

IV Kongres Statystyki Polskiej

2-4 lipca 2024 r., Warszawa



Bayes Risk Consistency of Nonparametric Classification Rules

for

Point Processes

Mirosław Pawlak

UM&AGH

mirek.pawlakk@gmail.com

Outline

I Spike Trains Data(Spiking NN) and Point Processes

II Classification for Point Processes

1. Bayes Rule and Risk

2. Nonparametric Plug-In Classification Rules

3. Kernel Classifiers

III Extensions

I Spike Trains Data(Spiking NN) and Point Processes

Spike Trains Data

Biological Systems

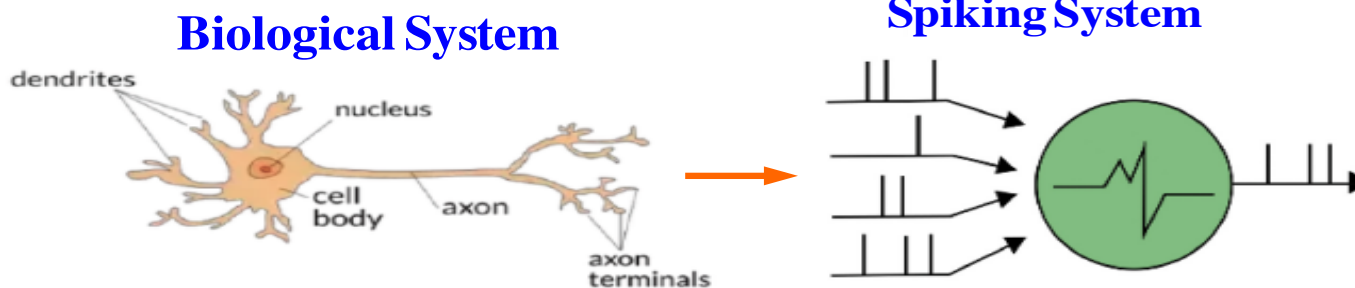
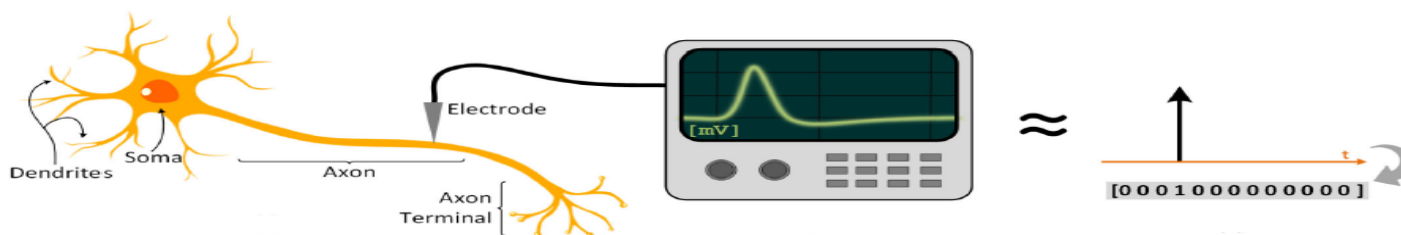
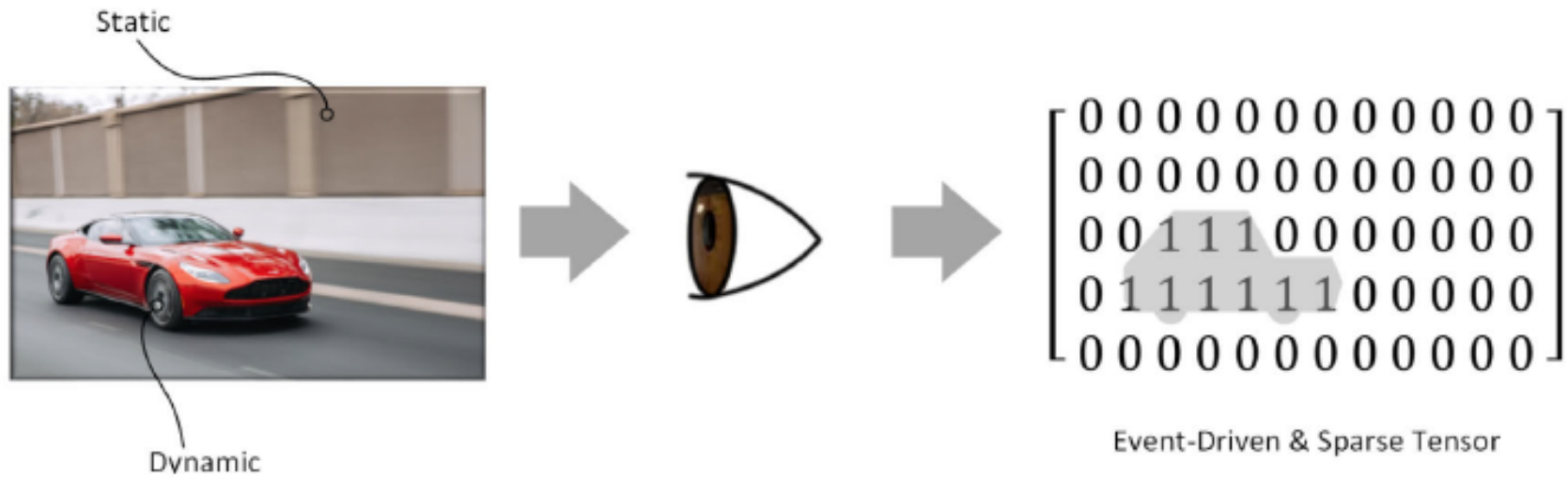
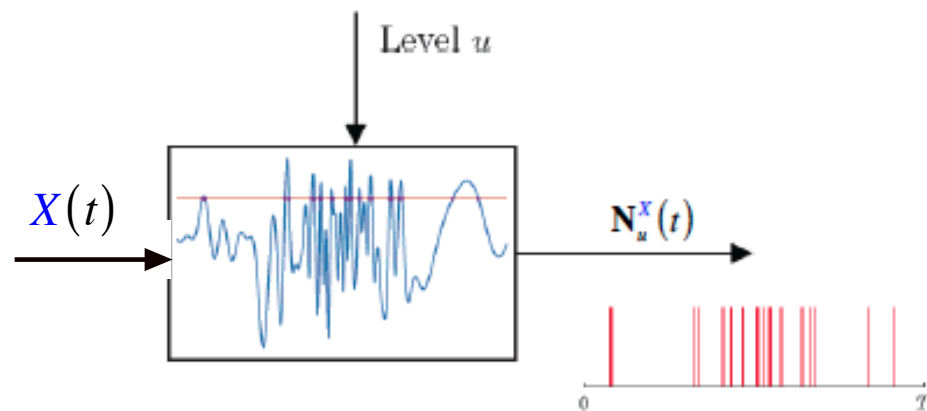
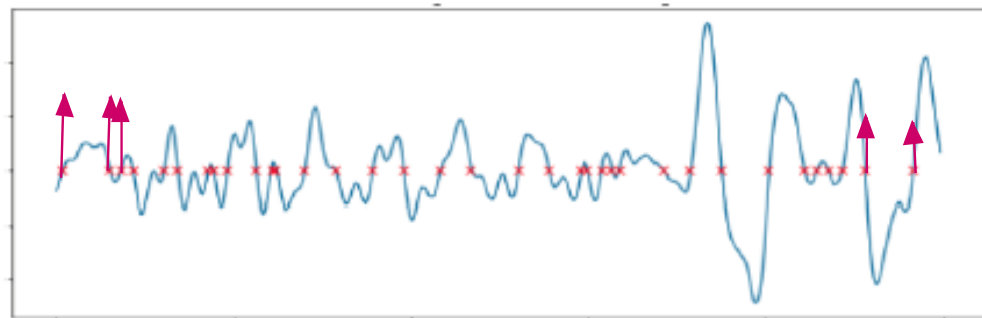
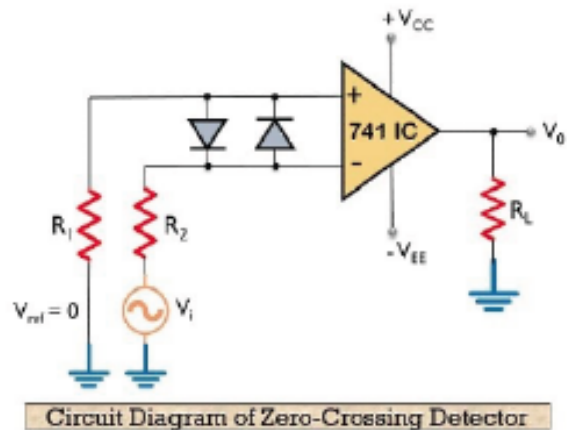
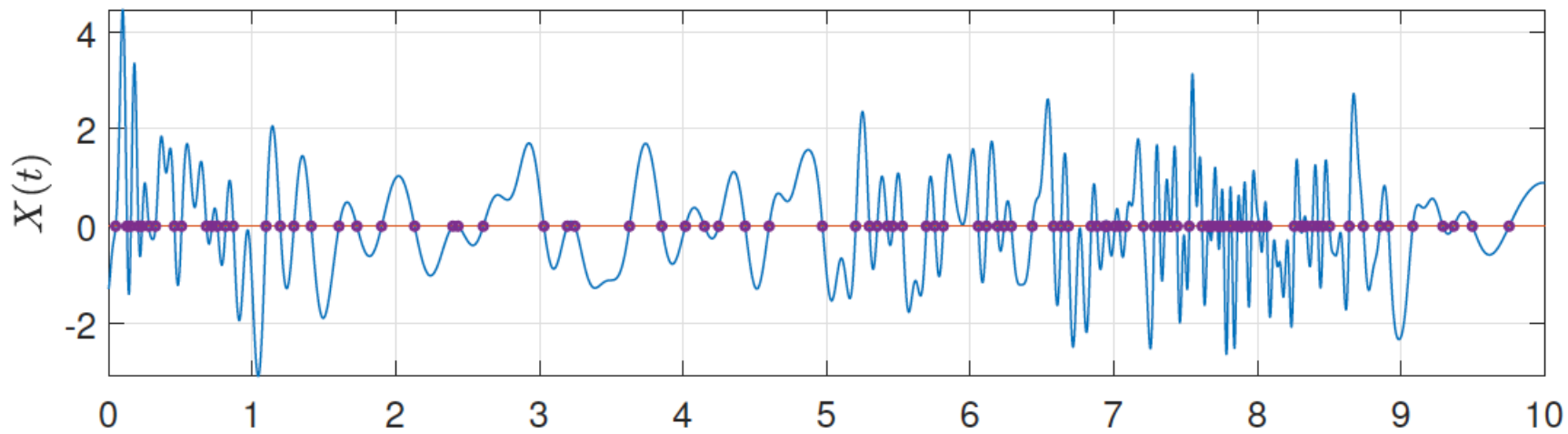


