POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Bayesian Inference on Poisson-Gamma Random Measures for Spatial Statistics

Bachelor's Thesis in
Mathematical Engineering - Ingegneria Matematica

Author: **Edoardo Tarcisio Visconti**

Student ID: 983121
Advisor: Prof. Federico Bassetti
Academic Year: 2023-24

Abstract

In this Bachelor's Thesis, we investigate the Poisson-Gamma Random Measure, a shot noise Cox process introduced by Wolpert and Ickstadt in [4]. We begin by providing some background on Random Measures, with a focus on Poisson Point Processes, Completely Random Measures, and Gamma Processes. Then, the Poisson-Gamma Random Measure is presented along with theorems for its simulation. Additionally, we offer illustrative examples along with numerical methodologies to facilitate implementation.

In the second section, an overview of Markov Chain Monte Carlo (MCMC) algorithms is provided, with an emphasis on methodologies relevant to the subsequent discussion.

In the third segment, we present our contribution: the development of a Bayesian inference algorithm tailored specifically for Poisson-Gamma Random Measures.

A general model for Bayesian inference on such random measures has already been proposed in [4]. Nonetheless, we introduce an alternative approach based on discretizing the underlying Gamma process, namely a discrete Gamma process. We are confident that our model offers computational advantages, increased stability, and enables sampling from the posterior while observing multiple realizations of the same process simultaneously (unlike in [4], where only one realization is considered). Additionally, we believe that a partial result can be proved to show the efficacy of our model even in cases where the latent Gamma process is non-discrete.

Finally, we demonstrate the practical application of our approach using georeferenced wildfire data. This example serves to showcase the properties of the model and inference techniques, both in a constant scale and on a time-varying scale.

The code and the dataset are available here: **Project's Github**.

Keywords: Poisson-Gamma Random Measure, Poisson Process, Bayesian inference, Markov Chain Monte Carlo.

Abstract in lingua italiana

Nella presente Tesi, investighiamo la Misura Aleatoria Poisson-Gamma, un processo di Cox introdotto da Wolpert e Ickstadt in [4]. Iniziamo fornendo alcune basi sulle Misure Aleatorie, concentrandoci su Processi Puntuali di Poisson, Misure Aleatorie Complete e Processi Gamma. Successivamente, viene presentata la Misura Aleatoria Poisson-Gamma insieme a teoremi per la sua simulazione. Inoltre, offriamo esempi illustrativi insieme a metodologie numeriche per facilitarne l'implementazione.

Nella seconda sezione, viene fornita una panoramica degli algoritmi di Markov Chain Monte Carlo (MCMC), con un'attenzione particolare alle metodologie rilevanti per la discussione successiva.

Nella terza parte, presentiamo il nostro contributo: lo sviluppo di un algoritmo di inferenza Bayesiana specifico per le Misure Aleatorie Poisson-Gamma.

Un modello generale per l'inferenza Bayesiana su tali Misure Aleatorie è già stato proposto in [4]. Tuttavia, introduciamo un approccio alternativo basato sulla discretizzazione del processo Gamma sottostante, noto come processo Gamma discreto. Siamo fiduciosi che il nostro modello offra vantaggi computazionali, maggiore stabilità e consenta di campionare dalla posterior osservando contemporaneamente più realizzazioni dello stesso processo (a differenza di [4], dove viene considerata solo una realizzazione). Inoltre, riteniamo che si possa dimostrare un risultato parziale per mostrare l'efficacia del nostro modello anche nei casi in cui il processo Gamma latente non è discreto.

Infine, dimostriamo l'applicazione pratica del nostro approccio utilizzando dati georeferenziati su incendi boschivi. Questo esempio serve a mostrare le proprietà del modello e le tecniche di inferenza, sia su scala costante che su scala variabile nel tempo.

Il codice e il dataset sono disponibili qui: **GitHub del progetto**.

Parole chiave: Misure Aleatorie Poisson-Gamma, Processo di Poisson, Inferenza Bayesiana, Markov Chain Monte Carlo

Contents

1 | Random measures and their simulation

1.1. The Poisson Random Measure

This section offers a brief but rigorous overview of the Poisson random measure, commonly referred to as the Poisson point process.

Notation: Given a measure space $(E, \mathcal{E})$ we consider the space $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ where

$$\mathcal{M} = \{\text{sigma-finite measures } \eta \text{ on E such that } \forall B \in \mathcal{E},\ \eta(B) \in \mathbb{N} \cup \{\infty\}\}$$

and $\mathcal{B}(\mathcal{M})$ is the σ-algebra generated by the sets

$$\{\eta \in \mathcal{M}\colon\ \eta(B) = k\}$$

for $B \subset E$ measurable and $k \in \mathbb{N}$

Definition 1.1.1. A point process on $(E, \mathcal{E})$ is a stochastic process

$$M = (M(B))_{B \in \mathcal{E}}$$

with values in $\mathbb{N} \cup \{\infty\}$ such that $M \in \mathcal{M}$ a.s.

Intuitively, a point process indicates how many points there are in each region B of the space, i.e. for fixed B, the random integer $M(B)$ represents the number of points in B. The condition $M \in \mathcal{M}$ a.s. ensures that all the numbers of points in different regions are compatible with each other.

Definition 1.1.2. A Poisson point process (or Poisson random measure) on $(E, \mathcal{E})$ with intensity measure μ ($\sim ppp(\mu)$) is a point process M such that:

1. for all $B_1, \ldots, B_k \subset E$ measurable and disjoint, $M(B_1), \ldots, M(B_k)$ are independent;
2. for all $B \subset E$ measurable, $M(B) \sim Poisson(\mu(B))$.

Remark 1.1.1. From 2. we deduce that $\mathbb{E}[M(B)] = \mu(B)\ \forall B$ measurable, which is the expected number of points in the region B.

Theorem 1.1.1. Let $M \sim ppp(\mu)$ on $(E, \mathcal{E})$. Let $\tau = M(E)$ i.e. the total number of points in E. There exists some random variables $X_n \in E, n > 0$ such that:

$$M = \sum_{n=1}^{\tau} \delta_{X_n} \text{ a.s.}$$

Theorem 1.1.2. Let $\mu(E) < \infty$ and $Z, (X_i)_{i\geq 1}$ be independent random variables such that:

$$Z \sim Poisson(\mu(E)), \quad X_i \sim \frac{\mu(\cdot)}{\mu(E)}.$$

Then: $M = \sum_{n=1}^{Z} \delta_{X_n}$ is a $ppp(\mu)$ on E.

Remark 1.1.2. If $\mu(E) < +\infty$, this theorem allows us to simulate $M \sim ppp(\mu)$ with the following scheme:

Algorithm 1.1 Simulation of $M \sim ppp(\mu)$

1: Evaluate $A = \mu(E)$
2: Draw one sample z from $Z \sim Poisson(A)$
3: Draw z independent samples from $X_i \sim \frac{\mu(\cdot)}{\mu(E)}$

Where the last step can be performed with an Acceptance-Rejection Method or any other suitable methods.

