Fast Bayesian Intensity Estimation for the Permanental Process

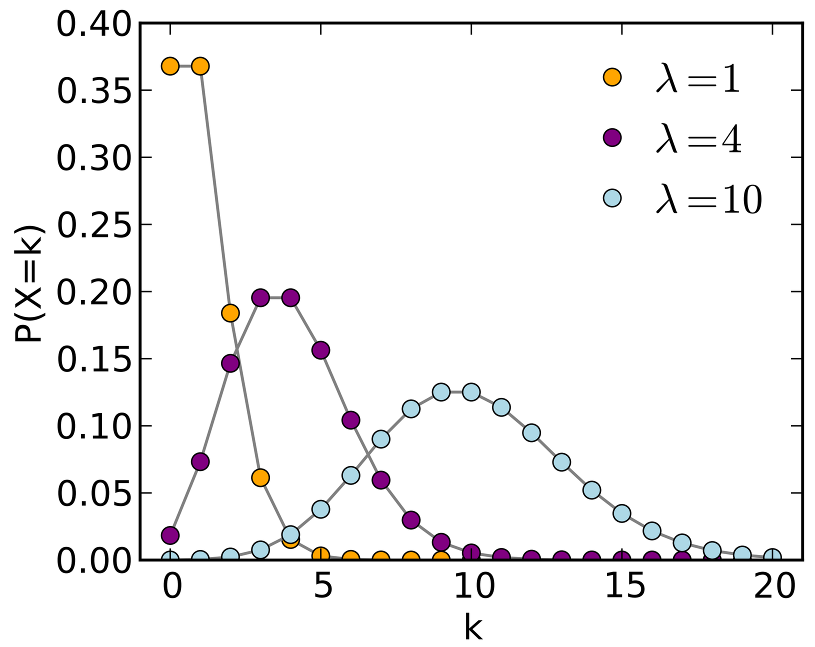
Christian J. Walder [1,2] and Adrian N. Bishop [1,2,3]

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

Abstract

The Cox process is a stochastic process which generalises the Poisson process by letting the underlying intensity function itself be a stochastic process. In this paper we present a fast Bayesian inference scheme for the permanental process, a Cox process under which the square root of the intensity is a Gaussian process. In particular we exploit connections with reproducing kernel Hilbert spaces, to derive efficient approximate Bayesian inference algorithms based on the Laplace approximation to the predictive distribution and marginal likelihood. We obtain a simple algorithm which we apply to toy and real-world problems, obtaining orders of magnitude speed improvements over previous work.