Image Analysis

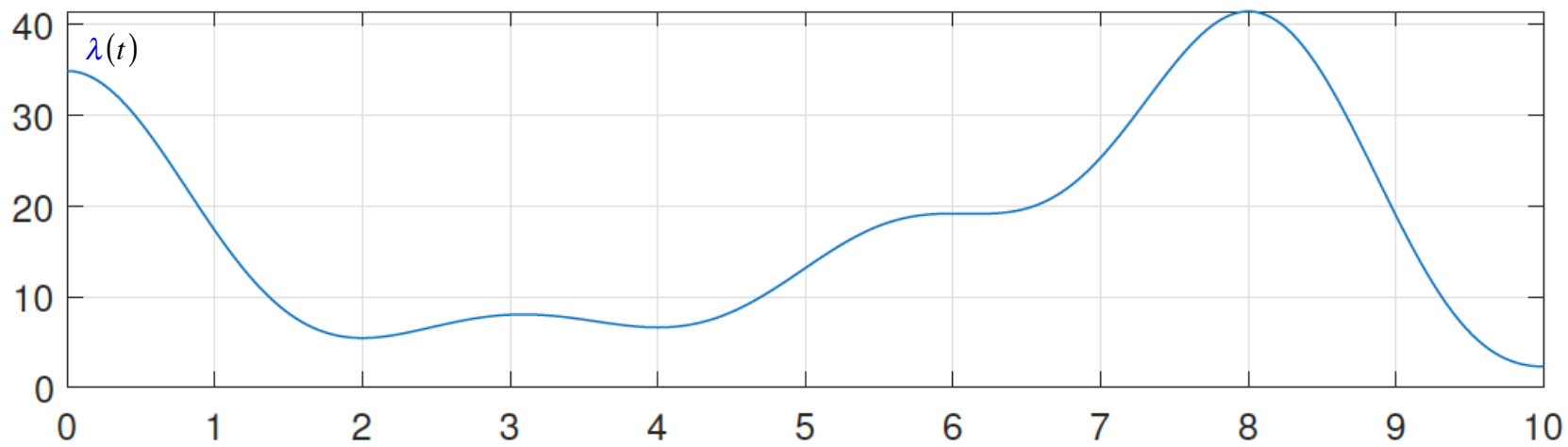


Event Driven Signal Analysis: Level Crossings

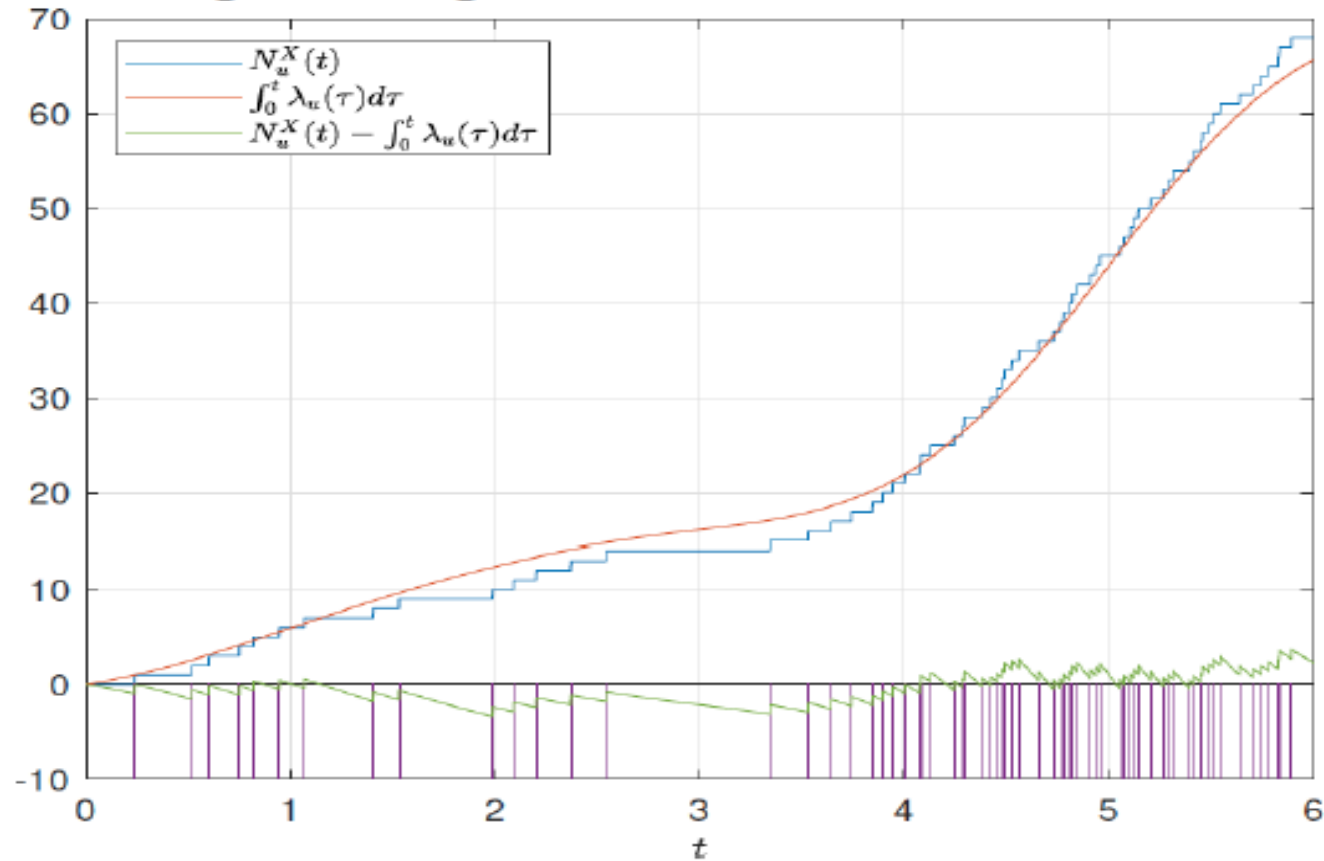




Local Intensity



▲ Level Crossings Counting Process



- Inference from Level Crossings

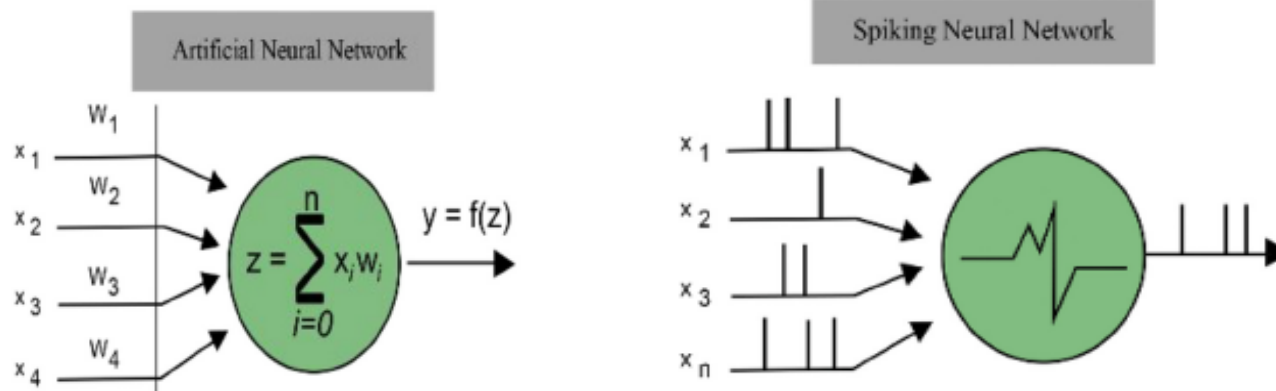
Lindgren: *Statistical Science*, 2019

Pawlak: *IEEE Signal Processing*, 2023

Spiking Neural Networks

Networks of Spiking Neurons: The Third Generation of Neural Network Models

Wolfgang Maass



2022 roadmap on neuromorphic computing and engineering

Dennis V Christensen¹ , Regina Dittmann², Bernabe Linares-Barranco³,
Abu Sebastian⁴ , Manuel Le Gallo⁴ , Andrea Redaelli⁵, Stefan Slesazeck⁶ , Thomas
Mikolajick^{6,7} , Sabina Spiga⁸ , Stephan Menzel⁹, Ilia Valov⁹ ,
Gianluca Milano¹⁰ , Carlo Ricciardi¹¹ , Shi-Jun Liang¹² , Feng Miao¹² ,
Mario Lanza¹³ , Tyler J Quill¹⁴, Scott T Keene¹⁵ , Alberto Salleo¹⁴, Julie Grollier¹⁶,
Danijela Marković¹⁶, Alice Mizrahi¹⁶ , Peng Yao¹⁷, J Joshua Yang¹⁷ ,
Giacomo Indiveri¹⁸ , John Paul Strachan¹⁹ , Suman Datta²⁰, Elisa Vianello²¹ ,
Alexandre Valentian²², Johannes Feldmann¹, Xuan Li²³, Wolfram H P Pernice^{24,25},
Harish Bhaskaran²³, Steve Furber²⁶, Emre Neftci²⁷, Franz Scherr²⁸,
Wolfgang Maass²⁸, Srikanth Ramaswamy²⁹ , Jonathan Tapson³⁰,
Priyadarshini Panda³¹, Youngeun Kim³¹, Gouhei Tanaka³², Simon Thorpe³³,
Chiara Bartolozzi³⁴, Thomas A Cleland³⁵, Christoph Posch³⁶, ShihChii Liu¹⁸,
Gabriella Panuccio³⁷ , Mufti Mahmud³⁸ , Arnab Neelim Mazumder³⁹,
Morteza Hosseini³⁹, Tinoosh Mohsenin³⁹, Elisa Donati¹⁸ , Silvia Tolu¹⁴ ,
Roberto Galeazzi⁴⁰, Martin Ejsing Christensen⁴¹, Sune Holm⁴², Daniele Ielmini⁴³  and
N Pryds¹ 

