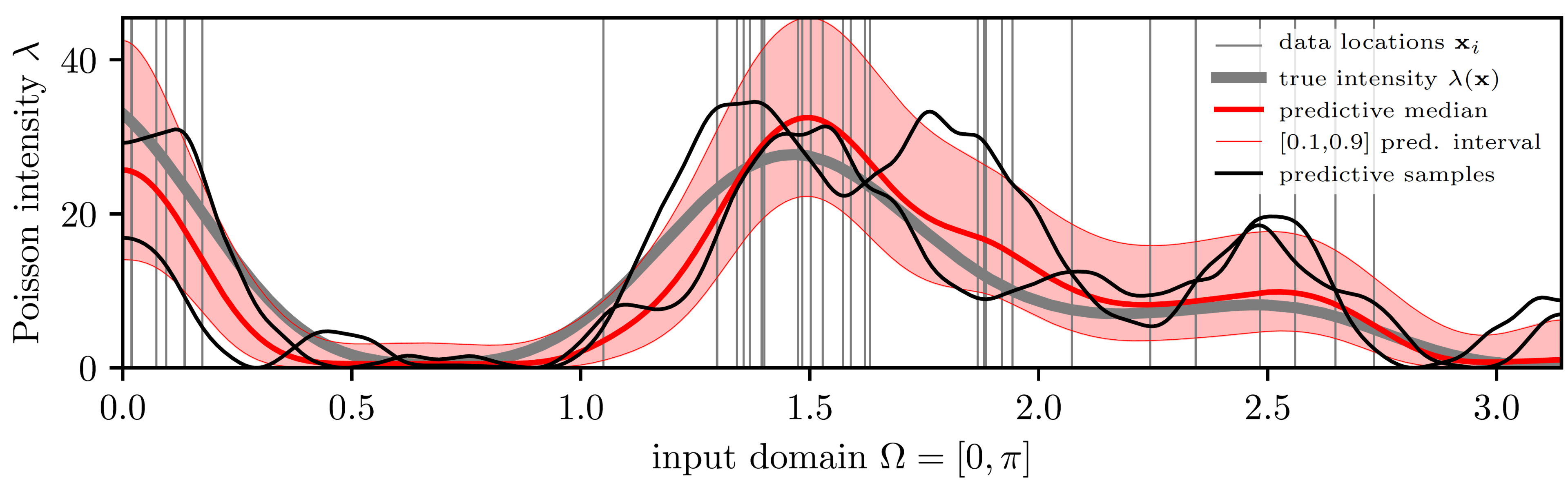
Fast Bayesian Intensity Estimation for the Permanental Process

Christian J. Walder[1,2] and Adrian N. Bishop[1,2,3]