Background: Poisson Random Variable

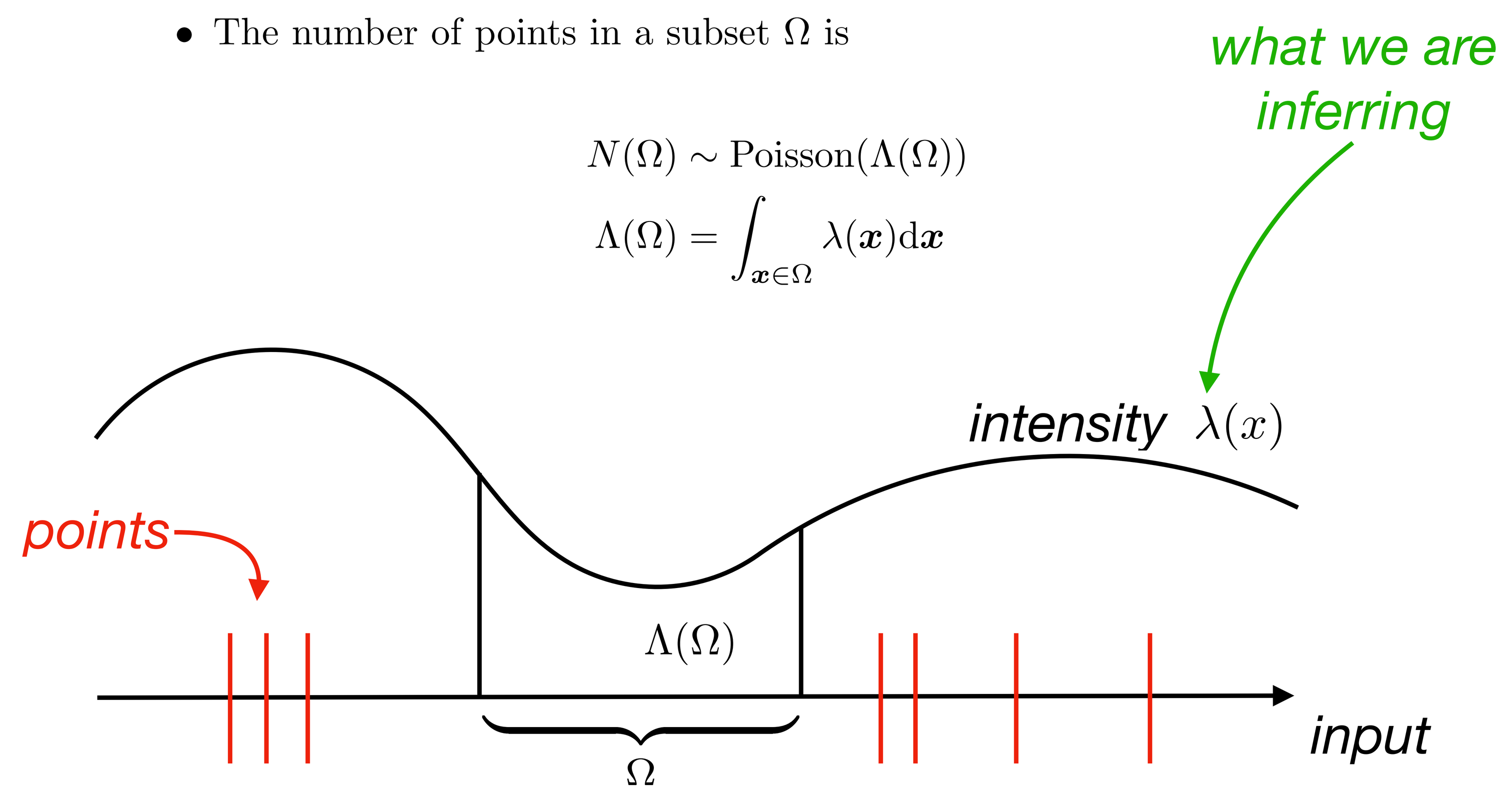


$$X|\lambda \sim \text{Poisson}(\lambda)$$
$$P(X=k|\lambda) = \lambda^k \exp(-\lambda)/k!$$

Background: Poisson Process

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

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



Challenge: Poisson Process Likelihood Function

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

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

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

and

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

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

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

So the likelihood simplifies to:

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

Key Idea: Squared Link Function

The trick:

1. let $\lambda(\boldsymbol{x}) = \frac{1}{2}f(\boldsymbol{x})$ so the integral in the likelihood becomes a squared norm
2. effectively "remove" the integral from the likelihood
3. include it in the regulariser / log-prior

Key Idea: Applied with Kernel Regularisation[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

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

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

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.

Contribution: Approximate Bayesian Scheme

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

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

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

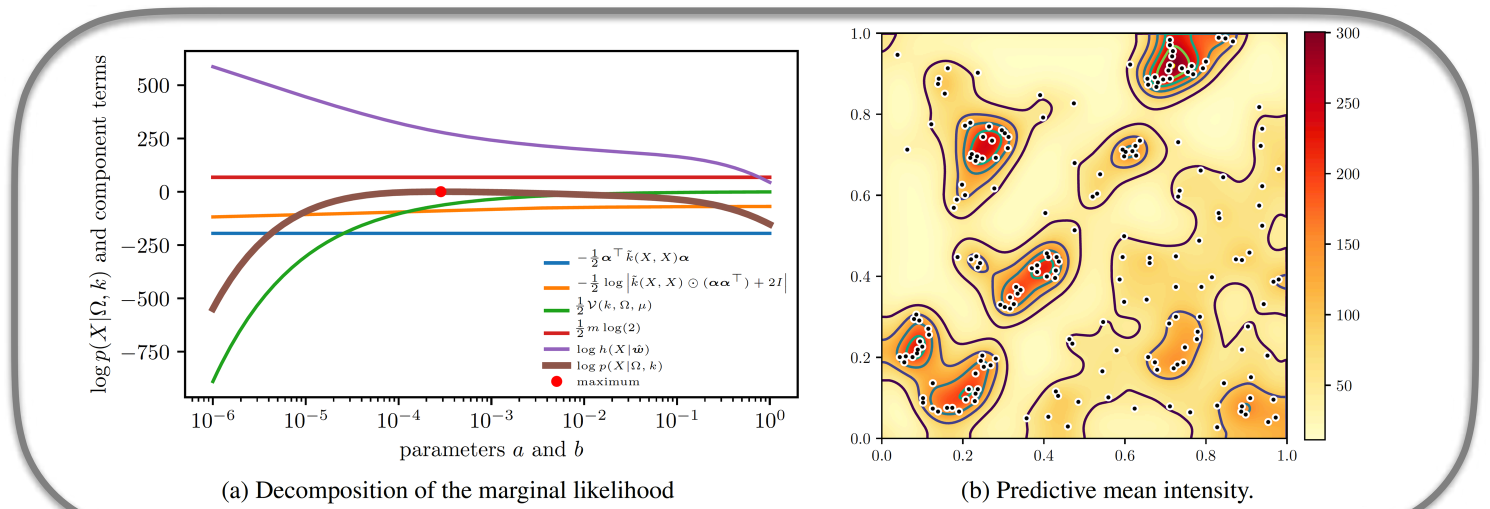
$$\mathbb{E}\left[f(\boldsymbol{x}^*)|X,\Omega,k\right] \approx \sum_{i=1}^{m}\alpha_i\tilde{k}(\boldsymbol{x}_i,\boldsymbol{x}^*)$$
$$\operatorname{Var}\left[f(\boldsymbol{x}^*)|X,\Omega,k\right] \approx \tilde{k}(\boldsymbol{x}^*,\boldsymbol{x}^*) - (\tilde{k}(\boldsymbol{x}^*,X)\odot\boldsymbol{\alpha})S^{-1}(\boldsymbol{\alpha}^\top\odot\tilde{k}(X,\boldsymb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