Training Spiking Neural Networks Using Lessons From Deep Learning

Spike Trains Data/Spiking NN \longleftrightarrow Statistics

▲ Neural Networks

attack problems with 10 million parameters and try to get an answer

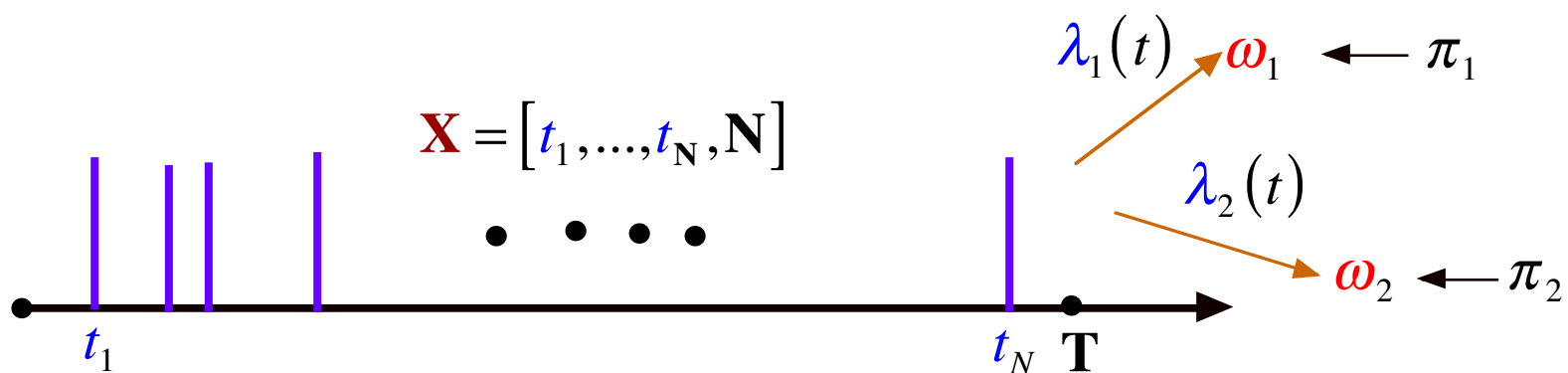
▲ Statistics

attack problems with a simple model and try to get it right

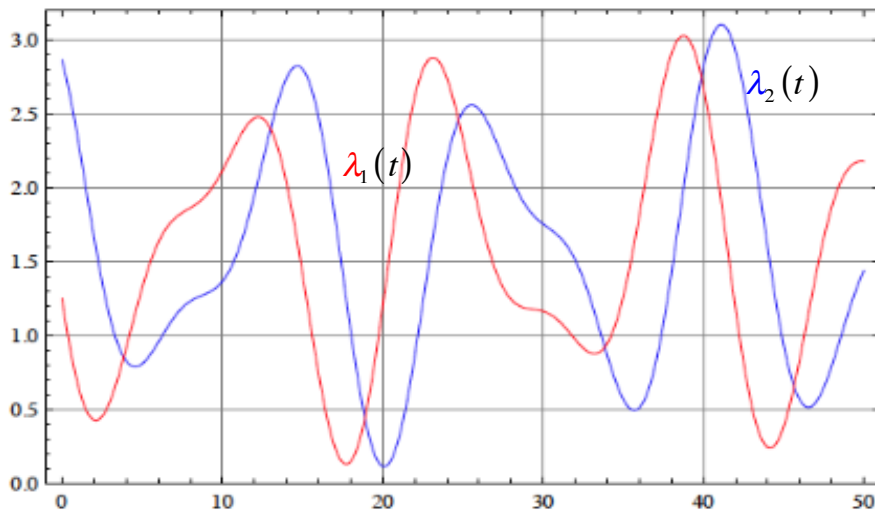
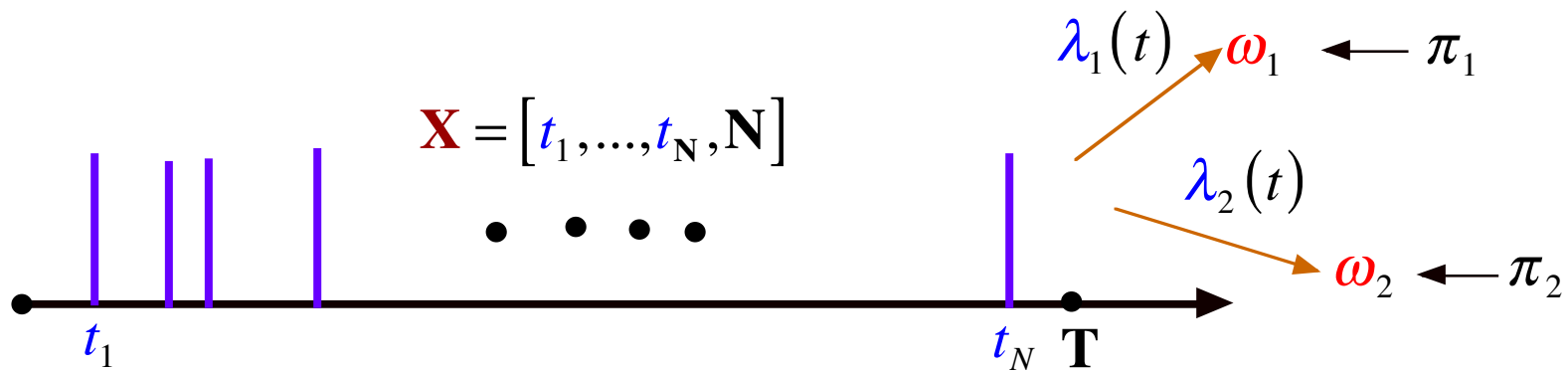
II Classification for Point Processes (Spike Trains Data)

1. Bayes Rule and Risk

● Classification Problem



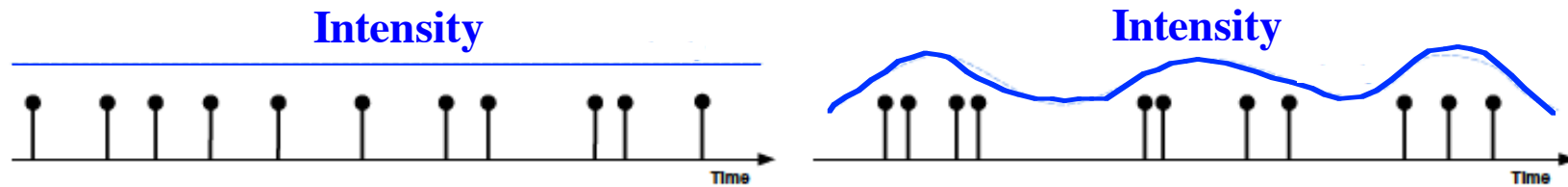
Classification Problem



$$\Pr[\text{spike in } (t, t + \Delta t) | \mathbf{X} \in \omega_1] \approx \lambda_1(t) \Delta t$$

$$\Pr[\text{spike in } (t, t + \Delta t) | \mathbf{X} \in \omega_2] \approx \lambda_2(t) \Delta t$$

Intensity Function



$$\mathbf{E}[\mathbf{N}(t, t + \Delta t) | \mathbf{H}(t)] \approx \lambda(t | \mathbf{H}(t)) \Delta t$$

$\mathbf{H}(t)$ - the history of all events up to time t

The Likelihood Function

The occurrence density of counting process $\mathbf{N}(t)$ with the intensity function $\lambda(t)$

$$\mathbf{X} = [t_1, \dots, t_N, \mathbf{N}]$$

$$\boxed{\mathbf{N} = \mathbf{N}(T)}$$

$$\boxed{f(\mathbf{x}) = \prod_{i=1}^{\mathbf{N}} \lambda(t_i) \exp\left(-\int_0^T \lambda(u) du\right)}$$

*Daley, Vere-Jones: **An Introduction to the Theory of Point Processes**

Bayes Rule

▲ Bayes Rule in Terms of Local Intensities

$$\boxed{\pi_1 = \pi_2}$$

$$\psi_T^* : \mathbf{X} \in \omega_1 \text{ if } \sum_{i=1}^N \log \left(\frac{\lambda_1(t_i)}{\lambda_2(t_i)} \right) \geq \gamma$$

↑

$$\gamma = \int_0^T (\lambda_1(u) - \lambda_2(u)) du$$

$$\bullet \gamma = \int_0^T (\lambda_1(u) - \lambda_2(u)) du + \log \left(\frac{\pi_2}{\pi_1} \right) \bullet$$

▲ Bayes Rule in Terms of Stochastic Integrals

$$\psi_T^* : \mathbf{X} \in \omega_1 \quad \text{if} \quad \int_0^T \log \left(\frac{\lambda_1(t)}{\lambda_2(t)} \right) d\mathbf{N}(t) \geq \gamma$$

- A1:**
- $\lambda_i(t)$ - deterministic functions, $i = 1, 2$
 - $d\mathbf{N}_i(t) = \lambda_i(t)dt + d\mathbf{M}_i(t)$
 - $\mathbf{M}_i(t) = \mathbf{N}_i(t) - \int_0^t \lambda_i(s)ds$ - martingale
- ↓
- Poisson processes with intensity functions $\lambda_i(t)$, $i = 1, 2$. (Watanabe Th.)