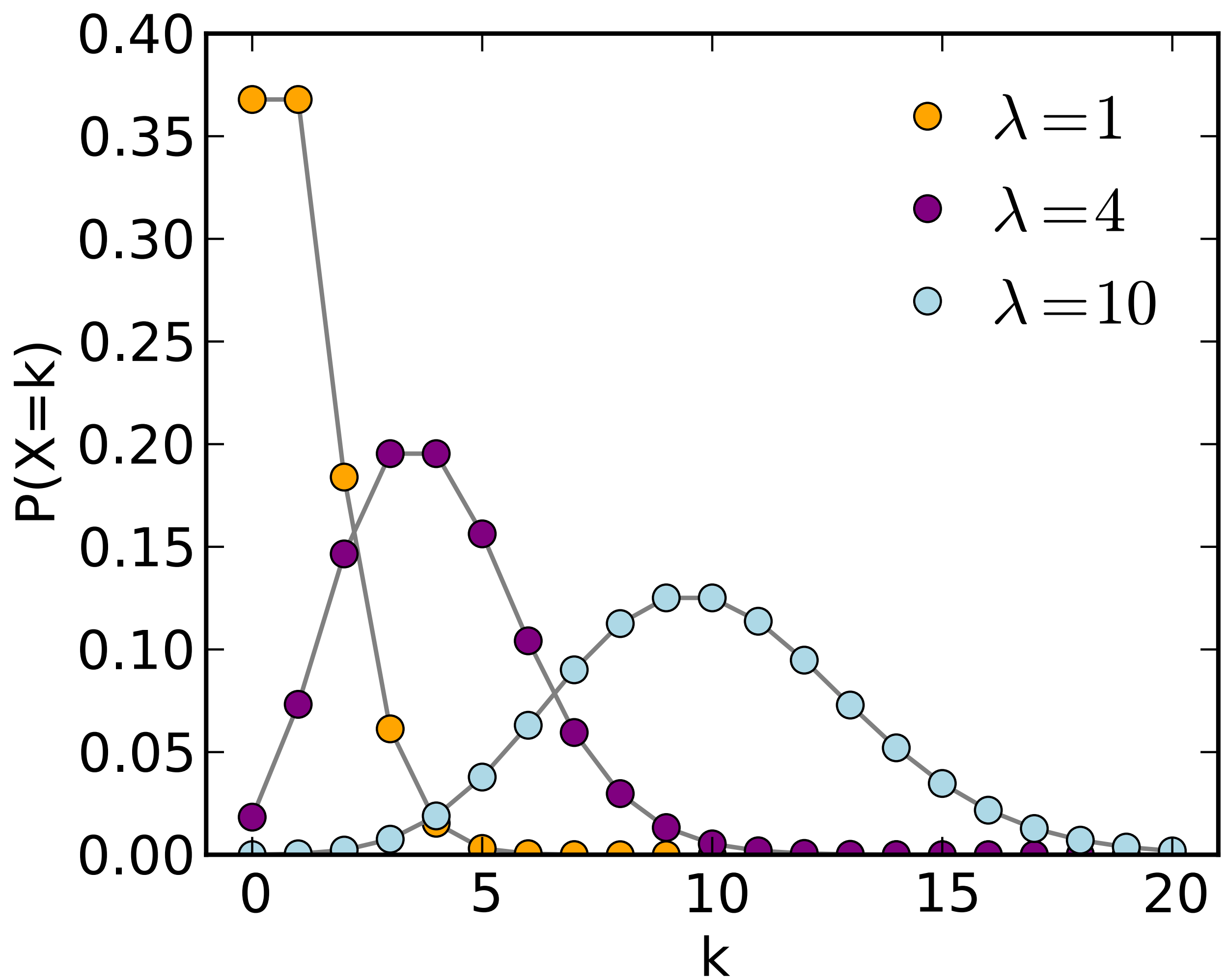
[1] Data61, CSIRO, Australia
[2] The Australian National University
[3] University of Technology Sydney



Overview

- Poisson distribution
- Poisson point process
  - Definition
  - Likelihood
- Squared link function:
  - Reproducing kernel Hilbert space norm regularisation
  - Gaussian process prior
- Experiments
- Summary

