Fast Bayesian Intensity Estimation for the Permanental Process

Christian J. Walder\textsuperscript{1,2} and Adrian N. Bishop\textsuperscript{1,2,3}

\textsuperscript{1} Data61, CSIRO, Australia
\textsuperscript{2} The Australian National University
\textsuperscript{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

\[ X \mid \lambda \sim \text{Poisson}(\lambda) \]

\[ P(X = k \mid \lambda) = \frac{\lambda^k \exp(-\lambda)}{k!} \]
Poisson Point Process

- Distribution over sets of points
- The number of points in a subset $\Omega$ is

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

$$\Lambda(\Omega) = \int_{x \in \Omega} \lambda(x) \, dx$$
Poisson Process: Likelihood Function

What is the density \( p(\mathcal{X}|\lambda, \Omega) \) for realisation \( \mathcal{X} = \{x_i\}_{i=1,2,...,m} \subset \Omega \)?

\[
p(\mathcal{X}|\lambda, \Omega) = P(|\mathcal{X}| = m|\lambda, \Omega) m! \prod_{i=1}^{m} p(x_i|\lambda, \Omega)
\]

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

and

\[
\frac{p(x_i|\lambda, \Omega)}{p(x_0|\lambda, \Omega)} = \lim_{\epsilon \to 0} \frac{1 - \text{Poisson}(0|\Lambda([x_i, x_i + \epsilon]))}{1 - \text{Poisson}(0|\Lambda([x_0, x_0 + \epsilon]))} = \frac{\lambda(x_i)}{\lambda(x_0)}
\]

\[
\Rightarrow \quad p(x_i|\lambda, \Omega) = \frac{\lambda(x_i)}{\Lambda(\Omega)}.
\]

So the likelihood simplifies to:

\[
p(\mathcal{X}|\lambda, \Omega) = \frac{\Lambda(\Omega)^m \exp(-\Lambda(\Omega))}{m!} m! \prod_{i=1}^{m} \frac{\lambda(x_i)}{\Lambda(\Omega)} = \exp(-\Lambda(\Omega)) \prod_{i=1}^{m} \lambda(x_i).
\]
Squared Link Function: Regularised Maximum Likelihood

Writing out the integral in the likelihood we get

\[ \ln p(\mathcal{X}|\lambda, \Omega) = \sum_{i=1}^{m} \log \lambda(x_i) - \int_{x \in \Omega} \lambda(x) dx \]

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

\[ \propto 2 \sum_{i=1}^{m} \log f(x_i) - \frac{1}{2} \int_{x \in \Omega} f^2(x) dx. \]

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

\[ f^\star \triangleq \arg \max_f = 2 \sum_{i=1}^{m} \log f(x_i) - \frac{1}{2} \left( \|f\|_{L_2(\Omega)}^2 + \|f\|_\mathcal{H}^2 \right) . \]

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.
To summarise, we handle the intractable integral by

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

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

Solution is then trivial given the reproducing kernel of $H(k, \Omega)$.
The norm is
\[ \|f\|_{H(k,\Omega)}^2 \triangleq \|f\|_{L^2(\Omega)}^2 + \|f\|_H^2. \]
define the regularisation operator
\[ \|f\|_H^2 \triangleq \|\psi f\|_{L^2(\Omega)}^2, \]
use the reproducing property
\[ f(x) \triangleq \langle f, \tilde{k}(x) \rangle_{H(k,\Omega)}, \]
we get the (typically partial differential) equation
\[ \tilde{k}(x, \cdot) + \psi^* \psi \tilde{k}(x, \cdot) = \delta(\cdot) \]
Depending on \( \psi \) 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.
Solrich, 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

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

Gaussian process distributed $f$ may therefore be written

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

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

$$\mathbb{E}[f(x^*)|X, \Omega, k] \approx \sum_{i=1}^{m} \alpha_i \tilde{k}(x_i, x^*)$$

$$\text{Var}[f(x^*)|X, \Omega, k] \approx \tilde{k}(x^*, x^*) - (\tilde{k}(x^*, X) \odot \alpha) S^{-1} (\alpha \odot \tilde{k}(X, x^*)),$$

where

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{m} \log \alpha_i^2 + \frac{1}{2} \alpha^\top \tilde{K} \alpha,$$

$$S = (\tilde{k}(X, X) \odot (\alpha \alpha^\top) + 2I).$$

no representer theorem required

the same "equivalent kernel"
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

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

=> ML-II model selection works
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