▲ Bayes Rule in Terms of Stochastic Integrals

$$\psi_T^* : \mathbf{X} \in \omega_1 \text{ if } U_T(\mathbf{X}) \geq \alpha_T$$

- $U_T(\mathbf{X}) = \int_0^T \log\left(\frac{\lambda_1(t)}{\lambda_2(t)}\right) d\mathbf{M}(t)$

zero mean **martingale**

- $\alpha_T = \int_0^T (\lambda_1(t) - \lambda_2(t)) dt + \int_0^T \log\left(\frac{\lambda_1(t)}{\lambda_2(t)}\right) \lambda(t) dt$

$$\lambda(t) = \begin{cases} \lambda_1(t) & \text{if } \mathbf{X} \in \omega_1 \\ \lambda_2(t) & \text{if } \mathbf{X} \in \omega_2 \end{cases}$$

▲ Shape Densities & Intensity Factors

$$\bullet \lambda_1(t) = \frac{p_1(t) \lambda_1(t)}{\int_0^T \lambda_1(u) du} \tau_1$$

$$\tau_1 = \mathbf{E}[\mathbf{N}_1(T)]$$

$$\bullet \lambda_2(t) = p_2(t) \tau_2$$

$$\tau_2 = \mathbf{E}[\mathbf{N}_2(T)]$$

Example: $\lambda_1(t) = \lambda(t)$ and $\lambda_2(t) = \mu\lambda(t)$

$$\mu > 0$$

$$\Rightarrow p_1(t) = p_2(t), \tau_2 = \mu\tau_1$$

▲ Bounds

$$U_T(\mathbf{X}) \geq \alpha_T$$

$$\mathbf{K}_T(p \parallel q) = \int_0^T \log\left(\frac{p(t)}{q(t)}\right) p(t) dt$$

$\mathbf{K}_T(\cdot \parallel \cdot)$: KL divergence

- $\mathbf{X} \in \omega_1$: $-\frac{(\tau_1 - \tau_2)^2}{\tau_2} - \tau_1 \mathbf{K}_T(p_1 \parallel p_2) \leq \alpha_T \leq -\tau_1 \mathbf{K}_T(p_1 \parallel p_2)$
- $\mathbf{X} \in \omega_2$: $\tau_2 \mathbf{K}_T(p_2 \parallel p_1) \leq \alpha_T \leq \frac{(\tau_1 - \tau_2)^2}{\tau_1} + \tau_2 \mathbf{K}_T(p_2 \parallel p_1)$

Note: $\tau_1 = \tau_2 = \tau \Rightarrow$

$$\omega_1: \alpha_T = -\tau \mathbf{K}_T(p_1 \parallel p_2)$$

$$\omega_2: \alpha_T = \tau \mathbf{K}_T(p_2 \parallel p_1)$$

▲ **Bounds**

$$U_T(\mathbf{X}) \geq \alpha_T$$

$$U_T(\mathbf{X}) = \int_0^T \log \left(\frac{\lambda_1(t)}{\lambda_2(t)} \right) dM(t)$$

$$\mathbf{V}_T(p \parallel q) = \int_0^T \log^2 \left(\frac{p(t)}{q(t)} \right) p(t) dt: \text{KL variation}$$

$$\mathbf{K}_T(p \parallel q) \leq \sqrt{\mathbf{V}_T(p \parallel q)}$$

- $\mathbf{X} \in \omega_1$:

$$\tau_1 \left\{ \mathbf{K}_T(p_1 \parallel p_2) + \log \left(\frac{\tau_1}{\tau_2} \right) \right\}^2 \leq \text{var}[U_T(\mathbf{X})] \leq \tau_1 \left\{ \sqrt{\mathbf{V}_T(p_1 \parallel p_2)} + \log \left(\frac{\tau_1}{\tau_2} \right) \right\}^2$$

- $\mathbf{X} \in \omega_2$:

$$\tau_2 \left\{ \mathbf{K}_T(p_2 \parallel p_1) + \log \left(\frac{\tau_2}{\tau_1} \right) \right\}^2 \leq \text{var}[U_T(\mathbf{X})] \leq \tau_2 \left\{ \sqrt{\mathbf{V}_T(p_2 \parallel p_1)} + \log \left(\frac{\tau_2}{\tau_1} \right) \right\}^2$$

Note:

A2: $0 < \delta \leq \lambda_i(t) \leq C$, for all $t \geq 0$

- $\mathbf{X} \in \omega_1$: $\tau_1 \log^2\left(\frac{\delta}{C}\right) \leq \text{var}[U_T(\mathbf{X})] \leq \tau_1 \log^2\left(\frac{C}{\delta}\right)$

- $\mathbf{X} \in \omega_2$: $\tau_2 \log^2\left(\frac{\delta}{C}\right) \leq \text{var}[U_T(\mathbf{X})] \leq \tau_2 \log^2\left(\frac{C}{\delta}\right)$

▲ Asymptotic

$$\mathbf{A3}: \frac{1}{T} \int_0^T \lambda_i(s) ds \rightarrow d > 0 \text{ as } T \rightarrow \infty, i = 1, 2$$

$$d \log^2 \left(\frac{\delta}{C} \right) \leq \underline{\lim}_T \text{var} \left[\frac{1}{\sqrt{T}} U_T(\mathbf{X}) \right] \leq \overline{\lim}_T \text{var} \left[\frac{1}{\sqrt{T}} U_T(\mathbf{X}) \right] \leq d \log^2 \left(\frac{C}{\delta} \right)$$

Bayes Risk

$$\mathbf{R}_T^* = \frac{1}{2} \mathbf{P}[U_T(\mathbf{X}) \geq \alpha_T \mid \mathbf{X} \in \omega_2] + \frac{1}{2} \mathbf{P}[U_T(\mathbf{X}) < \alpha_T \mid \mathbf{X} \in \omega_1]$$

Lemma (Le Guevel, 2021, CRM): Let $\mathbf{N}(t)$ be the counting process allowing the martingale decomposition

$$d\mathbf{N}(t) = \lambda(t)dt + d\mathbf{M}(t)$$

$$\mathbf{P}\left[\left|\int_0^T g(t)d\mathbf{M}(t)\right| \geq \varepsilon\right] \leq 2 \exp\left[-\frac{\varepsilon^2}{2v_T + \varepsilon u_T}\right]$$

$$\int_0^T g^2(t)\lambda(t)dt \leq v_T$$

$$|g(t)| \leq u_T$$

Theorem 1

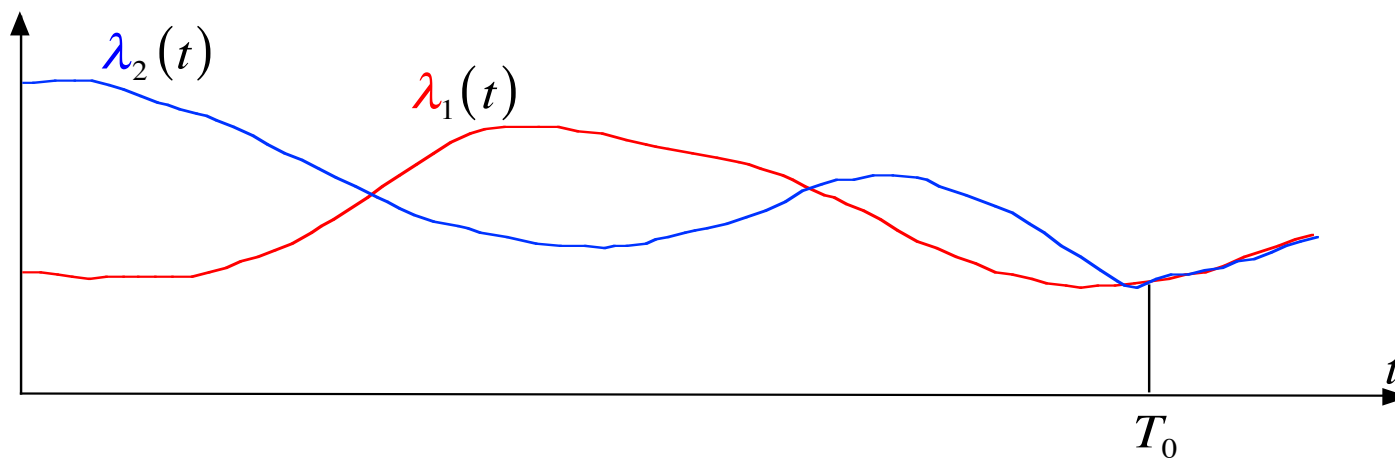
- Under **A1** + **A2** + **A3** + **A4**

$$\mathbf{R}_T^* \prec \frac{1}{2} \exp(-C_T T)$$

- $C_T \approx \frac{1}{3}d$

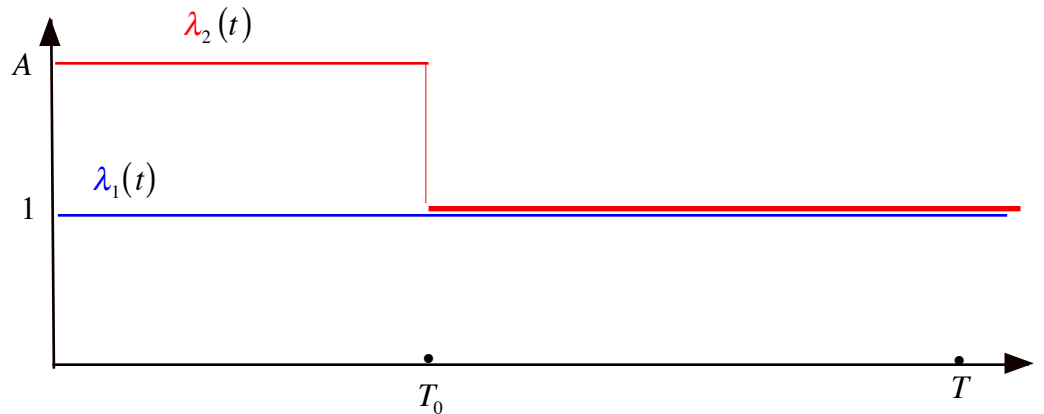
Assumption A4

$$\int_{T_0}^{\infty} |\lambda_1(s) - \lambda_2(s)| ds > 0$$



$$\int_{T_0}^{\infty} |\lambda_1(s) - \lambda_2(s)| ds = 0$$

Example



$$\mathbf{X} \in \omega_2 \quad \text{if} \quad \prod_{i=1}^{N(T)} \frac{\lambda_2(t_i)}{\lambda_1(t_i)} \exp \left[\int_0^T \lambda_1(s) ds - \int_0^T \lambda_2(s) ds \right] \geq 1$$

$$\alpha(T) = \mathbf{P} \left[\prod_{i=1}^{N(T)} \frac{\lambda_2(t_i)}{\lambda_1(t_i)} \exp \left[\int_0^T \lambda_1(s) ds - \int_0^T \lambda_2(s) ds \right] \geq 1 \mid \mathbf{X} \in \omega_1 \right]$$

$$\alpha(T) = \mathbf{P} \left[N(T_0) > \frac{(A-1)T_0}{\log(A)} \mid \mathbf{X} \in \omega_1 \right]$$