As for standard random variables, point processes can be characterised by their Laplace Functional:

Definition 1.1.3. Let M be a point process on $(E, \mathcal{E})$, for every $u : E \to \mathbb{R}_{>0}$ measurable we define the Laplace Functional as:

$$L_M(u) = \mathbb{E}\left[\exp\left(-\int_E u(dx)M(dx)\right)\right]$$

Remark 1.1.3. We can interpret $\int_E u(dx)M(dx)$ as: $\sum_{x^*|\text{ "points of M"}} u(x^*)$ counted with multiplicities.

Theorem 1.1.3. (Characterization via Laplace Functional). Let μ be a σ-finite measure on $(E, \mathcal{E})$. Let M be a point process on E. The following are equivalent:

1. $M \sim ppp(\mu)$
2. for every $u : E \to \mathbb{R}_{>0}$ measurable:

$$L_M(u) = \exp\left(-\int_E \left(1 - e^{-u(x)}\right) \mu(dx)\right)$$

This theorem can be seen as a definition for a Poisson point process.

Remark 1.1.4. A standard Poisson process on the real line is special case of Poisson random measure where $E = \mathbb{R}$ and $\mu(A) = \lambda|A|$. It is it possible to prove that it can be simulated by adding independent exponential random variables.

Proof for Theorem 1.1.3 and other material on the Poisson process is available in [1].

1.2. Completely Random Measures and the Gamma Process

In this subsection, a short overview of Completely Random Measures is provided. Then the Gamma process is introduced.

Technical details and the proofs of the theorems are in: [1], [2], [3].

Definition 1.2.1. A Completely Random Measure (CRM) on a Polish space $(E, \mathcal{E})$ is a random measure M such that, for all $B_1, \dots, B_k \subset E$ measurable and disjoint, $M(B_1), \dots, M(B_k)$ are independent.

Under general assumptions, a CRM M can be written as the sum of three independent random measures:

- a deterministic measure;
- an atomistic random measure M_D (with fixed atoms), $\sum_{i \geq 1} S_i \delta_{\theta_i^*}$ where the points $\theta_1^*, \theta_2^*, \dots$ are fixed in E and S_i are independent infinitely divisible positive random variables (e.g. a Gamma RV);
- the ordinary component M_O, i.e. a discrete random measure M_O without fixed atoms.

For our purposes, we can restrict to a CRM M without deterministic component.

Then, according to Kingman's representation theorem, $\forall u \in BM_+(E)$ (measurable positive, bounded functions on E), M is characterized by the Laplace functional (proper hypotheses in [2]):

$$L_M(u) = \exp\left(- \sum_i \psi_{S_i}(v) f(\theta_i^*) - \int_{\mathbb{R}_{>0} \times E} \left(1 - e^{-vu(x)}\right) \mu(dvdx) \right) \tag{1.1}$$

where $\psi_{S_i} = -\log\left(\mathbb{E}[\exp(-vS_j)]\right)$ and μ, the Levy measure, is a σ-finite diffuse measure on the Borel sets of $\mathbb{R}_{>0} \times E$, such that, for every bounded measurable set $A \subset E$, it holds:

$$\int_{\mathbb{R}_{>0} \times A} \min(1, v) \mu(dvdx) < \infty$$

We will write:

$$M \sim CRM(\mu, \underline{\theta}^*, \Psi)$$

In the following, we will allow μ to have an atomistic component, i.e. $\mu = \mu(dvd\theta) + \sum_{\theta_i^* \in \underline{\theta}^*} \mu(dv, \theta_i^*)$ and accordingly, we can write $M \sim CRM(\mu, \emptyset)$ if

$$L_M(u) = \exp\left(-\int_{\mathbb{R}_{>0}\times E} \left(1 - e^{-vu(x)}\right) \mu(dvdx)\right). \tag{1.2}$$

Note that the notation $M \sim CRM(\mu, \emptyset)$ does not necessarily mean that M has no fixed atoms, since, now, they may arise from μ. It can also be proved that the atoms that arise from μ, if they exist, they have necessarily infinite divisible weights.

Theorem 1.2.1. If M is a CRM with Laplace functional (1.2) then:

$$M(A) = \int_{\mathbb{R}_{>0}\times A} vN(dvdx) = \sum_{j\geq 1} v_j \delta_{x_j}(A)$$

for a suitable Poisson random measure $N = \sum_{j\geq 1} \delta_{(v_j, x_j)}$ on $\mathbb{R}_{>0} \times E$ with intensity measure μ.

Definition 1.2.2. A Homogeneous completely random measure of parameter (ν, H), for short hCRM(ν, H), is a CRM with Laplace functional (1.2) in the special case:

$$\mu(dvdx) = \nu(dv)H(dx),$$

where H is a boundedly finite measure on E and ν a σ-finite measure on $\mathbb{R}_{>0}$ such that

$$\int_{\mathbb{R}_{>0}} min(1, v)\nu(dv) < \infty$$

Remark 1.2.1. A Poisson random measure (or Poisson point process) M on E with intensity measure H, $M \sim ppp(H)$, can be seen as special case of hCRM where the Levy intensity is $\nu(dv) = \delta_1(dv)$. Then the Levy measure is $\mu(dvdx) = \delta_1(dv)H(dx)$.

Definition 1.2.3. A Gamma random measure (or Gamma process) W on E with base measure H and rate $\beta > 0$, written $\mathcal{G}aP(H, \beta)$, is a hCRM with Levy intensity $\nu(dv) = v^{-1}e^{-\beta v}dv$. In this case the Levy measure is $\mu(dvdx) = v^{-1}e^{-\beta v}dvH(dx)$ and the Laplace functional (1.2) reduces to:

$$L_W(f) = \exp\left(-\int_E \log\left(1 + \frac{f(x)}{\beta}\right) H(dx)\right).$$

Corollary 1.2.1. From Theorem 1.2.1, if H is boundedly finite and $W \sim \mathcal{G}aP(H, \beta)$, then:

$$W(A) = \int_{\mathbb{R}_{>0} \times A} vN(dvdx) = \sum_{j \geq 1} v_j \delta_{x_j}(A)$$

where $N = \sum_{j \geq 1} \delta_{(v_j, x_j)}$ is a Poisson Process on $\mathbb{R}_{>0} \times E$ with intensity measure $\mu(dvdx) = v^{-1} e^{-\beta v} dv H(dx)$.

From Definition 1.2.3, we can generalize the Gamma process to non-constant $\beta(x) > 0$. Note that now this process is no more a hCRM as the Levy measure does not factorise.

Definition 1.2.4. A generalised Gamma process (or non-homogeneous Gamma process) W with parameters H and $\beta(x) > 0$, written $W \sim \mathcal{G}aP(H(dx), \beta(x))$ is characterised by the Laplace functional:

$$L_W(f) = \exp\left(- \int_E \log\left(1 + \frac{f(x)}{\beta(x)}\right) H(dx) \right).$$

Note that W is a Completely Random Measure ($W \sim CRM(\mu, \emptyset)$) with $\mu(dvdx) = v^{-1} e^{-\beta(x) v} dv H(dx)$ if $\int_{\mathbb{R}_{>0} \times A} min(1, v) \mu(dvdx) < \infty$.

1.3. Simulation of the Gamma Process

Let's see now how to simulate a generalised Gamma process.