Poisson Random Variable



$$X|\lambda \sim \mathrm{Poisson}(\lambda)$$

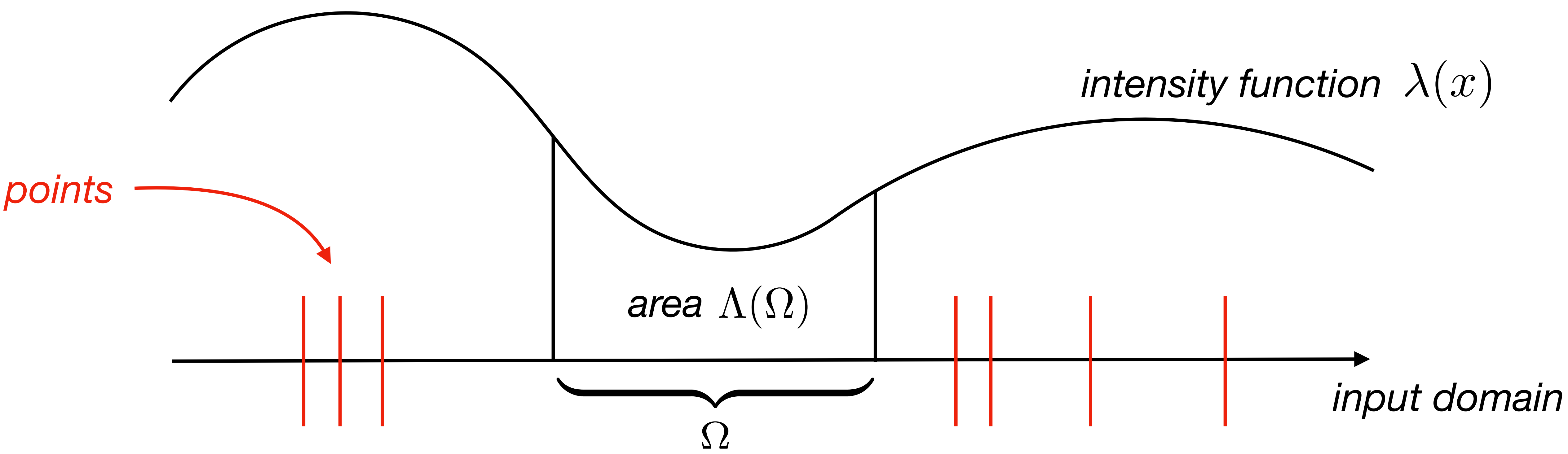
$$P(X = k|\lambda) = \lambda^k \exp(-\lambda)/k!$$

Poisson Point Process

- Distribution over sets of points
- The number of points in a subset Ω is

$$N(\Omega) \sim \text{Poisson}(\Lambda(\Omega))$$

$$\Lambda(\Omega) = \int_{\boldsymbol{x} \in \Omega} \lambda(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$



Poisson Process: Likelihood Function

What is the density $p(\mathcal{X}|\lambda, \Omega)$ for realisation $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1,2,\ldots,m} \subset \Omega$?

$$p(\mathcal{X}|\lambda, \Omega) = P(|\mathcal{X}| = m|\lambda, \Omega)\, m! \prod_{i=1}^{m} p(\boldsymbol{x}_i|\lambda, \Omega)$$

where $\quad P(|\mathcal{X}| = m|\lambda, \Omega) \triangleq \text{Poisson}(m|\Lambda(\Omega)),$

and

$$\frac{p(\boldsymbol{x}_i|\lambda, \Omega)}{p(\boldsymbol{x}_0|\lambda, \Omega)} = \lim_{\epsilon \to 0} \frac{1 - \text{Poisson}(0|\Lambda([\boldsymbol{x}_i, \boldsymbol{x}_i + \epsilon]))}{1 - \underbrace{\text{Poisson}(0|\Lambda([\boldsymbol{x}_0, \boldsymbol{x}_0 + \epsilon]))}_{\text{Pr[zero points near } x_0]}} = \frac{\lambda(\boldsymbol{x}_i)}{\lambda(\boldsymbol{x}_0)}$$

$$\Rightarrow \quad p(\boldsymbol{x}_i|\lambda, \Omega) = \frac{\lambda(\boldsymbol{x}_i)}{\Lambda(\Omega)}.$$

So the likelihood simplifies to:

$$p(\mathcal{X}|\lambda, \Omega) = \frac{\Lambda(\Omega)^m \exp(-\Lambda(\Omega))}{m!} m! \prod_{i=1}^{m} \frac{\lambda(\boldsymbol{x}_i)}{\Lambda(\Omega)} = \exp(-\Lambda(\Omega)) \prod_{i=1}^{m} \lambda(\boldsymbol{x}_i).$$

Squared Link Function: Regularised Maximum Likelihood [1]

Writing out the integral in the likelihood we get

$$\ln p(\mathcal{X}|\lambda, \Omega) = \sum_{i=1}^{m} \log \lambda(\boldsymbol{x}_i) - \int_{\boldsymbol{x}\in\Omega} \lambda(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

intractable integral in the likelihood

which, by parameterising $\lambda(\boldsymbol{x}) = \frac{1}{2} f^2(\boldsymbol{x})$, becomes

permanental process for f ~ GP

$$\propto 2 \sum_{i=1}^{m} \log f(\boldsymbol{x}_i) - \frac{1}{2} \underbrace{\int_{\boldsymbol{x}\in\Omega} f^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}}_{\triangleq \|f\|^2_{L_2(\Omega)}} .$$