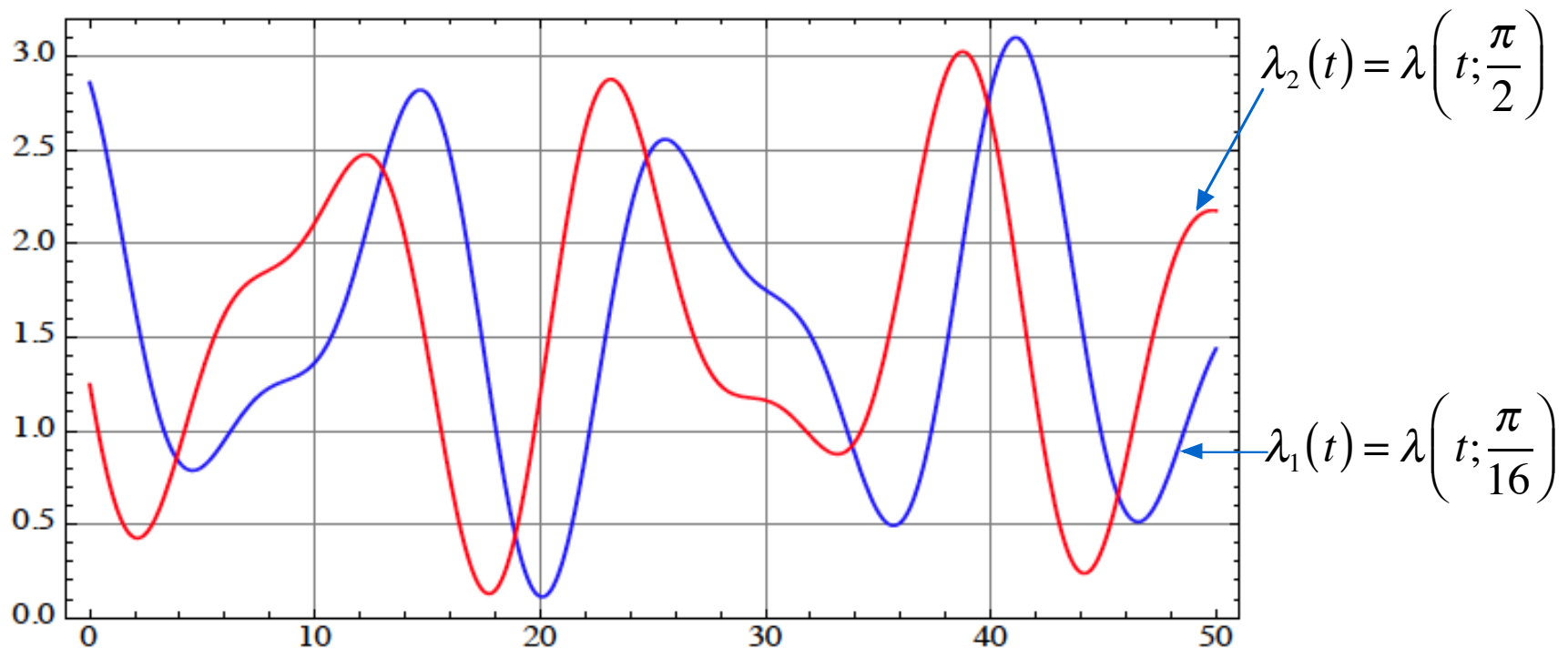
⇓

$$\alpha(T) = \sum_{j > \frac{(A-1)T_0}{\log(A)}} \frac{1}{j!} T_0^j e^{-T_0}$$

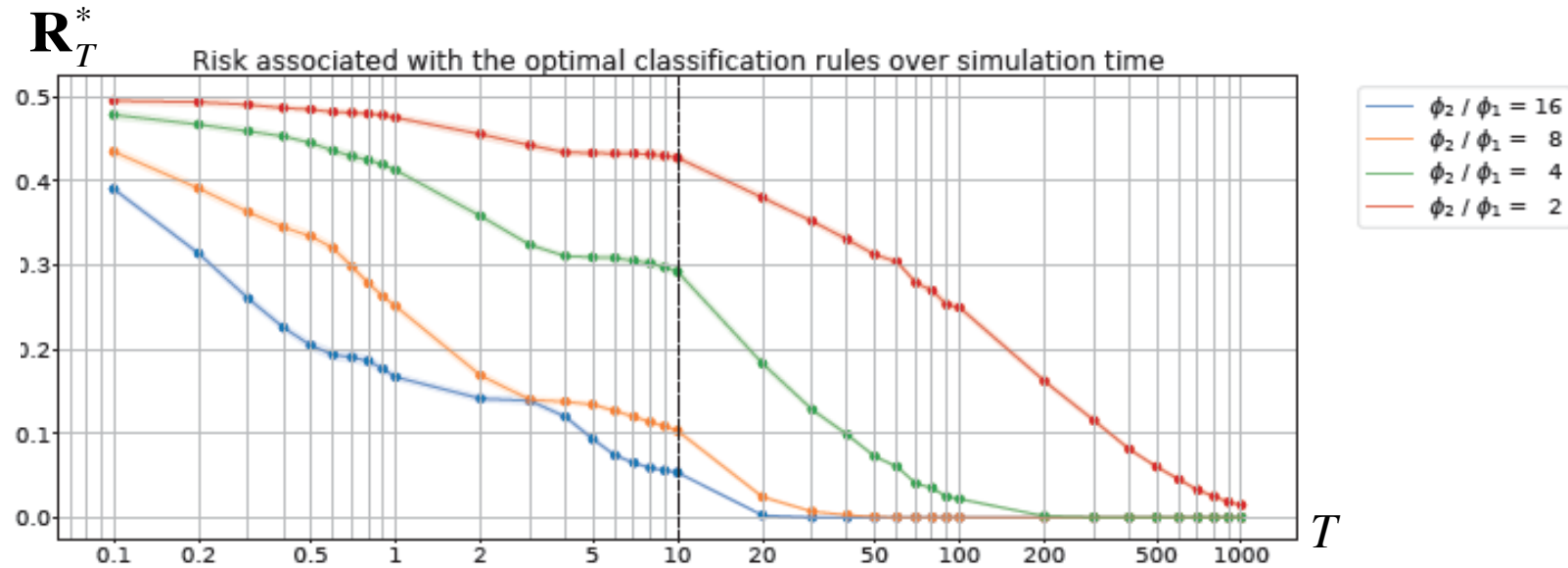
$\alpha(T)$ is a constant for $T > T_0$

Example: • $\lambda(t; \phi) = 1.6 + \cos\left(\frac{\pi}{4\sqrt{3}}t + \phi\right) + 0.5 \cos\left(\frac{\pi}{3\sqrt{2}}t + \frac{\pi}{4} + \phi\right)$

$$\lambda_1(t) = \lambda(t; \phi_1), \lambda_2(t) = \lambda(t; \phi_2)$$



$$\lambda_1(t) = \lambda(t; \phi_1), \lambda_2(t) = \lambda(t; \phi_2)$$



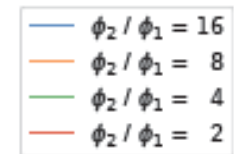
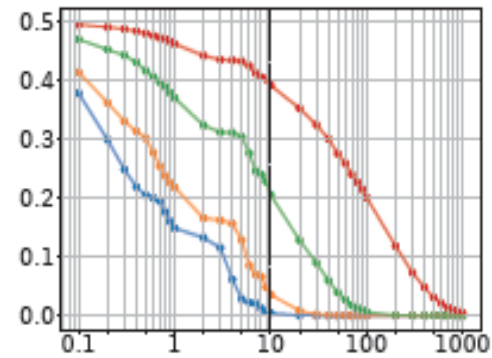
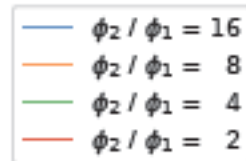
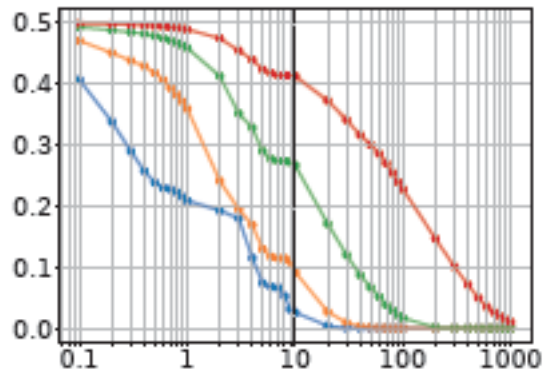
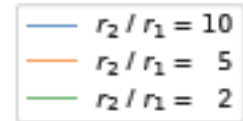
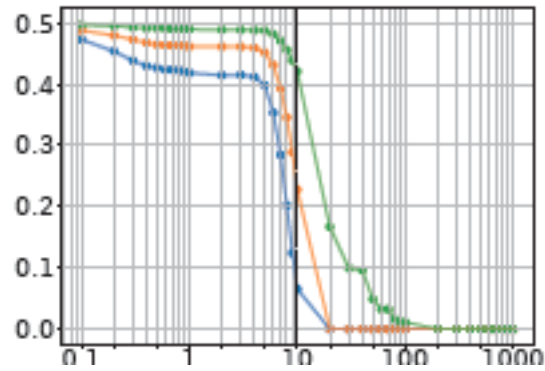
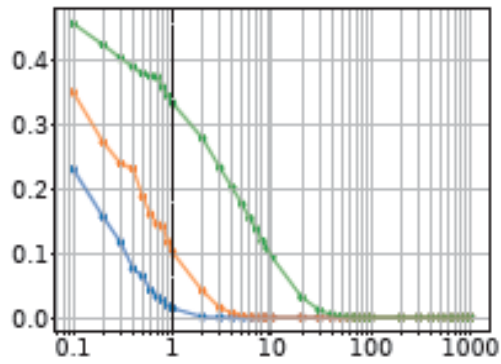
- The slow decay of \mathbf{R}_T^* for **close** intensities: $\phi_2 / \phi_1 = 2$.
- The fast rate for **distant** intensities: $\phi_2 / \phi_1 = 16$.