Context for Theorem 1.3.1 and and its proof are available in [4], while more on Remark 1.3.1 can be found in [5].

In the following we will use: $E_1(t) = \int_t^\infty e^{-u}u^{-1}$.

Theorem 1.3.1. Let $\alpha(x) \geq 0$ and $\beta(x) > 0$ be measurable functions on the space E. Let $\{\sigma_m\}_{m\in\mathbb{N}}$ be identical distributed draws for any probability $\Pi(dx)$ on E, and let $\{\tau_m\}_{m\in\mathbb{N}}$ be the successive jumps of a standard Poisson Process (on the real line). Set $\tau(u,x) = E_1\left[u\beta(x)\right]\alpha(x)$, and $v_m = \sup\{u \geq 0 : \tau(u,\sigma_m) \geq \tau_m\}$, that is:

$$v_m = E_1^{-1}\left[\frac{\tau_m}{\alpha(\sigma_m)}\right]\beta(\sigma_m)^{-1}$$

or $v_m = 0$ if $\alpha(\sigma_m) = 0$.
Then, the random measure defined by $\Gamma(dx) := \sum_{m\in\mathbb{N}} v_m\delta_{\sigma_m}(dx)$ is a generalised Gamma process $\Gamma \sim \mathcal{G}aP(\alpha(dx),\beta(x))$ where $\alpha(dx) = \Pi(dx)\alpha(x)$

In our setting, we will refer to the $\{v_m\}_{m\leq M}$ as the weights of the Gamma process and to the $\{\sigma_m\}_{m\leq M}$ as the locations.

Corollary 1.3.1. The Gamma process $\Gamma \sim \mathcal{G}aP(\alpha(dx),\beta(x))$ can be approximated in distribution to arbitrarily high accuracy by the following:

Algorithm 1.2 Inverse Levy measure algorithm

1: Initialisation: Fix a large integer M and choose $\Pi(dx)$, a convenient distribution on E, such that $\alpha(x) = \frac{\alpha(dx)}{\Pi(dx)}$
2: Generate M independent identically distributed draws $\{\sigma_m\}_{m\leq M}$ from $\Pi(dx)$.
3: Generate $\{\tau_m\}_{m\leq M}$, the first M jumps of a standard Poisson process on the real line.
4: Set $v_m = E_1^{-1}\left[\frac{\tau_m}{\alpha(\sigma_m)}\right]\beta(\sigma_m)^{-1}$.
5: Set $\Gamma(dx) \approx \Gamma_M(dx) = \sum_{m\leq M} v_m\delta_{\sigma_m}(dx)$.

Remark 1.3.1. For numerical simulation it is useful to know that E_1 can be seen as:

$$E_1(x) = \lim_{a\to 0}\left[\int_x^\infty e^{-t}t^{a-1}dt\right]$$

and therefore it can be computed as the limit of rescaled χ^2 tail probabilities as the degrees

of freedom $d = 2a$ tend to zero:

$$E_1(x) = \lim_{d\to 0} \Gamma(d/2) Pr[\chi^2_d > 2x] = \lim_{d\to 0} (2/d) Pr[\chi^2_d > 2x]$$

Therefore, if we know the inverse $CDF_d(x)$ of a χ^2_d, i.e. $PPF_d(x)$ (it is implemented in many scripting languages):

$$E_1^{-1}(x) = \lim_{d\to 0} \frac{1}{2} PPF_d(1 - (d/2)y)$$

Let's see now a numerical example.

Example 1.3.1. Choosing: $E = [0, 2] \times [0, 2]$, $M = 1000$, $\Pi(dx) = \frac{1}{4}dx$ (i.e. the uniform measure on the square), $\alpha(x) = 10||x||_1$, $\beta(x) = 2 + x_1$, we can draw a sample from $W \sim \Gamma$ with Algorithm 1.2. A graphical representation of the sample is shown in the picture below (Figure 1.1).

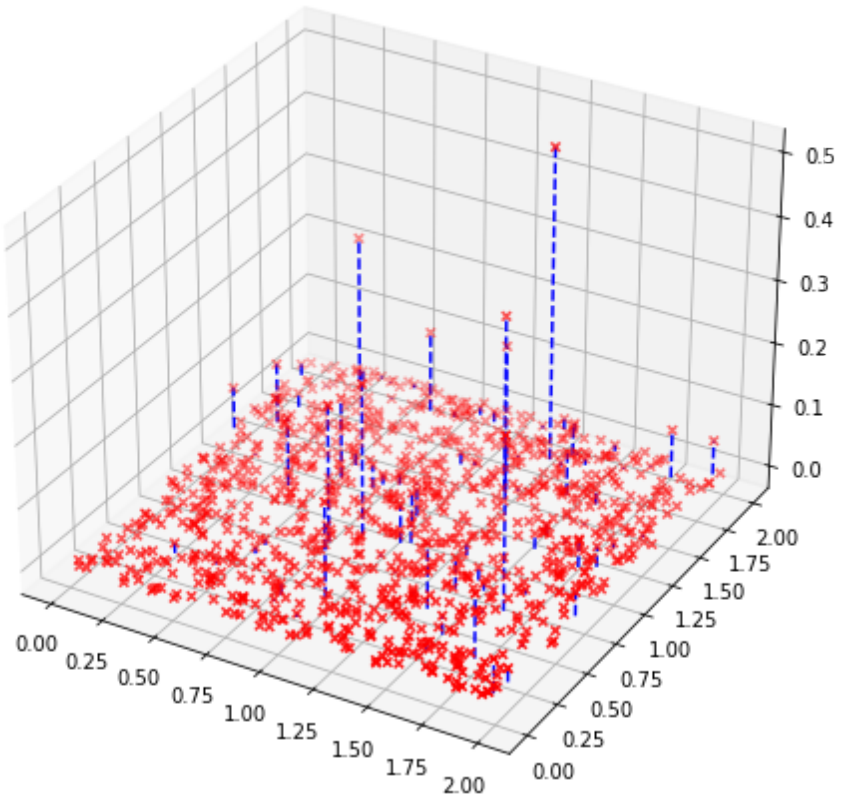

Figure 1.1: A sample from the Gamma process W on E as described in the example above.

1.4. Cox process and the Poisson-Gamma Random Measure

The Poisson-Gamma random measure is a special case of Cox point process:

Definition 1.4.1. Let $\Lambda(dx)$ be a non-negative, a.s. locally integrable, random measure on E. Then a point Process N defined on E, is a Cox point process driven by Λ, if the conditional distribution of N given Λ is a Poisson point process with intensity function Λ. In short:

$$N(dx)|\Lambda \sim ppp(\Lambda(dx)).$$

The Cox process is also known as doubly-stochastic Poisson process, as two degrees of uncertainty are introduced: first, there is a underlying random measure Λ, then only a Poisson Point Process N is observed (whose intensity measure is Λ).

In particular, we need to define shot noise Cox processes, where $\Lambda(dx)$ is built as in the following:

1. First we consider a point process W on E to create a shot noise term (shot noise is just the sum of some function over all the points of a process).
2. Then we introduce a (non-negative) kernel function $k_\psi(\cdot,\cdot)$ on E.
3. Now, $\Lambda(x) = \int_E k_\psi(x,y)W(dy)$ and $\Lambda(dx) = \Lambda(x)w(dx)$ for some reference measure $w(dx)$ on E (we will use the Lebesgue measure or a rescaled version of it).