Regularised maximum likelihood with regulariser (log prior) $\|f\|^2_{\mathcal{H}}$ gives

$$f^* \triangleq \arg\max_f = 2 \sum_{i=1}^{m} \log f(\boldsymbol{x}_i) - \frac{1}{2} \underbrace{\left(\|f\|^2_{L_2(\Omega)} + \|f\|^2_{\mathcal{H}} \right)}_{\|f\|^2_{\triangleq \mathcal{H}(k,\Omega)}} .$$

modified RKHS

Can easily solve with the theory of reproducing kernel Hilbert spaces.

[1] Flaxman, S, Teh, YW, and Sejdinovic, D, **Poisson Intensity Estimation with Reproducing Kernels.** AISTATS 2017.

Summary so Far

To summarise, we handle the intractable integral by

1. letting $\lambda(\boldsymbol{x}) = \frac{1}{2} f^2(\boldsymbol{x})$ so that the integral becomes a function norm
2. effectively "removing" the integral from the likelihood
3. including it in the in the regulariser via

$$\|f\|^2_{\mathcal{H}(k,\Omega)} \triangleq \|f\|^2_{L_2(\Omega)} + \|f\|^2_{\mathcal{H}} \,.$$

Solution is then trivial given the reproducing kernel of $\mathcal{H}(k, \Omega)$.

Regularisation Operator Approach

The norm is

$$\|f\|^2_{\mathcal{H}(k,\Omega)} \triangleq \|f\|^2_{L_2(\Omega)} + \|f\|^2_{\mathcal{H}} \,.$$

define the regularisation operator

$$\|f\|^2_{\mathcal{H}} \triangleq \|\psi f\|^2_{L_2(\Omega)} \,,$$

use the reproducing property

$$f(\boldsymbol{x}) \triangleq \langle f, \tilde{k}(\boldsymbol{x}) \rangle_{\mathcal{H}(k,\Omega)},$$

we get the (typically partial differential) equation

$$\tilde{k}(\boldsymbol{x}, \cdot) + \psi^* \psi \tilde{k}(\boldsymbol{x}, \cdot) = \delta(\cdot)$$

Depending on ψ this is *e.g.* a Poisson or Klein-Gordon equation, *etc.*

- Leads to useful closed form expressions and algorithms from physics.
- Unfortunately it's unclear how to make it probabilistic (Gaussian process)!

Duffy, D. ***Green's Functions with Applications.*** 2015 book.
Thomas-Agnan, C. ***Computing a Family of Green's Functions for Statistical Applications.*** 1993 tech report.
Sollich, P and Williams, C. K. I. ***Understanding Gaussian Process Regression Using the Equivalent Kernel***. NIPS 2005.

Squared Link Function: Gaussian Process Prior

By Mercer's theorem, we may decompose the covariance k as

$$k(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{N} \lambda_i \phi_i(\boldsymbol{x}) \phi_i(\boldsymbol{y})$$

Gaussian process distributed f may therefore be written

$$\begin{aligned} f(x) &= \boldsymbol{w}^\top \Phi(\boldsymbol{x}) \\ \text{where} \quad \boldsymbol{w} &\sim \mathcal{N}(0, \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)). \end{aligned}$$

We can then derive the (Laplace) approximate predictive mean and variance

$$\begin{aligned} \mathbb{E}\left[f(\boldsymbol{x}^*) | X, \Omega, k\right] &\approx \sum_{i=1}^{m} \alpha_i \tilde{k}(\boldsymbol{x}_i, \boldsymbol{x}^*) \\ \operatorname{Var}\left[f(\boldsymbol{x}^*) | X, \Omega, k\right] &\approx \tilde{k}(\boldsymbol{x}^*, \boldsymbol{x}^*) - \left(\tilde{k}(\boldsymbol{x}^*, X) \odot \boldsymbol{\alpha}\right) S^{-1} \left(\boldsymbol{\alpha}^\top \odot \tilde{k}(X, \boldsymbol{x}^*)\right), \end{aligned}$$