- $$\lambda_1(t) = \lambda(t; \phi_1), \lambda_2(t) = \lambda(t; \phi_2)$$

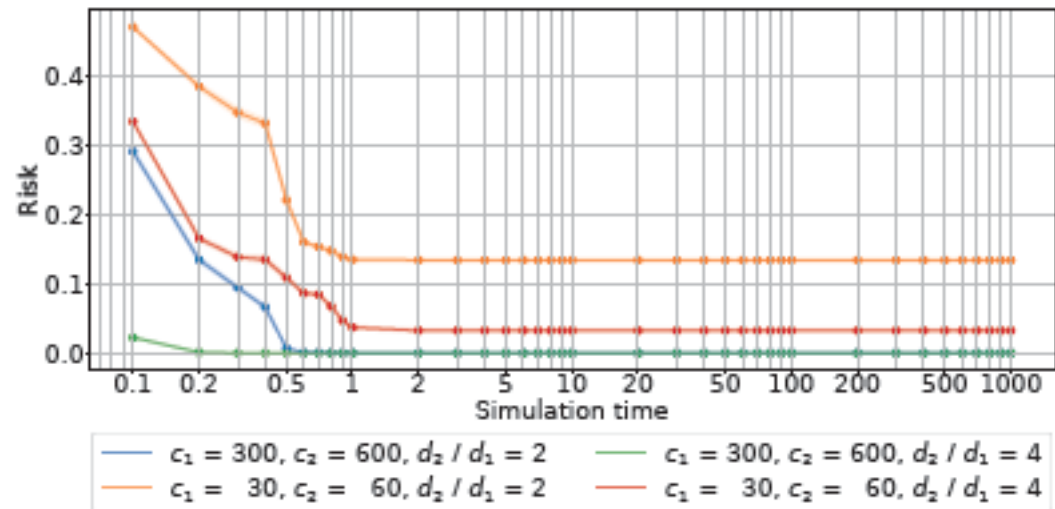
$$\lambda_1(t) = \lambda(t; r_1), \lambda_2(t) = \lambda(t; r_2)$$

a)	$\lambda(t; r) = r$
b)	$\lambda(t; r) = (r + 1) + r \cos\left(\frac{2\pi}{30}t + 2.6\right) + \cos\left(\frac{2\pi}{28}t + 4.5\right)$
c)	$\lambda(t; \phi) = \left[3.1 + 3 \cos\left(\frac{\pi}{3\sqrt{2}}t + \phi\right)\right]^{1/2}$
d)	$\lambda(t; \phi) = 1.3 \exp\left[\cos\left(\frac{\pi}{3\sqrt{2}}t + \phi\right)\right]$

R_T^* versus T



- $\lambda(t; c, d) = c \exp[-d(t - 0.5)^2] \leftarrow \int_0^\infty \lambda_i(t) dt < \infty$



- $\underline{\lim}_T \mathbf{R}_T^* > 0$

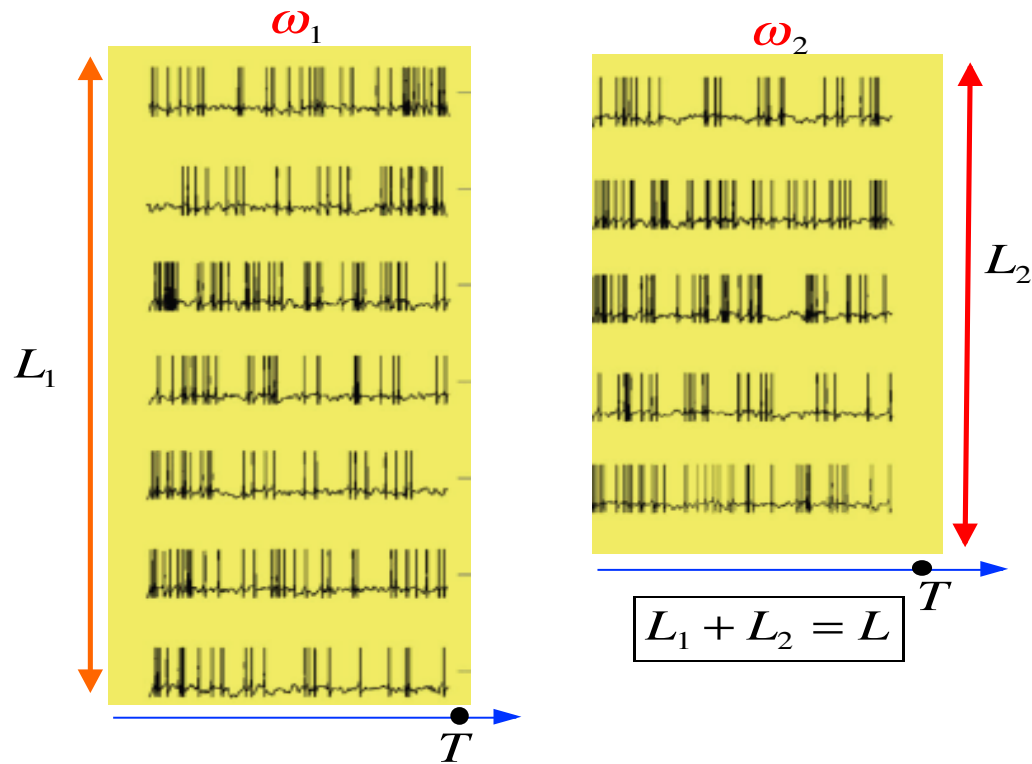
3. Nonparametric Classification Rules

▲ **Bayes Rule** $\psi_T^* : \mathbf{X} \in \omega_1$ if $\mathbf{W}_T(\mathbf{X}) \geq \eta_T$

- $\eta_T = \tau_1 - \tau_2 + N \log \left(\frac{\tau_2}{\tau_1} \right)$

- $\mathbf{W}_T(\mathbf{X}) = \sum_{i=1}^N \log \left(\frac{p_1(t_i)}{p_2(t_i)} \right)$

▲ Training Set: $\mathbf{D}_L = \{(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_L, Y_L)\}$



$$\mathbf{X}_j = [t_1^{[j]}, \dots, t_{N^{[j]}}^{[j]}, \mathbf{N}^{[j]}] , \quad Y_j \in \{\omega_1, \omega_2\}$$

▲ Plug-In Rules

$$\hat{\psi}_{L,T} : \mathbf{X} \in \omega_1 \text{ if } \hat{\mathbf{W}}_{L,T}(\mathbf{X}) \geq \hat{\eta}_{L,T}$$

$$\mathbf{X} = [t_1, \dots, t_N, \mathbf{N}]$$

- $\hat{\eta}_{L,T} = \hat{\tau}_1 - \hat{\tau}_2 + N \log \left(\frac{\hat{\tau}_2}{\hat{\tau}_1} \right)$

- $\hat{\mathbf{W}}_{L,T}(\mathbf{X}) = \sum_{i=1}^N \log \left(\frac{\hat{p}_1(t_i)}{\hat{p}_2(t_i)} \right)$

▲ Estimates Based on Data Aggregation

- $\hat{\tau}_1 = \frac{1}{L_1} \sum_{j=1}^L N^{[j]} \mathbf{1}(Y_j = \omega_1)$

$$\hat{\tau}_2 = \frac{1}{L_2} \sum_{j=1}^L N^{[j]} \mathbf{1}(Y_j = \omega_2)$$

- $\hat{p}_1(t) = \frac{1}{L_1} \sum_{j=1}^L \hat{p}_1^{[j]}(t) \mathbf{1}(Y_j = \omega_1)$

$$\hat{p}_2(t) = \frac{1}{L_2} \sum_{j=1}^L \hat{p}_2^{[j]}(t) \mathbf{1}(Y_j = \omega_2)$$

▲ Bayes Risk Consistency

- The Conditional Risk

$$\mathbf{R}_{L,T} = \mathbf{P} \left[\hat{\psi}_{L,T}(\mathbf{X}) \neq Y \mid \mathbf{D}_L \right]$$

- BRC

$$\mathbf{R}_{L,T} \rightarrow \mathbf{R}_T^* \text{ as } L \rightarrow \infty \text{ (} P \text{)}$$

Theorem 2

Consider the plug-in classification rules such that

$$\sup_{t \in [0, T]} \left| \hat{p}_i(t) - p_i(t) \right| \rightarrow 0 \text{ as } L \rightarrow \infty \quad (P) \quad i = 1, 2$$

Then

$$\mathbf{R}_{L, T} \rightarrow \mathbf{R}_T^* \text{ as } L \rightarrow \infty \quad (P)$$

