Permanental Process with RFF

Sellier, Jeremy
jeremy.sellier.18@ucl.ac.uk

May 2, 2021

1 Preliminaries

1.1 The Poisson Process

A Poisson process is a random sequence of events occurring on a continuous domain $\mathcal{X}$. The general approach to infer point process is to estimate its intensity function $\lambda(x) : \mathcal{X} \mapsto \mathbb{R}^+$. The intensity $\lambda(x)$ can be interpreted heuristically as the instantaneous probability of occurrence of an event around a location x, i.e.

$$\lambda(x) := \lim_{\mu(dx) \to +0} \frac{E\left[N(dx)\right]}{\mu(dx)}$$

where dx is a small region around x with measure $\mu(dx)$ and $N(A)$ is the random number of points within a sub-region $A \subset \mathcal{X}$. Then given some observations $\mathbf{x} = \{x_i\}_{i=1}^N$ we can define the likelihood of data as

$$p(\mathbf{x}|\lambda(\cdot)) = \exp\left(-\int_{\mathcal{X}} \lambda(x)\, d\mu(x)\right) \prod_{i=1}^{N} \lambda(x_i). \tag{1}$$

1.2 Gaussian Cox Process

A Cox process, also known as a doubly stochastic Poisson process is a point process which is a generalization of a Poisson process where the intensity is itself a stochastic process. A Gaussian Process Cox process is obtained by placing a GP prior on λ. we also need to add a deterministic "link" function $\ell : \mathbb{R} \mapsto \mathbb{R}^+$ to that ensure the intensity function remains non-negative, leading to

$$\begin{aligned} &\lambda(\cdot) := \ell(f(\cdot)) \\ &f \sim \mathcal{GP}(0, K(x, x')). \end{aligned} \tag{2}$$

We have the posteriors

$$p(\mathbf{x}|\mathbf{f}) = \exp\left(-\int_{\mathcal{X}} \ell(f(x))\, dx\right) \prod_{i=1}^{N} \ell(f(x_i)) \tag{3}$$

$$p(\mathbf{f}|\mathbf{x}) = \frac{\exp\left(-\int_{\mathcal{X}} \ell(f(x))\, dx\right)\left[\prod_{i=1}^{N} \ell(f(x_i))\right]\mathcal{GP}(f)}{\int \exp\left(-\int_{\mathcal{X}} \ell(f(x))\, dx\right)\left[\prod_{i=1}^{N} \ell(f(x_i))\right]\mathcal{GP}(f)\, df}$$

which are often described as "doubly-intractable" because of the integral of the intensity function and the nested integral in the denominator, which typically cannot be calculated explicitly.

2 Permanental Process

We now consider the general case spatial where $\mathcal{X} = \mathbb{R}^2$ and restrict the domain of observations to a bounded window $\mathcal{S} \subset \mathbb{R}^2$. The absence of a closed form for the integral in (3) makes the fitting of Cox point process models to point pattern data difficult. Fitting even simple Cox processes has typically used MCMC methodology, and has been extremely computationally expensive.

However, a more flexible class of Gaussian Cox process, called *permanental process*, provides exception in that analytical expression for their density are available. They are obtained by defining the intensity in (3) as a square of Gaussian processes i.e setting $l(\cdot) = (\cdot)^2$. We expose here some brief results regarding permanental process taken from McCullagh and Moller [1].

Integral expression via Random Fourrier Feature Alternatively we can decompose the kernel via *Random Fourrier Feature* proposed by Rahimi and Recht [2]. The approach is a consequence of Bochner's theorem, which states that any bounded, continuous and shift-invariant kernel $k(x, x') := k(x - x')$ is a Fourier transform of a bounded positive measure. Assuming k has a spectral density p, it can be rewritten as

$$\begin{aligned} k(x - x') &= \int_{\mathbb{R}^d} \exp(-iw^\top(x - x'))p(w)dw \\ &= E_w\left[\exp(-iw^\top(x - x'))\right] \\ &\approx \frac{1}{R}\sum_{i}^{R} \exp(-iw_i^\top(x - x')) \\ &:= \tilde{k}(x - x') \end{aligned} \tag{4}$$

where $w_1, \ldots, w_R$ are iid samples following $p(w)$. We can now define an explicit feature mapping $\phi : \mathcal{X} \to \mathbb{R}^R$

$$\varphi(x) = \frac{1}{\sqrt{R}}\left[\exp(-iw_1^\top x), \ldots, \exp(-iw_R^\top x)\right]^\top .$$

and express (9) as $\tilde{k}(x - x') = \varphi(x)^\top\varphi(x')$.