We are now ready for:

Definition 1.4.2. A Poisson-Gamma random measure is a shot noise Cox process $N(dx)$, where W is a Gamma process, and $\Lambda(dx)$ is derived as above from $\Lambda(x) = \int_E k_\psi(x,y)W(dy)$ for some integral kernel $k_\psi(\cdot,\cdot)$.

Note that ψ is an arbitrary and fixed parameter that is associated with the choice of the integral kernel $k_\psi(\cdot,\cdot)$.

Example 1.4.1. Setting $E = [0, 100] \times [0, 100]$, $M = 1000$, $\beta^{-1} = 1.5 * 10^5$, $H(dx) = \alpha(dx) = 70 * \beta$;

First, we simulate the Gamma process $W \sim \mathcal{G}aP(\alpha(dx), \beta(x))$, with Algorithm 1.2 (represented in Figure 1.2).

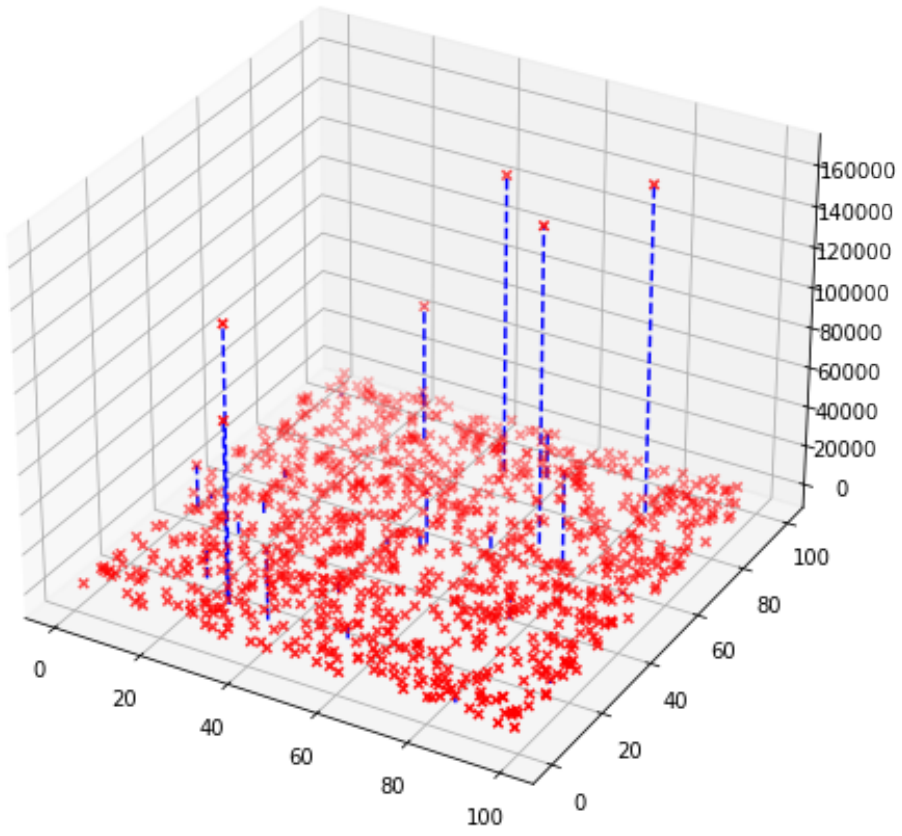

Figure 1.2: The sample we have drawn from W as described above.

In this project, as in [4], we will model $k_\psi(x, y)$ as a Gaussian kernel density: $k_\psi(x, y) = \frac{1}{\pi\psi^2} \exp\left(-\frac{|x-y|^2}{\psi^2}\right)$. Therefore, Λ can be seen as a "mixture" of Gaussian kernels whose weights and means are the weights and the locations of the underlying Gamma process.

Choosing $\psi = 5$, $w(dx) = 10^{-4}dx$, we can now compute $\Lambda(dx)$ (as explained after Definition 1.4.1) and plot it in Figure 1.3:

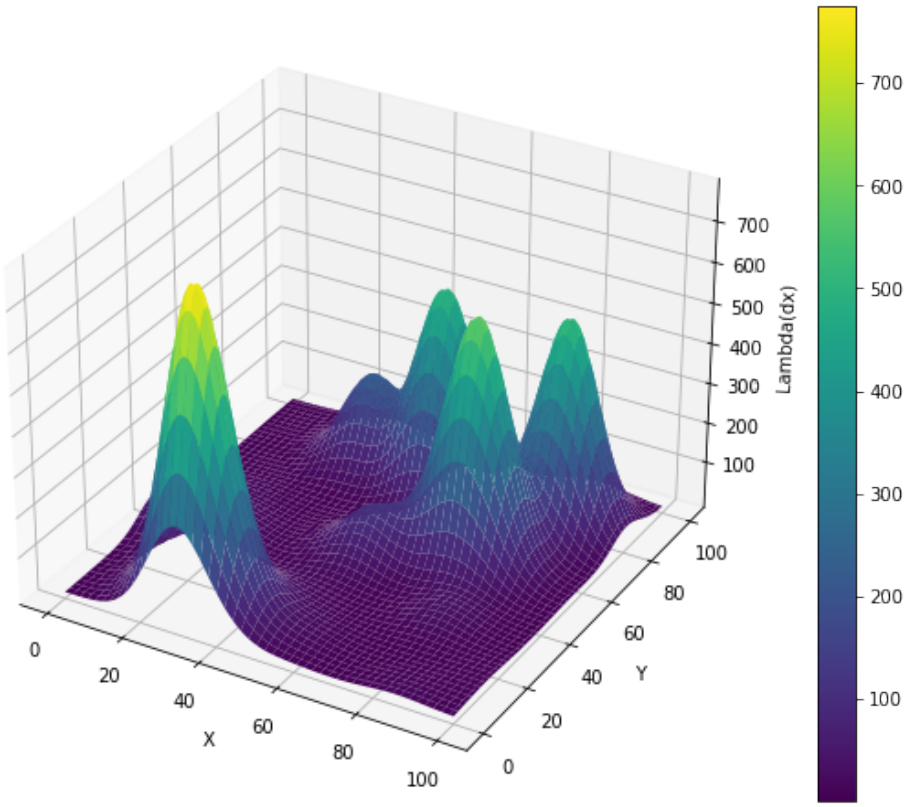

Figure 1.3: $\Lambda(x)$ computed as explained after Definition 1.4.1.

In the end, using Algorithm 1.1, we simulate $N \sim ppp(\Lambda)$. In particular:

1. to compute $\Lambda(E)$, instead of a bi-dimensional quadrature, it's easier and faster to take advantage of the geometry of the domain and of the properties of Gaussian kernels. Therefore, if $E = [a, b] \times [c, d]$ and $\Phi(x)$ is the CDF of the standard normal distribution:

$$\Lambda(E) = \sum_{m \leq M} \underbrace{\left\{ v_m \left[\Phi\left(\frac{b - \sigma_m^x}{\psi}\right) - \Phi\left(\frac{a - \sigma_m^x}{\psi}\right) \right] \left[\Phi\left(\frac{d - \sigma_m^y}{\psi}\right) - \Phi\left(\frac{c - \sigma_m^y}{\psi}\right) \right] \right\}}_{w_m}$$

where $\{v_m\}_{m \leq M}$ are the weights and $\{\sigma_m\}_{m \leq M} = \{(\sigma_m^x, \sigma_m^y)\}_{m \leq M}$ are the locations of latent the Gamma process.