Note: Helly's Theorem for the Stieltjes Integral

$$f_n(x) \rightarrow f(x) \text{ uniformly on } [0, T] \text{ and } g \in \mathbf{BV}[0, T]$$

$$\Rightarrow \int_0^T f_n(x) dg \rightarrow \int_0^T f(x) dg$$

4. Kernel Classifiers

- The kernel estimate from the single realization

$$\hat{p}_i^{[j]}(t) = \frac{1}{N^{[j]}} \sum_{l=1}^{N^{[j]}} K_h(t - t_l^{[j]})$$

$$K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right)$$

- The kernel estimate from the aggregated data

$$\hat{p}_i(t) = \frac{1}{L_i} \sum_{j=1}^{L_i} \hat{p}_i^{[j]}(t) \mathbf{1}(Y_j = \omega_i), \quad i = 1, 2$$

- **Uniform Convergence for the Kernel Estimate**

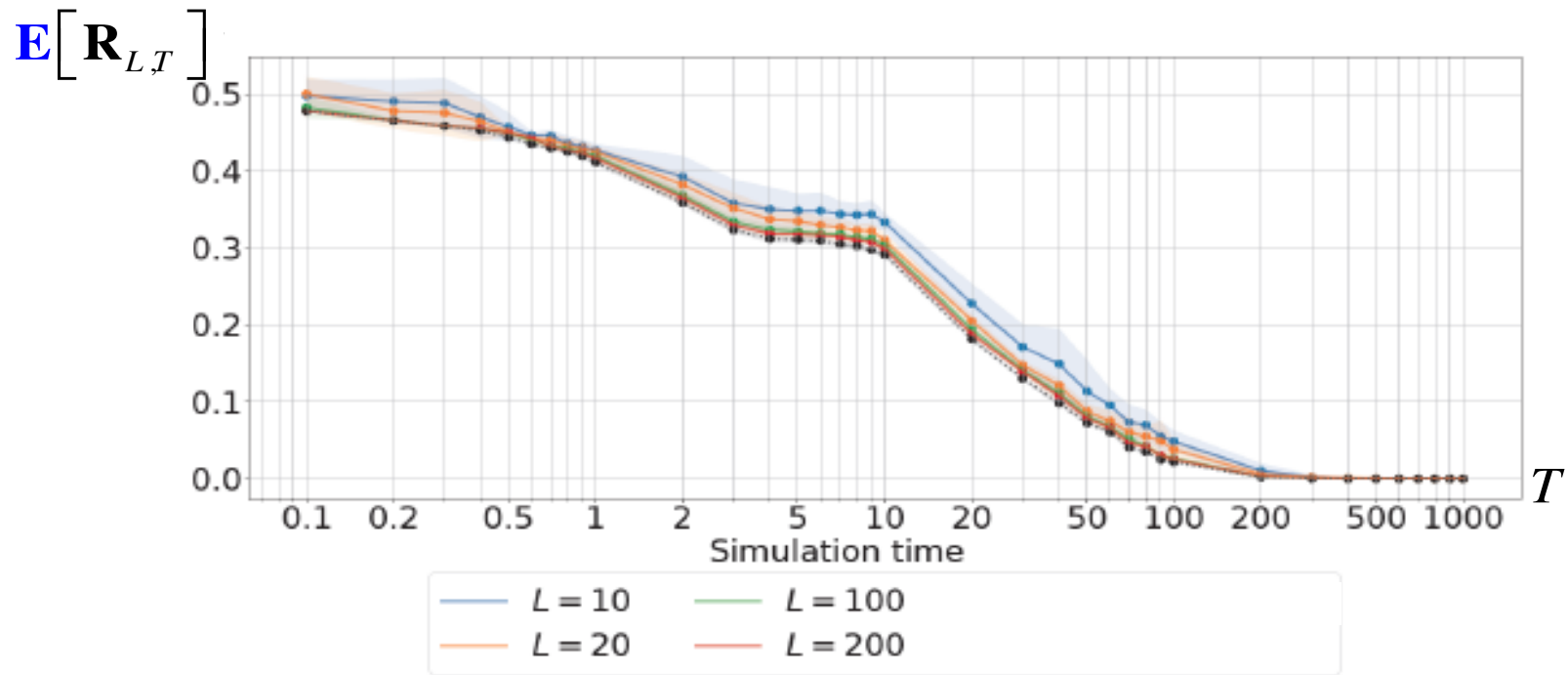
- $\lambda_i(t)$ - Lipschitz on \mathbb{R}_+
- $K(t)$ - Lipschitz on $[-1,1]$
- $h(L) \rightarrow 0$ and $\frac{Lh(L)}{\log(L)} \rightarrow \infty$

Then

$$\sup_{t \in [\varepsilon, T-\varepsilon]} \left| \hat{p}_i(t) - p_i(t) \right| \rightarrow 0 \text{ as } L \rightarrow \infty \quad (P) \quad i = 1, 2$$

Note: Need the boundary corrected kernels

Example: $\lambda_1(t) = \lambda(t; \phi_1)$, $\lambda_2(t) = \lambda(t; \phi_2)$, $\phi_2 / \phi_1 = 4$.



- $h \rightarrow$ data driven choice by max log-likelihood

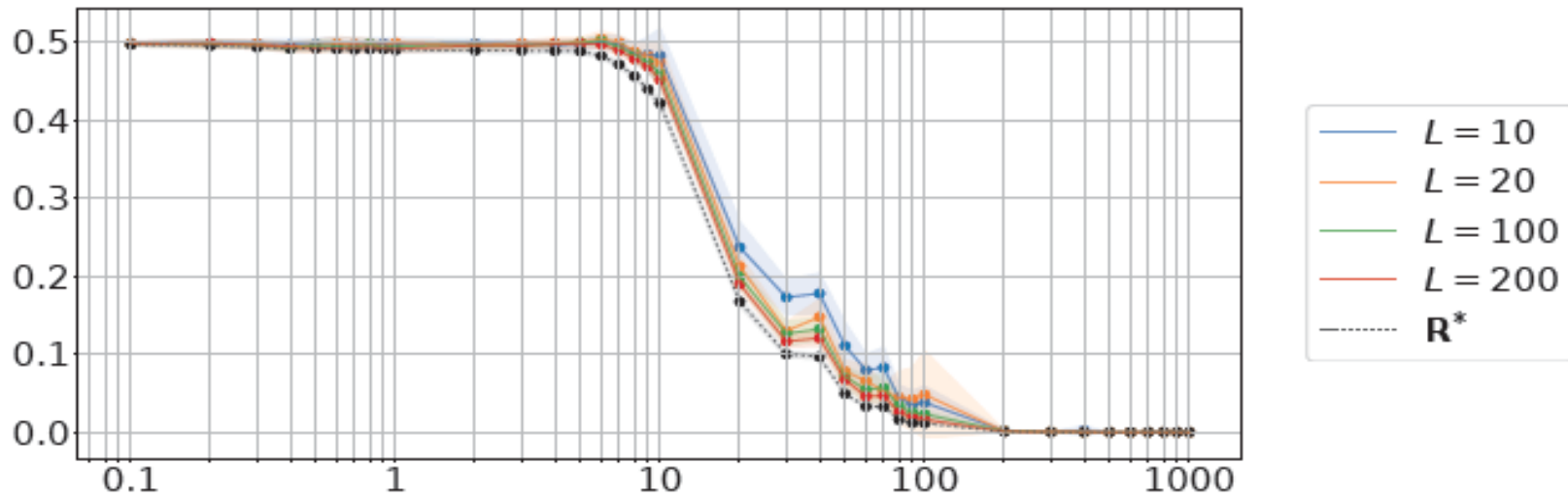
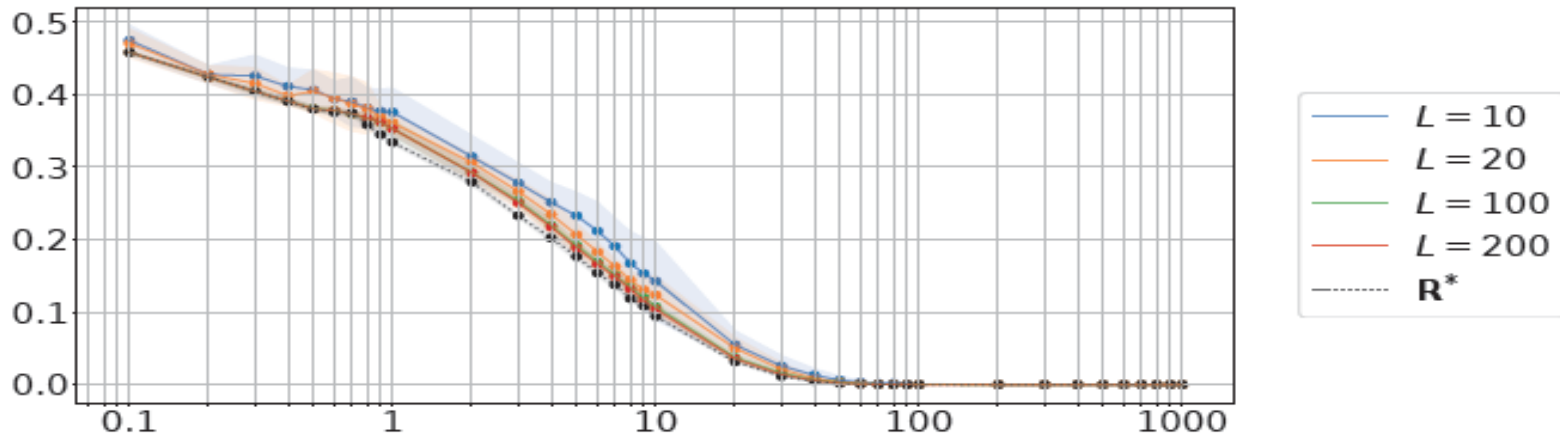