As previously, we can now reformulate $f \sim \mathcal{GP}$ to an equivalent linear form approximation

$$f(x) \approx \sum_{i=0}^{R} \beta_i\varphi_i(x) \tag{5}$$

where $\beta \sim \mathcal{N}(0, I)$. Indeed,

$$\mathrm{Cov}(f(x), f(x')) = \sum_{i=0}^{R}\sum_{j=0}^{R} E\left[\beta_i\beta_j\right]\varphi_i(x)\varphi_j(x') = \sum_{i=0}^{\infty}\varphi_i(x)\varphi_i(x') = \tilde{k}(x, x')$$

Gaussian Kernel For Gaussian kernel where $k(x - x') = \exp\left(-\frac{||x-x'||_2^2}{2\sigma^2}\right)$, the corresponding spectral density $p(w)$ defined in (9) is Gaussian $\mathcal{N}(0, \sigma^{-2}I)$. For dimension $d > 1$ the extended Gaussian kernel $k(x - x') = \prod_{i=1}^{d}\exp\left(-\frac{(x_i - x_i')^2}{2\sigma_i^2}\right)$ admits a spectral density of $p(w) \sim \mathcal{N}(0, \mathrm{diag}(\gamma))$ where $\gamma = [1/\sigma_1^2, \ldots, 1/\sigma_d^2]$.

Appendix A Integral Expression via RFF

Let f the be the RFF approximation of the GP as in (5) i.e. $f(x) = \sum_{i=0}^{R} \beta_i \varphi_i(x)$. We further assume without loss of generality that $d = 2$ for spatial-points data and $\mathcal{S} = [-a, a]^2$. We also let $x = (x^1, x^2)$, $w_i = (w_i^1, w_i^2)$ for all $i = 1, \dots, R$.

Calculation 1

The kernels are in practise real valued thus the imaginary part of (9) can be discarded as follows

$$\begin{aligned} k(x - x') &= E_w \left[\exp(-iw^\top (x - x'))\right] \\ &= E_{w,b} \left[2\cos(w^\top x + b)\cos(w^\top x' + b)\right] \\ &\approx \varphi(x)^\top \varphi(x) \end{aligned} \tag{6}$$

where $w \sim p(w)$ and $b \sim \mathcal{U}[0, 2\pi]$ and the explicit feature mapping is defined as

$$\varphi(x) = \frac{\sqrt{2}}{\sqrt{R}} \left[\cos(w_1^\top x + b_1), \dots, \cos(w_R^\top x + b_R)\right]^\top .$$

The integral of f over $\mathcal{S}$ can now be expressed as

$$\begin{aligned} \int_{[-a,a]^2} f(x)^2 dx &= \frac{2}{R} \sum_{i,j} \beta_i \beta_j \int_{[-a,a]^2} \varphi_i(x)\varphi_j(x')\, dx \\ &= \frac{2}{R} \sum_{i,j} \beta_i \beta_j \int_{[-a,a]^2} \cos(w_i^\top x + b_i)\cos(w_j^\top x + b_j)\, dx \\ &= \frac{1}{R} \sum_{i,j} \beta_i \beta_j \int_{[-a,a]^2} \left[\cos((w_i + w_j)^\top x + (b_i + b_j)) + \cos((w_i - w_j)^\top x + b_i - b_j)\right] dx \quad (7) \end{aligned} \tag{8}$$

where we used in the second line the trigonometric identity $2\cos(a)\cos(b) = \cos(a + b) + \cos(a - b)$. Thus, $\int_{\mathcal{S}} f(x)^2 dx = \frac{1}{R} \beta^\top \tilde{M} \beta$ where $\tilde{M}$ is the matrix with entries $\tilde{m}_{i,j} = \int_{[-a,a]^2} \cos(w_i^\top x)\cos(w_j^\top x)\, dx$. We now express the elements of $\tilde{M}$.

Case 1: $w_i \neq w_j$ Let's define $\tilde{b} = b_i + b_j$ and $\tilde{w}^1 = w_i^1 + w_j^1$. Then,

$$\begin{aligned} &\int_{[-a,a]^2} \cos[(w_i + w_j)^\top x + \tilde{b}]\, dx \\ &\quad = \iint_{x_1, x_2 \in [-a,a]} \cos[(w_i^1 + w_j^1)x^1 + (w_i^2 + w_j^2)x^2 + \tilde{b}]\, dx_1\, dx_2 \\ &\quad = \iint_{x_1, x_2 \in [-a,a]} \cos(\tilde{w}^1 x^1 + \tilde{w}^2 x^2 + \tilde{b})\, dx_1\, dx_2 \\ &\quad = \int_{x_2 \in [-a,a]} \frac{1}{\tilde{w}^1} \left[\sin(\tilde{w}^1 x^1 + \tilde{w}^2 x^2 + \tilde{b})\right]_{-a}^{a} dx_2 \\ &\quad = \int_{x_2 \in [-a,a]} \frac{1}{\tilde{w}^1} \left(\sin(a\tilde{w}^1 + \tilde{w}^2 x^2 + \tilde{b}) - \sin(\tilde{w}^2 x^2 - a\tilde{w}^1 + \tilde{b})\right) dx_2 \\ &\quad = -\frac{1}{\tilde{w}^1 \tilde{w}^2} \left[\cos(a\tilde{w}^1 + \tilde{w}^2 x^2 + \tilde{b}) - \cos(\tilde{w}^2 x^2 - a\tilde{w}^1 + \tilde{b})\right]_{-a}^{a} \\ &\quad = \frac{1}{\tilde{w}^1 \tilde{w}^2} \left(\cos[a(\tilde{w}^1 - \tilde{w}^2) + \tilde{b}] + \cos[a(\tilde{w}^2 - \tilde{w}^1) + \tilde{b}] - \cos[a(\tilde{w}^1 + \tilde{w}^2) + \tilde{b}] - \cos[-a(\tilde{w}^1 + \tilde{w}^2) + \tilde{b}]\right) \end{aligned}$$