where

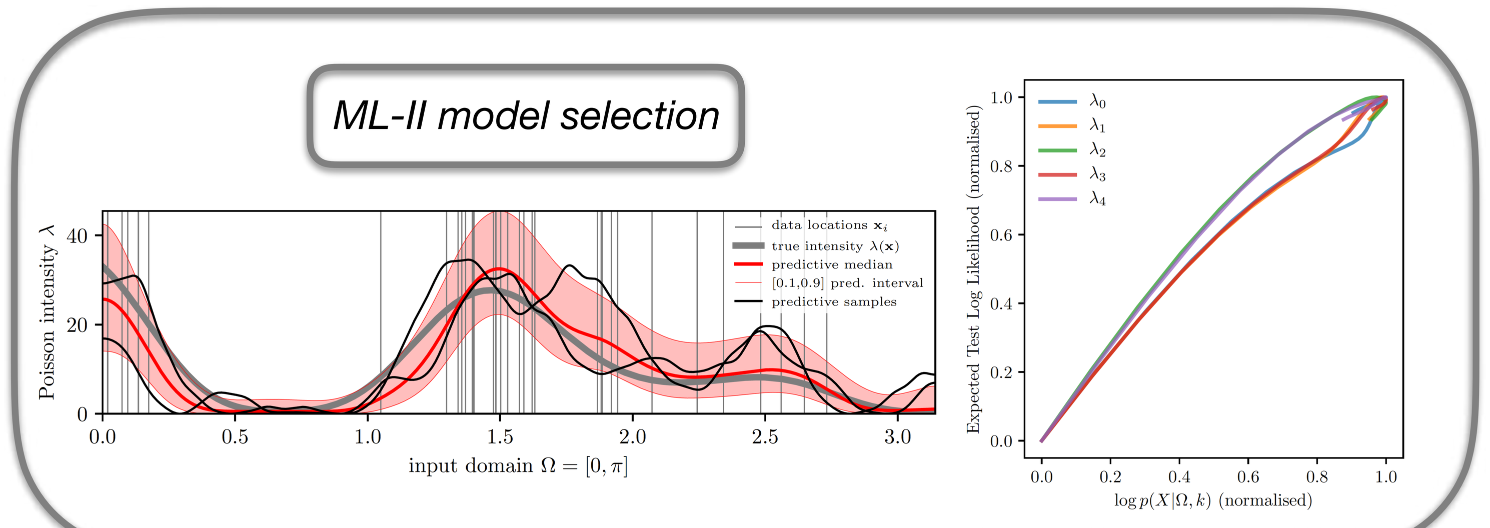
$$\hat{\boldsymbol{\alpha}} = \operatorname*{argmin}_{\boldsymbol{\alpha}}\sum_{i=1}^{m}\log\alpha_i^2 + \frac{1}{2}\boldsymbol{\alpha}^\top\tilde{K}\boldsymbol{\alpha},$$
$$S = (\tilde{k}(X,X)\odot(\boldsymbol{\alpha}\boldsymbol{\alpha}^\top) + 2I).$$

modified "covariance"

Numerical Examples



(a) Decomposition of the marginal likelihood (b) Predictive mean intensity.



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

References

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

[2] Lloyd, CM, Gunter, T, Osborne, MA, and Roberts, SJ, **Variational inference for gaussian process modulated poisson processes.** ICML 2015.