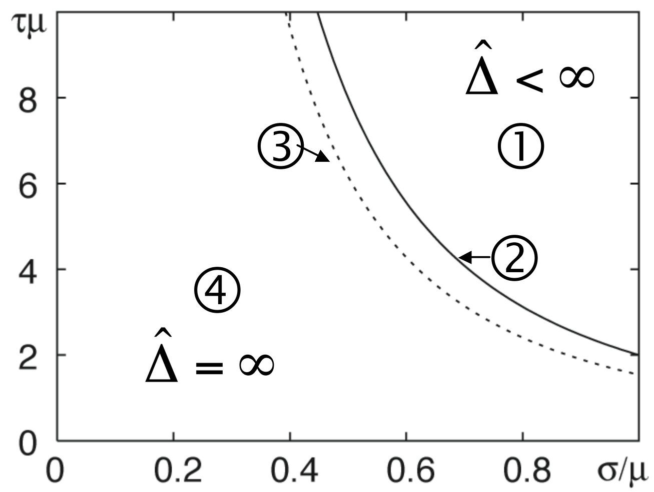
$$\text{MISE} \sim \frac{1}{\sqrt{2}} \sigma^2 z^{-\frac{1}{2}} \quad \text{as} \quad z \equiv \frac{\tau\sigma^2}{\mu} >> 1$$

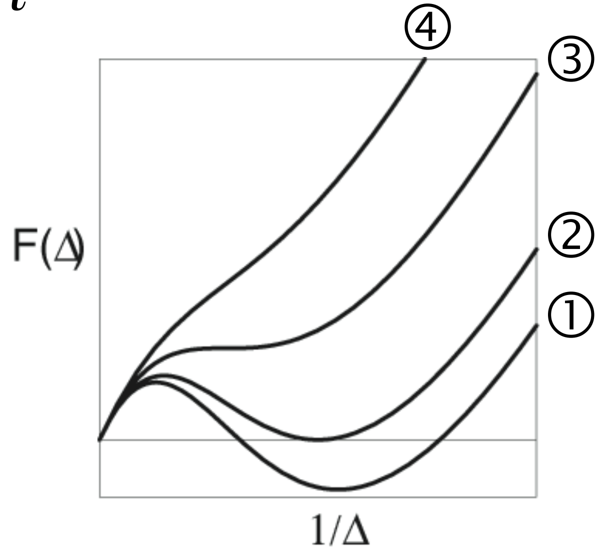
$$\text{MISE} \sim \frac{2}{\sqrt{3}} \sigma^2 z^{-\frac{1}{2}}$$ Optimal histogram

(Shimazaki and Shinomoto, NIPS'06)

Result 2: Sinusoidal modulation

$$\lambda(t) = \mu + \sigma \sin \frac{t}{\tau}$$





Rate of convergence (MISE)

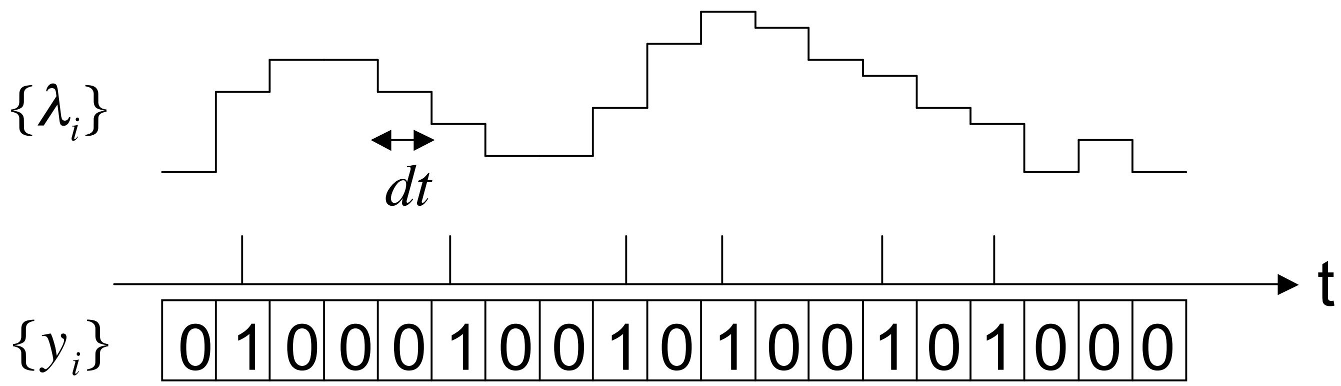
$$\mathrm{MISE} \sim 2^{-\frac{5}{3}} \sigma^2 z^{-\frac{2}{3}} \quad \text{as} \quad z \equiv \frac{\tau\sigma^2}{\mu} >> 1$$

$$\mathrm{MISE} \sim 2^{-\frac{5}{3}} 3^{\frac{2}{3}} \sigma^2 z^{-\frac{2}{3}}$$ Optimal histogram

(Shimazaki and Shinomoto, NIPS'06)

Practical algorithm (using EM)

Discrete time approximation:



State space model

Observation:

$$p(y_i \mid \lambda_i) = (\lambda_i dt)^{y_i} \exp(-\lambda_i dt)$$

Hidden process:

$$p(\lambda_{i+1} \mid \lambda_i; \Delta) = \frac{\Delta}{\sqrt{2\pi dt}} \exp\left[-\frac{(\lambda_{i+1} - \lambda_i)^2}{2dt/\Delta^2}\right]$$

Gaussian approximated EM algorithm

1. Set initial value $\Delta^{(0)}$ and $\Delta^{(l)} \leftarrow \Delta^{(0)}$.

2. E-step

   2-i. Filtering

Laplace approximation

$$p(\lambda_i \mid y_{1:i};\Delta^{(l)}) \propto p(y_i \mid \lambda_i)\,p(\lambda_i \mid y_{1:i-1};\Delta^{(l)}) \approx N(\lambda_{i/i}, v_{i/i})$$

$$p(\lambda_{i+1} \mid y_{1:i};\Delta^{(l)}) = \int p(\lambda_{i+1} \mid \lambda_i;\Delta^{(l)})\,p(\lambda_i \mid y_{1:i};\Delta^{(l)})\,d\lambda_{i+1} \approx N(\lambda_{i+1/i}, v_{i+1/i})$$ for i=1…N.