no representer theorem required

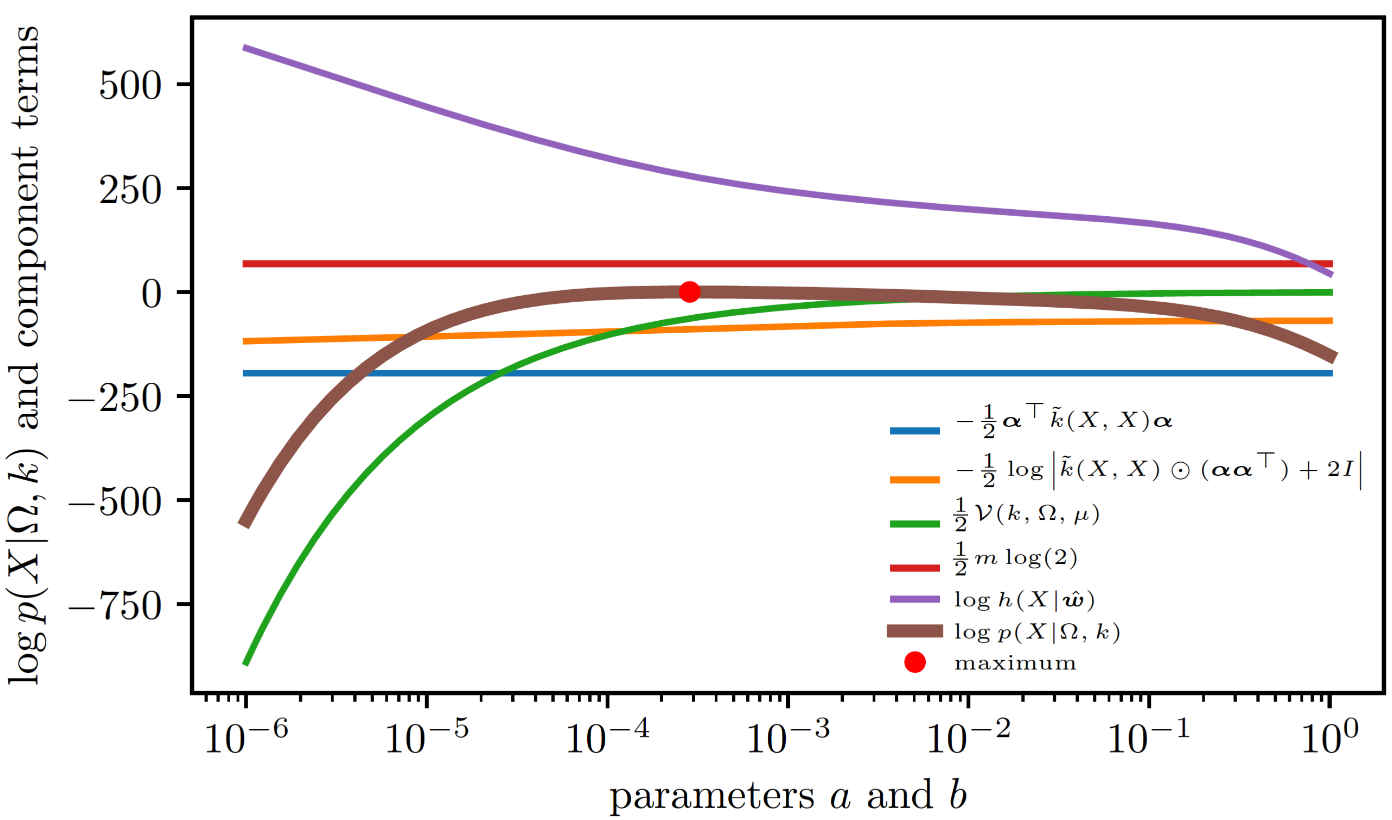
the same "equivalent kernel"