2. To sample from $X_i \sim \frac{\Lambda(\cdot)}{\Lambda(E)}$ one can use a common Acceptance-Rejection method; however, it is (again) faster and easier to take advantage of the structure of Λ and follow the scheme:

Algorithm 1.3 Sampling from $\frac{\Lambda(\cdot)}{\Lambda(E)}$

1: Sample $i \leq M$ with probability $p_m = \frac{w_m}{\Lambda(E)}$
2: **while** $x = (x_i, y_i) \notin E$ **do**
3: Sample $x_i \sim \mathcal{N}(\sigma_i^x, \psi^2)$
4: Sample $y_i \sim \mathcal{N}(\sigma_i^y, \psi^2)$
5: **end while**

Note that step 1 of the algorithm above can be seen as the choice of which point of the Gamma process generates the i-th observation of the Cox process.

Eventually, in Figure 1.4, we plot the points we have drawn from the Poisson-Gamma Process $N(dx)$ on E over the contour plot of $\Lambda(dx)$.

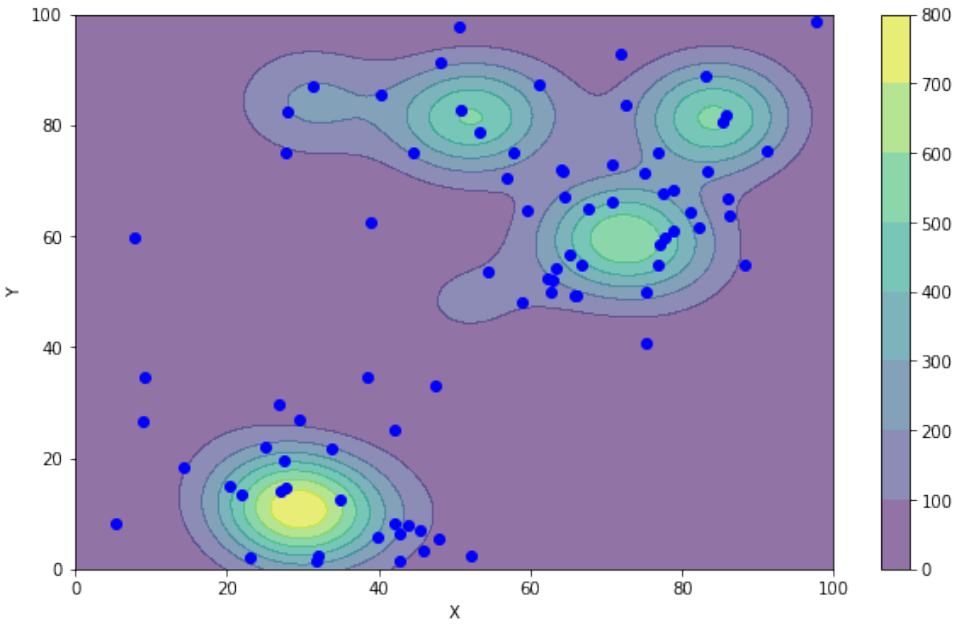

Figure 1.4: Points of the Poisson-Gamma process $N(dx)$ on E over the contour plot of $\Lambda(x)$.

2 | Bayesian Inference

2.1. The Bayesian framework

Given an iid sample $\underline{X} = (X_1, \ldots, X_n)$ from a density $f_\theta(x)$, with an unknown parameter $\theta \in \Theta$, the associated likelihood function is:

$$\mathcal{L}(\underline{X}|\theta) = \prod_i^n f_\theta(x_i).$$

This quantity is a fundamental entity for the analysis of the information provided about θ by the sample $\underline{X}$, and and Bayesian analysis relies on this function to draw inference on θ. The major input of the Bayesian approach, compared with a standard likelihood approach, is that it modifies the likelihood into a posterior distribution, which is a probability distribution on Θ defined by:

$$\pi(\theta|\underline{X}) = \frac{\mathcal{L}(\underline{X}|\theta)\pi(\theta)}{\int_\Theta \mathcal{L}(\underline{X}|\theta)\pi(\theta)d\theta}.$$

The factor $\pi(\theta)$ is called the "prior" and it has to be "chosen" to start the analysis. A first motivation for this approach is that the prior distribution summarizes the information that is available on θ prior to the observation of the sample $\underline{X}$. However, the choice of $\pi(\theta)$ is often decided on practical grounds rather than strong subjective beliefs or prior information.

The concept that is at the core of Bayesian analysis is that one should provide an inferential assessment conditional on the realized value of $\underline{X}$, and Bayesian inference gives a proper probabilistic meaning to this conditioning by allocating to θ a probability distribution.

2.2. Markov chain Monte Carlo

The posterior distribution is often not directly computable in closed form. Therefore, to sample from such a probability distribution, Markov Chain Monte Carlo (MCMC) methods are employed.

By creating a Markov chain $\{\theta^{(t)}\}_{t\geq 1}$ where the desired distribution $\pi(\theta)$ is its equilibrium distribution, we can generate a sample of π by recording states from the chain. If an algorithm that generates such a chain can be constructed, the ergodic theorem guarantees that, in almost all settings, the empirical average (over time) $\frac{1}{T}\sum_{t=1}^{T} g(\theta^{(t)})$ converges to $\mathbb{E}(g(\theta))$, independently from the starting value.

Several algorithms exist for constructing such chains, including Metropolis-Hastings and Gibbs sampling, among others. These algorithms iteratively generate states of the Markov chain, ensuring that over time, the states converge to the desired distribution.

2.2.1. Metropolis-Hastings algorithm

We want to sample from a target distribution $\pi(\theta)$ (known up to a normalizing constant). Typically (in this project) it will be the posterior, that is $\pi(\theta|\underline{X})$ but for simplicity we omit the dependence on $\underline{X}$. Note that the method is suitable to sample from any distribution.

To do so, we need to:

1. define a proposal kernel, i.e. a probability density of moving from θ_{old} to θ_{new}:

$$q(\theta_{new}|\theta_{old})$$

2. define an acceptance probability:

$$P(\theta_{new}, \theta_{old}) = \min\left(1, \frac{\pi(\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})q(\theta_{new}|\theta_{old})}\right)$$

Then, the Metropolis–Hastings algorithm is as follows:

Algorithm 2.1 Generic Metropolis–Hastings Sampler

1: Initialisation: Choose an arbitrary starting value $\theta^{(0)}$
2: **for** $t \geq 1$ **do**
3: Given $\theta^{(t-1)}$, sample $\hat{\theta}$ from $q\left(\theta|\theta^{(t-1)}\right)$
4: Compute $P = P\left(\hat{\theta}, \theta^{(t-1)}\right)$ as above
5: With probability P accept $\hat{\theta}$ and set $\theta_t = \hat{\theta}$; otherwise, reject $\hat{\theta}$ and set $\theta^{(t)} = \theta^{(t-1)}$
6: **end for**

Remark 2.2.1. While theoretical guarantees that the algorithm converges are high, the choice of q remains essential in practice. Poor choices of q may indeed result either in a very high rejection rate, meaning that the Markov chain hardly moves, or in a myopic exploration of the support of π, that is, in a dependence on the starting value θ_0 such that the chain is stuck in a neighborhood region of θ_0. A particular choice of proposal q may thus work well for one target density but be extremely poor for another one. While the algorithm is indeed universal, it is impossible to prescribe application-independent strategies for choosing q.