   2-ii. Smoothing

$$p(\lambda_i \mid y_{1:N};\Delta^{(l)}) = N(\lambda_{i/N}, v_{i/N})$$

where

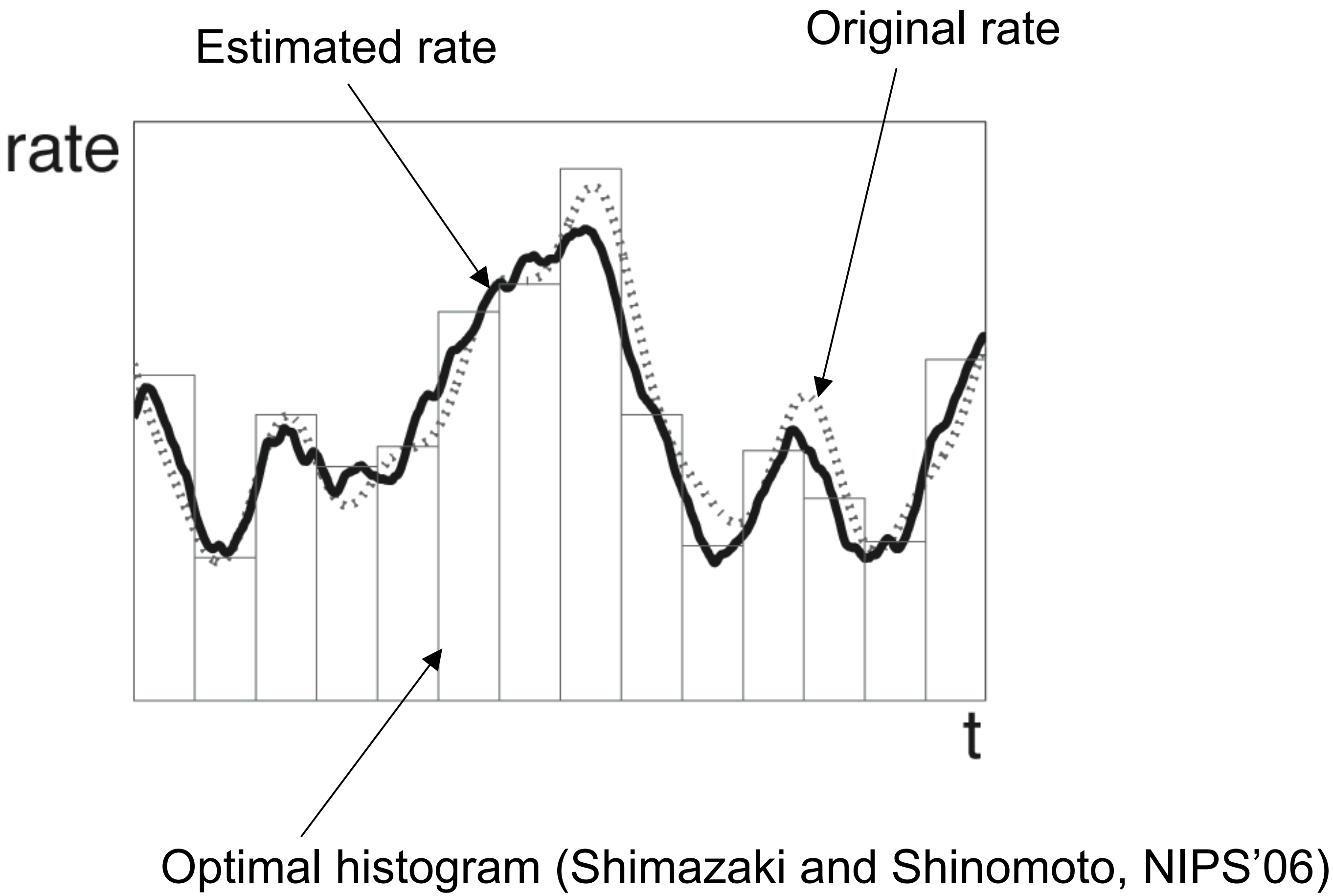
$$\lambda_{i/N} = v_{i/i}/v_{i+1/i}(\lambda_{i+1/N} - \lambda_{i+1/i})$$
$$v_{i/N} = (v_{i/i}/v_{i+1/i})^2(v_{i+1/N} - v_{i+1/i})$$ for i=N…1.

3. M-step

$$\Delta^{(l+1)} = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\mathrm{E}[(\lambda_{i+1} - \lambda_i)^2 \mid y_{1:N};\Delta^{(l)}]}{dt} \right)^{-1}$$

4. $\Delta^{(l)} \leftarrow \Delta^{(l+1)}$ and go to step 2.

Demonstration



Summary

- Empirical Bayes inference of a time-dependent rate of Poisson process
- Theoretical Analysis (path integral method)
- Practical algorithm (if I had enough time)
- Future works: non-Poisson model (see Koyama and Shinomoto, J.Phys.A, 2005), other Gaussian process prior

Acknowledges

- Thanks to JSPS, organizers of this workshop, and the great mountain in Whistler!!