The second term of the integral is computed in a similar manner. We finally get

$$\begin{aligned}\tilde{m}_{i,j} =& \frac{\cos[a(\tilde{w}^1 - \tilde{w}^2) + \tilde{b}] + \cos[a(\tilde{w}^2 - \tilde{w}^1) + \tilde{b}] - \cos[a(\tilde{w}^1 + \tilde{w}^2) + \tilde{b}] - \cos[-a(\tilde{w}^1 + \tilde{w}^2) + \tilde{b}]}{\tilde{w}^1 \tilde{w}^2} \\ &+ \frac{+\cos[a(\bar{w}^1 - \bar{w}^2) + \bar{b}] \cos[a(\bar{w}^2 - \bar{w}^1) + \bar{b}] - \cos[a(\bar{w}^1 + \bar{w}^2) + \bar{b}] - \cos[-a(\bar{w}^1 + \bar{w}^2) + \bar{b}]}{\bar{w}^1 \bar{w}^2}\end{aligned}$$

where $\bar{b} = b_i - b_j$ and $\bar{w}^1 = w_i^1 - w_j^1$

Case 2: $w_i = w_j$

$$\begin{aligned}\tilde{m}_{i,i} &= \int_{[-a,a]^2} \cos(2(w_i^\top x + b_i))\, dx + 4a^2 \\ &= \frac{\cos[2a(w_i^1 - w_i^2) + 2b_i] + \cos[2a(w_i^2 - w_i^1) + 2b_i] - \cos[2a(w_i^1 + w_i^2) + 2b_i] - \cos[-2a(w_i^1 + w_i^2) + 2b_i]}{4(w_i^1 w_i^2)} + 4a^2\end{aligned}$$

Calculation 2

Alternatively, we might also define the kernel approximation as

$$\begin{aligned}k(x - x') &= E_w \left[\exp(-iw^\top (x - x'))\right] \\ &= E_w \left[\cos(w^\top x)\cos(w^\top x') + \sin(w^\top x)\sin(w^\top x')\right] \qquad (9) \\ &\approx \frac{1}{R}\sum_{i=1}^{R} \cos(w_i^\top x)\cos(w^\top x') + \sin(w_i^\top x)\sin(w_i^\top x') \\ &= \varphi(x)^\top \varphi(x') \qquad (10)\end{aligned}$$

where the explicit feature mapping are now defined as

$$\varphi_i(x) = \frac{1}{\sqrt{R}} \begin{bmatrix} \cos(w_i^\top x) \\ \sin(w_i^\top x) \end{bmatrix}.$$

The integral of f over $\mathcal{S}$ is now

$$\begin{aligned}\int_{[0,a]^2} f(x)^2 dx &= \frac{1}{R}\sum_{i,j} \beta_i \beta_j \int_{[0,a]^2} \varphi_i(x)\varphi_j(x')\, dx \\ &= \frac{1}{R}\sum_{i,j} \beta_i \beta_j \int_{[0,a]^2} \left(\cos(w_i^\top x) + \sin(w_i^\top x)\right)\left(\cos(w_j^\top x) + \sin(w_j^\top x)\right)\, dx\end{aligned}$$

Thus, $\int_{\mathcal{S}} f(x)^2 dx = \frac{1}{R}\beta^\top \tilde{M}\beta$ where $\tilde{M}$ is the matrix with entries given by the integral term above as before. Notice that now, the *'sin only'* and the *'cossin'* part can be expressed as before from

$$\int_{[0,a]^2} \sin(w_i^\top x)\sin(w_j^\top x)\, dx = \frac{1}{2}\int_{[0,a]^2} \left[\cos((w_i - w_j)^\top x) - \cos((w_i + w_j)^\top x)\right]\, dx$$

And

$$\int_{[0,a]^2} \cos(w_i^\top x)\sin(w_j^\top x)\, dx = \frac{1}{2}\int_{[0,a]^2} \left[\sin((w_i - w_j)^\top x) + \sin((w_i + w_j)^\top x)\right]\, dx$$

References

[1] Jesper Moller Peter McCullagh. The permanental process, 2006.

[2] Rahimi and Recht. Random features for large-scale kernel machines, 2007.