Choosing particular $q(\theta_{new}|\theta_{old})$ we can have:

- The **Independence Sampler**: when q is independent from θ_{old}, i.e.

$$q(\theta_{new}|\theta_{old}) = q(\theta_{new}),$$

and then P simplifies to

$$P(\theta_{new}, \theta_{old}) = \min\left(1, \frac{\pi(\theta_{new})q(\theta_{old})}{\pi(\theta_{old})q(\theta_{new})}\right).$$

- The **Random Walk Sampler**: when q is a symmetric transition kernel i.e.:

$$q(\theta_{new}|\theta_{old}) = f(\theta_{new} - \theta_{old}),$$

where f is a symmetric density function and then P simplifies to

$$P(\theta_{new}, \theta_{old}) = \min\left(1, \frac{\pi(\theta_{new})}{\pi(\theta_{old})}\right).$$

For example f can be a normal density with mean zero and given variance (covariance matrix).

2.2.2. Gibbs Sampler

Let $\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$ and set $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p)$.

In this case, to sample from $\pi(\theta)$, we should be able to compute (by hand) all the full conditional distributions:

$$\pi_i(\theta_i|\theta_{-i}) \quad \forall i = 1, \ldots, p$$

Algorithm 2.2 Gibbs Sampler

1: Initialisation: Choose an arbitrary starting value $\theta^{(0)} = \left(\theta_1^{(0)}, \ldots, \theta_p^{(0)}\right)$
2: **for** $t \geq 1$ **do**
3: Given $\theta^{(t-1)}$, we sample θ^t with the following scheme:
4: 1) $\theta_1^{(t)} \sim \pi_1\left(\theta_1|\theta_2^{(t-1)}, \ldots, \theta_p^{(t-1)}\right)$
5: 2) $\theta_2^{(t)} \sim \pi_2\left(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_p^{(t-1)}\right)$
6: ...
7: p) $\theta_p^{(t)} \sim \pi_p\left(\theta_p|\theta_1^{(t)}, \ldots, \theta_{p-1}^{(t)}\right)$
8: **end for**

Remark 2.2.2. Metropolis-within-Gibbs algorithm
Several MCMC algorithms can be mixed together within a single algorithm using either a circular or a random design. While this construction is often suboptimal (in that the inefficient algorithms in the mixture are still used on a regular basis), it almost always brings an improvement compared with its individual components. A special case where a mixed scenario is used is the Metropolis-within-Gibbs algorithm: when building a Gibbs sampler, it may happen that it is difficult or impossible to simulate from one or several of the conditional distributions. In that case, a single Metropolis step associated with this conditional distribution (as its target) can be used instead.

3 | Bayesian inference for Poisson-Gamma Random measures

A Poisson-Gamma random measure $N(dx)$ can be modelled as Cox process with the following hierarchy:

$$\alpha, \beta \sim \pi(\alpha, \beta) \qquad W(dx)|\alpha, \beta \sim \mathcal{G}aP(\alpha(dx), \beta)$$

$$\Lambda(dx) = \int_E k_\psi(dx, y)W(dy) \qquad N(dx)|\alpha, \beta, W \sim ppp(\Lambda(dx))$$

Now, given one or more observation of N we want to sample from the posterior distribution of (α, β). On the other side, we derive ψ, the spatial interaction, from the physical nature of the phenomenon we would like to study.

A model for Bayesian inference on Poisson-Gamma random measures has already been proposed in [4], where the inference is performed with no restrictions to $\alpha(dx)$. However, usually, no prior information is available on the latent process $W(dx)$ and $\alpha(dx)$ is chosen equal to the Lebesgue measure on E rescaled by a factor $\alpha \in \mathbb{R}$. Therefore the inference reduces to estimating $\alpha, \beta \in \mathbb{R}$.

In the following we propose a different method, where $\alpha(dx)$ is modelled as a discrete measure and we will suggest an idea that (partially) connects the two methods. Moreover, we believe our model, although "simpler", is computationally less expensive, more stable and will allow us to sample from the posterior observing more realizations of the same process at once (while in [4] only one realization is taken in consideration at a time). In the following we also refer to this version of the latent Gamma process (with $\alpha(dx)$ discrete), as a discrete Gamma process (which is still a Gamma process and a CRM).

3.1. On the discrete Gamma Process

Definition 3.1.1. A Gamma process on a space E ($\sim \mathcal{G}aP(H_D, \beta)$) is a discrete Gamma process if $H_D(dx)$ is a discrete measure on E.

Proposition 3.1.1. A discrete Gamma process $W(dx) \sim \mathcal{G}aP(H_D, \beta)$ on a space E where $H_D(dx) = \sum_{m=1}^{M} \alpha_m \delta_{\sigma_m}(dx)$ and $\{\sigma_m\}_{m \leq M}$ are fixed points of E, is equivalent to:

$$W(dx) = \sum_{m=1}^{M} v_m \delta_{\sigma_m}(dx)$$

where $v_m \sim \mathcal{G}a(\alpha_m, \beta)$ $\forall m = 1, \ldots, M$ and $v_1, \ldots, v_M$ are independent.

Proof. Since the Gamma process W is a CRM, for all $A_1, \ldots, A_k \subset E$ measurable and disjoint, $W(A_1), \ldots, W(A_k)$ are independent.

Moreover, $\forall A \in E$ and $\forall t > 0$, $tW(A) = \int_E t\mathbb{I}_A(x)W(dx)$.

Therefore:

$$\begin{aligned}
L_{W(A)}(t) &= \mathbb{E}\left[e^{-tW(A)}\right] \\
&= \mathbb{E}\left[e^{-\int_E t\mathbb{I}_A(x)W(dx)}\right] \\
&= \exp\left\{-\int_E \log\left(1 + \frac{t\mathbb{I}_A(x)}{\beta}\right) H_D(dx)\right\} \\
&= \exp\left\{-\log\left(1 + \frac{t}{\beta}\right) H_D(A)\right\} \\
&= \left(1 + \frac{t}{\beta}\right)^{-H_D(A)}
\end{aligned}$$

Which is equivalent to: $W(A) \sim \mathcal{G}a(H_D(A), \beta)$ if $H_D(A) > 0$ or $W(A) = 0$ a.s. if $H_D(A) = 0$.