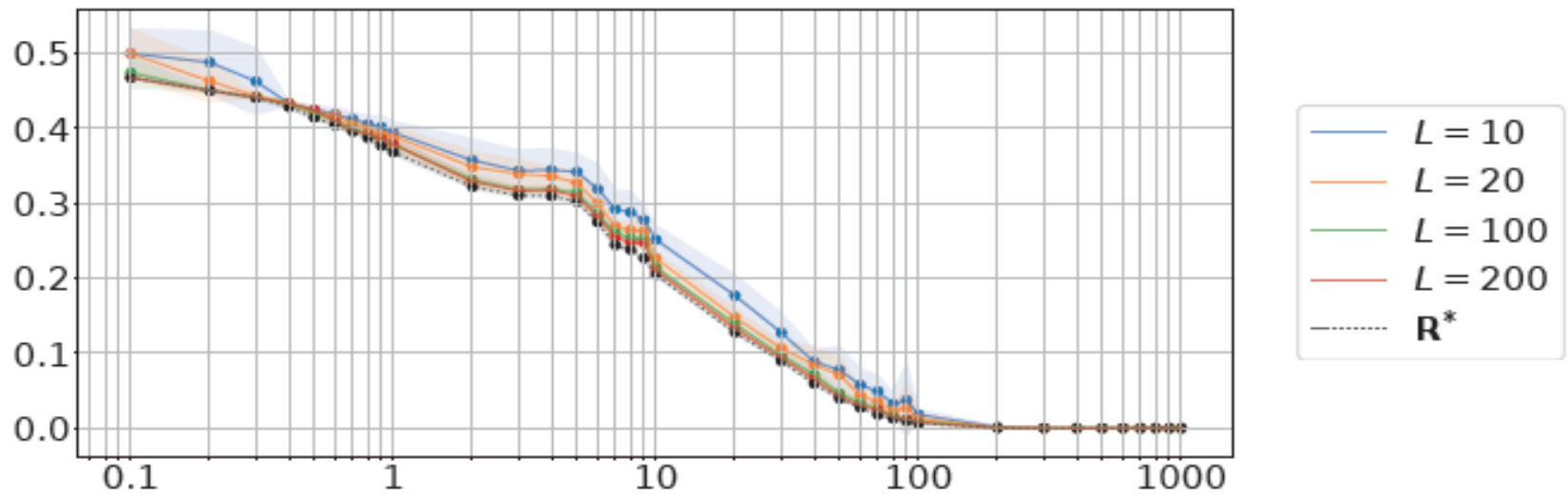
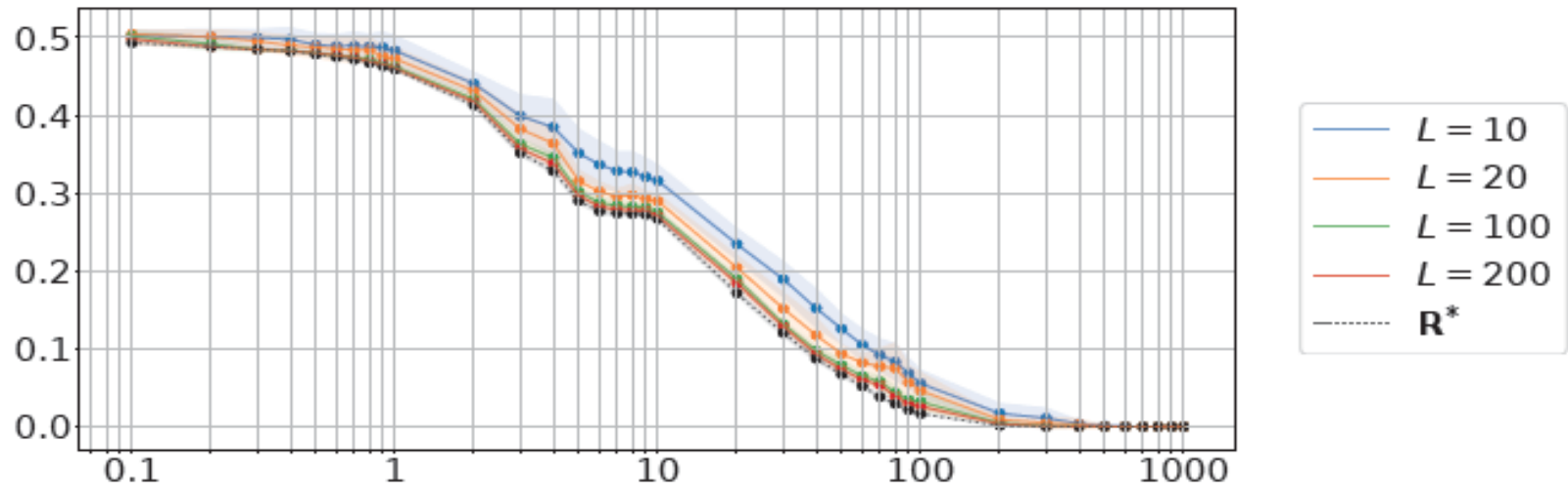
$$\lambda_1(t) = \lambda(t; \phi_1), \lambda_2(t) = \lambda(t; \phi_2)$$

$$\lambda_1(t) = \lambda(t; r_1), \lambda_2(t) = \lambda(t; r_2)$$

a)	$\lambda(t; r) = r$
b)	$\lambda(t; r) = (r + 1) + r \cos\left(\frac{2\pi}{30}t + 2.6\right) + \cos\left(\frac{2\pi}{28}t + 4.5\right)$
c)	$\lambda(t; \phi) = \left[3.1 + 3 \cos\left(\frac{\pi}{3\sqrt{2}}t + \phi\right)\right]^{1/2}$
d)	$\lambda(t; \phi) = 1.3 \exp\left[\cos\left(\frac{\pi}{3\sqrt{2}}t + \phi\right)\right]$

$E[\mathbf{R}_{L,T}]$ versus T





5. Extensions

▲ Rates !

$$\mathbf{R}_{L,T} = \mathbf{R}_T^* + O_P(L^{-\alpha})$$

▲ Inference from the Single Realization

- Aalen Multiplicative Models

O. Aalen. Nonparametric inference for a family of counting processes. The Annals of Statistics, 1978.

$$\lambda_i(t) = \mu \gamma_i(t), \mu \nearrow$$

▲ **Multi-Class Problem:** $(\omega_1, \lambda_1(t)), \dots, (\omega_c, \lambda_c(t))$

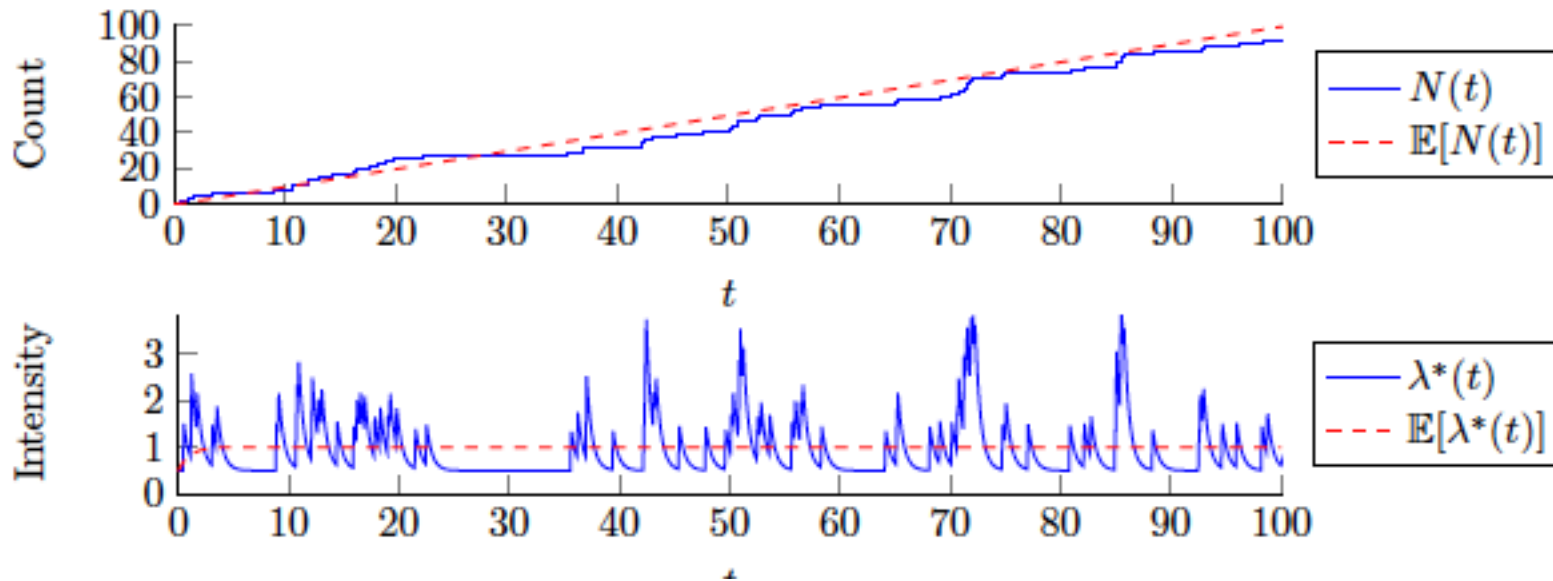
$$\psi_T^* : \mathbf{X} \in \omega_i \text{ if } \sum_{s=1}^N \log \left(\frac{\lambda_i(t_s)}{\lambda_k(t_s)} \right) \geq \gamma_{ik} \text{ for all } k \neq i$$

↑

$$\gamma_{ik} = \int_0^T (\lambda_i(u) - \lambda_k(u)) du + \log \left(\frac{\pi_k}{\pi_i} \right)$$

▲ Random Intensities

$$\lambda^*(t) = \lambda_0 + \sum_{t_i < t} g(t - t_i) = \lambda_0 + \int_0^t g(t - s) d\mathbf{N}(s)$$



R. Lima. **Hawkes** processes modeling, inference, and control: an overview. **SIAM Review**, 2023.

References

- **A. Cholaquidis, L. Forzani, P. Llop, L. Moreno. On the classification problem for Poisson point processes. Journal of Multivariate Analysis, 2017.**
- **X. Rong and V. Solo. On the error rate of classifying point processes. 60th IEEE Conference on Decision and Control, 2021.**
- **M.Pawlak. Bayes risk consistency of nonparametric rules for spike trains data, IEEE IT, 2024.**

Dziękuję !