where

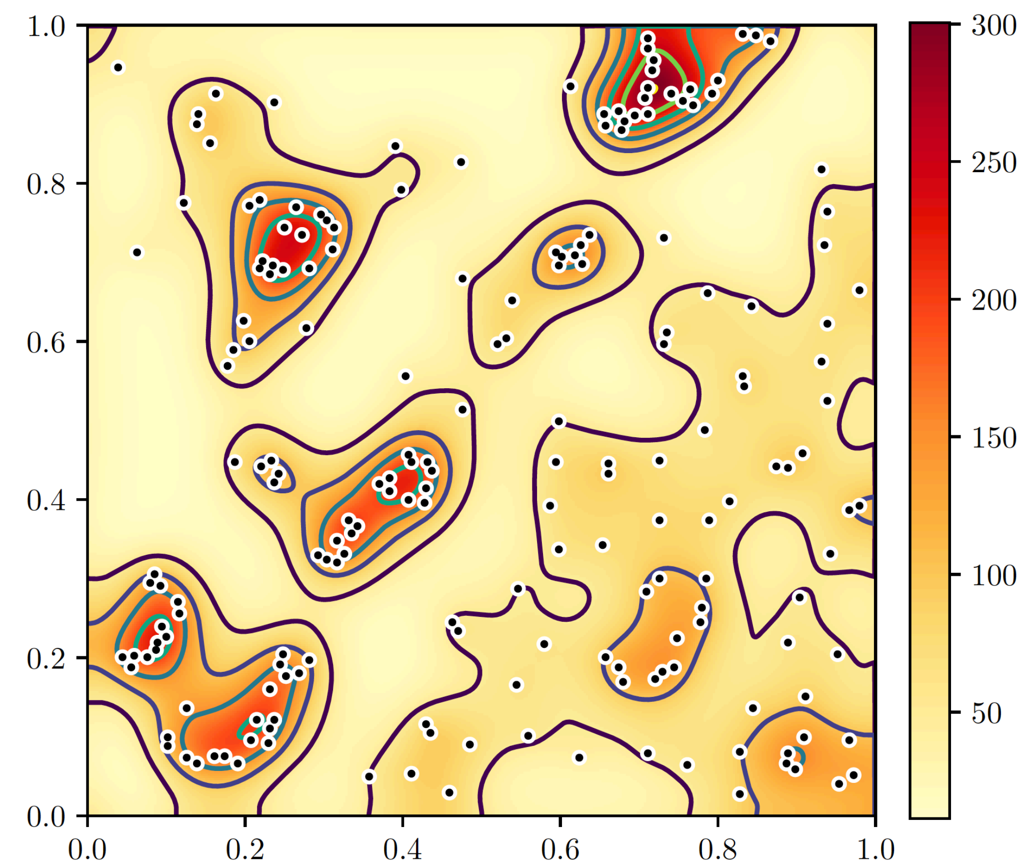
$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \operatorname*{argmin}_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \log \alpha_i^2 + \frac{1}{2} \boldsymbol{\alpha}^\top \tilde{K} \boldsymbol{\alpha}, \\ S &= \left(\tilde{k}(X, X) \odot (\boldsymbol{\alpha}\boldsymbol{\alpha}^\top) + 2I\right). \end{aligned}$$

Model Selection: Marginal Likelihood

The marginal likelihood is more cumbersome to write out, so we visualise a decomposition of it here:

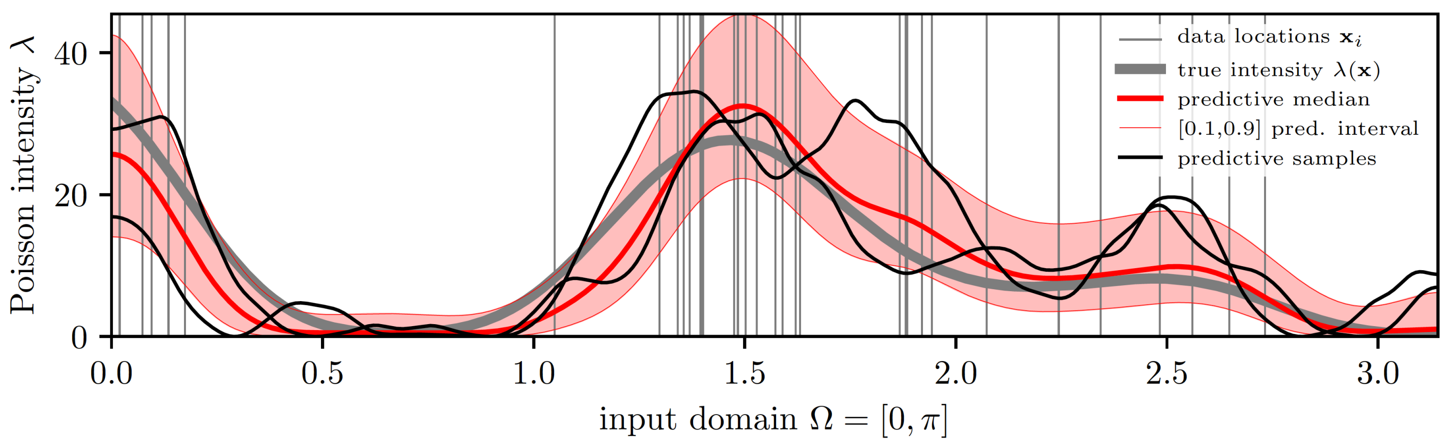


(a) Decomposition of the marginal likelihood

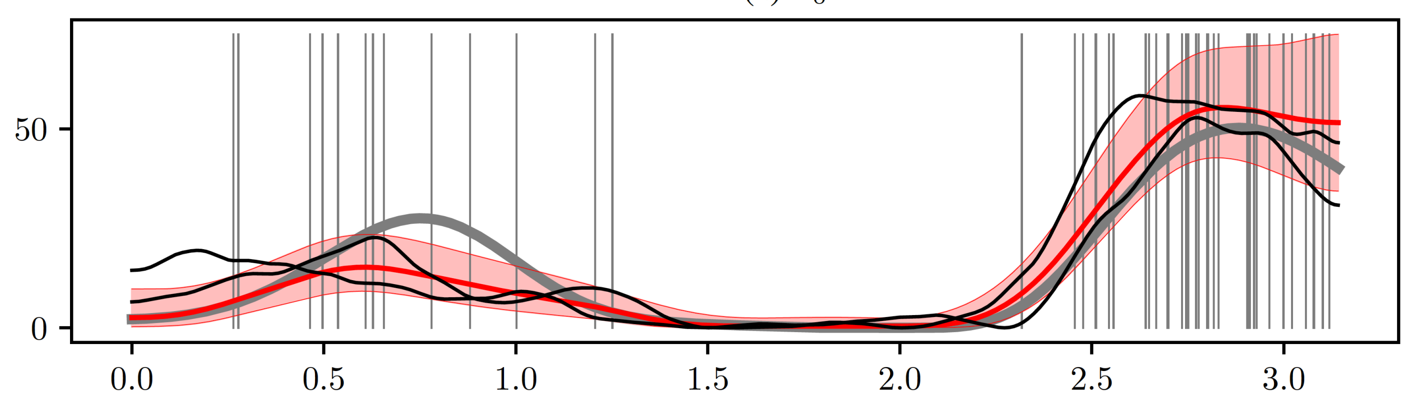


(b) Predictive mean intensity.