Now, since $H_D(dx) = \sum_{m=1}^{M} \alpha_m \delta_{\sigma_m}(dx)$:

$$\begin{aligned}
&\Rightarrow H_D(A) = \sum_{m:\sigma_m \in A} \alpha_m \\
&\Rightarrow W(A) = 0 \text{ a.s.} \quad \text{if } \sigma_m \notin A \quad \forall m = 1, \ldots, M \\
&\Rightarrow W_\omega(A) = \sum_{m=1}^{M} v_m(\omega)\delta_{\sigma_m}(A) \text{ a.s.}
\end{aligned}$$

which means that $v_m = W(A_m)$ if $\sigma_m \in A_m$ and $\sigma_k \notin A_m$ $\forall k \neq m$.

Therefore, $v_m \sim \mathcal{G}a(\alpha_m, \beta)\ \forall m = 1, \dots, M$ and $v_1, \dots, v_M$ are independent. □

When modeling a Poisson-Gamma-random measure with the previous hierarchy, no prior information is usually evident or available on the latent Gamma process, and a reasonable choice is: $\alpha(dx) = \alpha \cdot m(dx)$ (where $m(dx)$ is the Lebesgue measure on E). This is equivalent to believing that, a priori, no subset of E will have a bigger probability to "generate" points of the process $N(dx)$.

In this situation, we believe it can also be chosen to model $\alpha(dx)$ as: $\alpha_D(dx) = \sum_{m=1}^{M} \frac{\alpha}{M} \delta_{\sigma_m}(dx)$ where $\{\sigma_m\}_{m \leq M}$ are equally spaced on E and $\alpha \in \mathbb{R}$ is a fixed parameter.

In fact, if:

$$W_D \sim \mathcal{G}aP(\alpha_D(dx), \beta)$$

$$W \sim \mathcal{G}aP(\alpha \cdot m(dx), \beta)$$

then $\alpha_D(x) \longrightarrow \alpha \cdot m(dx)$ weakly,

and therefore follows that:

$$\mathbb{E}\left[e^{-tW_D(A)}\right] \longrightarrow \mathbb{E}\left[e^{-tW(A)}\right]$$

for every $f : E \to \mathbb{R}_{>0}$ measurable.

An in-depth theoretical analysis is not the intent of this Thesis, however, we believe that what we have pointed out shows that we can expect a similar behaviour of W_D and W when M is chosen big.

Further results on the convergence of the process can follow from the results in chapter 14 of [8].

3.2. Sampling from the discrete Gamma Process

We propose an alternative way of sampling from the discrete Gamma process (instead of Algorithm 1.2) and from a Poisson-Gamma random measure with it as the latent process. The properties shown here are key to compute the likelihood of the model.

Let's see in detail how it is done.

We start by recalling some properties of the Gamma random variable:

If

$$v_1, \dots, v_M \overset{\text{IND}}{\sim} \mathcal{G}a(\alpha_m, \beta) \quad \forall m = 1, \dots, M$$

and

$$\overline{v} := \sum_{m=1}^{M} v_m \ , \quad D_m := \frac{v_m}{\overline{v}} \quad \forall m = 1, \dots, M$$

then:

- $\overline{v}$ and $(D_1, \dots, D_M)$ are independent,
- $\overline{v} \sim \mathcal{G}a\left(\sum_{m=1}^{M} \alpha_m, \beta\right)$
- $(D_1, \dots, D_M) \sim DIR(\alpha_1, \dots, \alpha_M)$,

  where DIR is the notation for the Dirichlet distribution, or multivariate Beta distribution, which is a generalisation on the n-simplex of the Beta distribution.

For completeness, the pdf of $v \sim \mathcal{G}(a, b)$ is:

$$f_v(v) = \frac{b^a}{\Gamma(a)} v^{a-1} e^{-bv}, \text{ where } v > 0$$

and the pdf of $(d_1, \dots, d_M) \sim DIR(a_1, \dots, a_M)$ is:

$$f_d(d_1, \dots, d_M) = \frac{\Gamma\left(\sum_{i=1}^{M} a_i\right)}{\prod_{i=1}^{M} \Gamma(a_i)} \prod_{i=1}^{M} d_i^{a_i - 1}, \text{ where } \sum_{i=1}^{M} d_i = 1 \text{ and } d_i > 0 \quad \forall i.$$

Therefore, for our purpose the latent (discrete Gamma) process can be rewritten as:

$$\begin{aligned} W(dx) &= \sum_{m=1}^{M} v_m \delta_{\sigma_m}(dx) \\ &= \overline{v} \sum_{m=1}^{M} D_m \delta_{\sigma_m}(dx) \end{aligned}$$

where $v_m, \overline{v}, D_m$ are distributed as above.

Now, drawing samples from the Poisson-Gamma random measure $N(dx)$ reduces to:

0. Choose arbitrarely:

   - M, the number of fixed locations, and $\{\sigma_m\}_{m \leq M}$ their coordinates;
   - $\alpha_i \quad \forall m = 1, \dots, M$, the shape parameter of each v_m and set $\alpha = \sum_{m=1}^{M} \alpha_m$;

- β, the (same) rate parameter of each v_m;
- ψ, the standard deviation of the Gaussian Kernel $k_\psi(x,y) = \frac{1}{\pi\psi^2}\exp\left(-\frac{|x-y|^2}{\psi^2}\right)$.

1. Sample
$$\overline{v} \sim \mathcal{G}a\,(\alpha,\beta)$$
2. Sample
$$(D_1,\dots,D_M) \sim DIR(\alpha_1,\dots,\alpha_M)$$
3. Sample
$$N \sim Pois(\Lambda(E))$$
where $\Lambda(E) = \overline{v}\sum_{m=1}^{M} D_m k_\psi(E|\sigma_m)$.
4. Choose the locations $S_1,\dots,S_n$ that generates each of the N points, with probability:
$$P\left(S_i = \sigma_m | N, D_1,\dots,D_M\right) = \frac{D_m k_\psi(E|\sigma_m)}{\sum_{m=1}^{M} D_m k_\psi(E|\sigma_m)} \quad \forall i = 1,\dots,N \quad \forall m = 1,\dots,M.$$
5. Sample the points $\{X_i\}_{i=1}^N = \{(x_i,y_i)\}_{i=1}^N$ independently with probability:
$$P(X_i \in dx | N, S_i) = \frac{k_\psi(x|S_i)dx}{k_\psi(E|S_i)} \quad \forall i = 1,\dots,N.$$

Example 3.2.1. Choosing: $E = [0,10]\times[0,10]$, $M = 64$, $\{\sigma_m\}_{m\le M}$ equally spaced on E, $\psi = 0.5$, $\beta = 2.5\cdot 10^{-2}$, $\alpha = 20$ and $\alpha_m = \alpha/M \quad \forall m = 1,\dots,M$, we can sample as above and plot the process (see Figure 3.1).