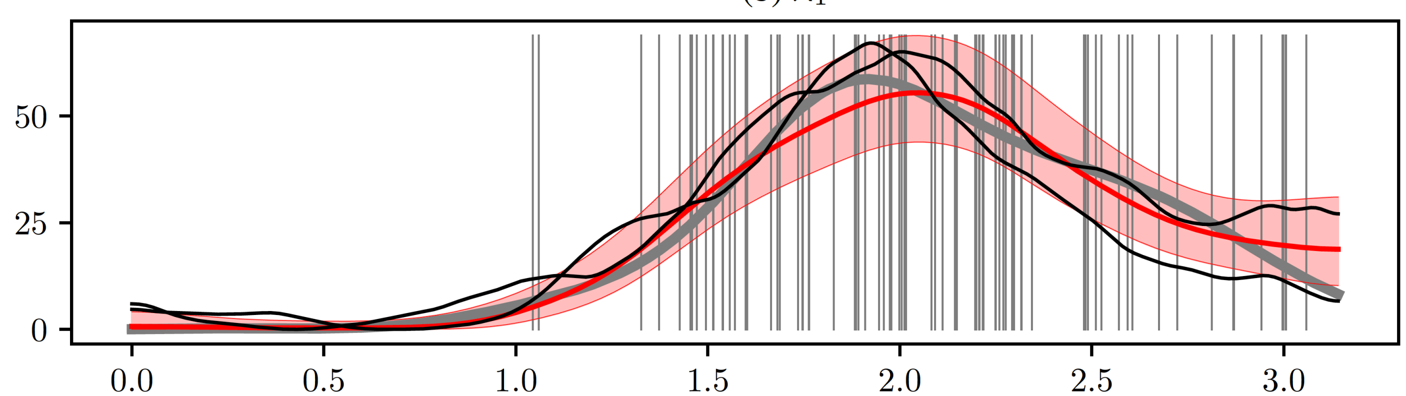
Model Selection: Marginal Likelihood



(a) λ_0



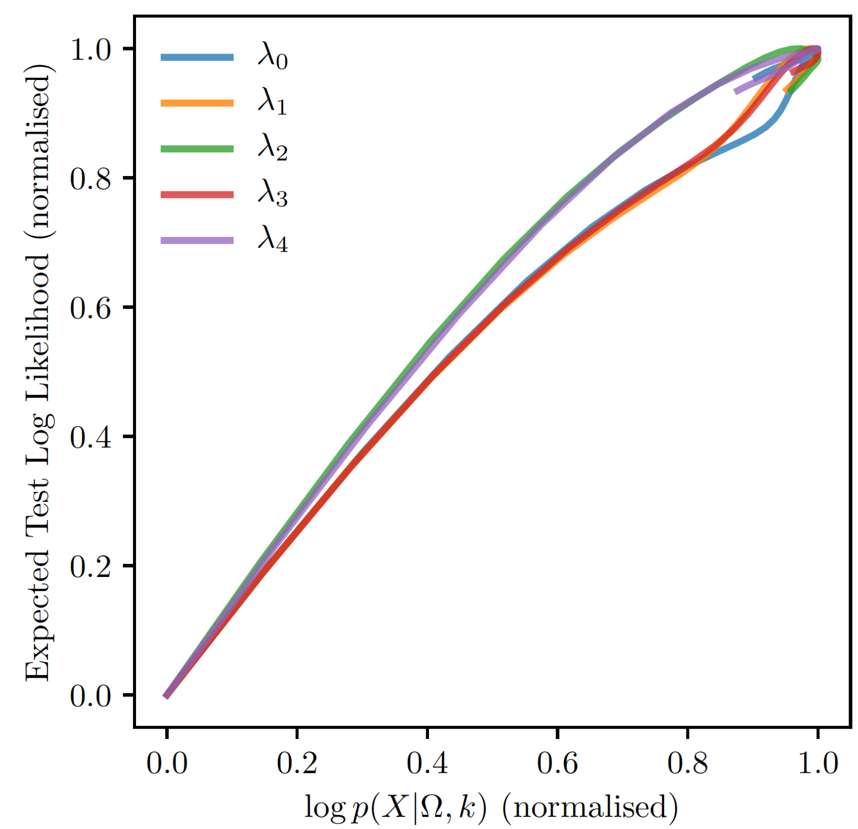
(b) λ_1



(c) λ_2

we observe a strong relationship between the marginal likelihood and the empirical predictive power

=> ML-II model selection works



(b) Expected log-loss *vs.* marginal likelihood.

Summary

- Previous work on log-Gaussian Cox processes has been hampered by computational problems
- We considered the poisson point process with intensity which is the square of a Gaussian process
- We demonstrated the advantages of a squared link function for the Cox process with Gaussian process prior
- The result is a simple and fast Bayesian method
- This is one of several recent papers which redress the balance w.r.t. the extensively studied log-Gaussian Cox process