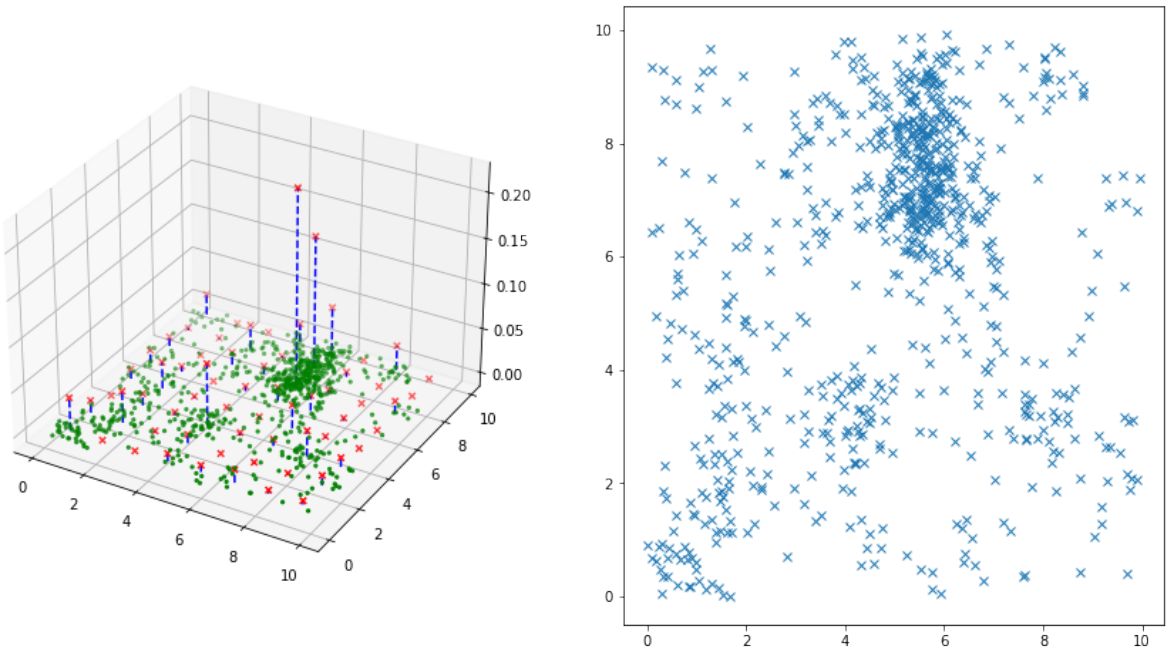

Figure 3.1: A realization of the simplified process as described in the example 3.2.1. On the left, the red crosses are the locations of the underlying Gamma Process, the blue dashed lines their weights while the green points are the points of the Poisson-Gamma random measure. On the right, the light blue crosses represent the points of the Poisson-Gamma random measure.

3.3. The Metropolis within Gibss algorithm

We are now interested in computing the likelihood of the model:

$$\begin{aligned}\mathcal{L}(\underline{X}|\alpha_1,\dots,\alpha_M,\beta) = f_{\overline{v}}(\overline{v}) f_D(D1,\dots,D_M) f_N(N|\Lambda(E)) \\ \times f_{X,S}(X_1,\dots,X_N,S_1,\dots,S_N|N,D)\end{aligned} \tag{3.1}$$

where:

$$\alpha = \sum_{m=1}^{M} \alpha_m$$

$$f_{\overline{v}}(\overline{v}) = \frac{\overline{v}^{\alpha-1} e^{-\beta\overline{v}} \beta^{\alpha}}{\Gamma(\alpha)}$$

$$f_D(D1,\dots,D_M) = \Gamma(\alpha) \cdot \prod_{m=1}^{M} \frac{D_m^{\alpha_m-1}}{\Gamma(\alpha_m)}$$

$$\Lambda(E) = \overline{v} \sum_{m=1}^{M} D_m k_\psi(E|\sigma_m)$$

$$f_N(N|\Lambda(E)) = \frac{\Lambda(E)^N e^{-\Lambda(E)}}{N!} \mathbb{I}_{\mathbb{N}}(N)$$

$$f_{X,S}(X_1, \dots, X_N, S_1, \dots, S_N | N, D) = \prod_{i=1}^{N} \frac{D_i k_\psi(E|S_i)}{\sum_{m=1}^{M} D_m k_\psi(E|\sigma_m)} \frac{k_\psi(X_i|S_i)}{k_\psi(E|S_i)}.$$

Therefore, supposing $(\alpha_1, \dots, \alpha_M)$ and β independent, simplifying and rearranging the terms, the joint likelihood becomes:

$$\begin{aligned} &\mathcal{L}(\underline{X}|\alpha_1, \dots, \alpha_M, \beta)\pi(\alpha_1, \dots, \alpha_M, \beta) = \pi(\alpha_1, \dots, \alpha_M)\pi(\beta)\overline{v}^{\alpha-1}\overline{v}^N e^{-\beta\overline{v}} \\ &\times \exp\left\{-\overline{v} \sum_{m=1}^{M} D_m k_\psi(E|\sigma_m)\right\} \times \beta^\alpha \frac{1}{N!} \prod_{m=1}^{M} \frac{D_m^{\alpha_m - 1}}{\Gamma(\alpha_m)} \times \prod_{i=1}^{N} D_i k_\psi(X_i|S_i). \end{aligned} \tag{3.2}$$

We can now compute the full conditionals:

1. $S_i|All$:

$$\mathbb{P}(S_i = \sigma_m) = \frac{D_i k_\psi(X_i|S_i)}{\sum_{m=1}^{M} D_m k_\psi(X_i|\sigma_m)} \quad \forall i = 1, \dots, N \quad \forall m = 1, \dots, M$$

2. $(D_1, \dots, D_M)|All$:

$$\begin{aligned} &\propto \prod_{m=1}^{M} \left(D_m^{\alpha_m - 1} D_m^{n_m}\right) \cdot e^{-\Lambda_D(E)} \\ &\propto DIR(\alpha_1 + n_1, \dots, \alpha_M + n_M) \cdot e^{-\Lambda_D(E)} \end{aligned}$$

where $n_m = \#\{j = 1, \dots, N | S_j = \sigma_m\}$.

This full conditional is not known in closed form; nonetheless, we will sample it with a step of an Independence Sampler. Choosing as a proposal $q(D_{new}|D_{old}) = q(D_{new}) = DIR(\alpha_1 + n_1, \dots, \alpha_M + n_M)$, then the acceptance probability becomes:

$$P(D_{new}, D_{old}) = \min\left(1, \frac{e^{-\Lambda_{D_{new}}(E)}}{e^{-\Lambda_{D_{old}}(E)}}\right).$$

3. $\overline{v}|All$:

$$\propto \overline{v}^{\alpha-1}\overline{v}^{N}e^{-\beta\overline{v}} \times \exp\left\{-\overline{v}\sum_{m=1}^{M} D_m k_\psi(E|\sigma_m)\right\}$$

and therefore we recognise that:

$$\overline{v}|All \sim \mathcal{G}a\left(\alpha+N, \beta+\sum_{m=1}^{M} D_m k_\psi(E|\sigma_m)\right).$$

4. $\beta|All$:

$$\propto \pi(\beta)e^{-\beta\overline{v}}\beta^{\alpha}$$

Choosing $\pi(\beta) \sim \mathcal{G}a(a_0, b_0)$ the model is conjugate and

$$\beta|All \sim \mathcal{G}a(a_0+\alpha, b_0+\overline{v}).$$

5. For simplicity we will suppose $\alpha_m = \frac{\alpha}{M} \quad \forall m = 1, \ldots, M$ and we will sample from the posterior of α.

$\alpha|All$:

$$\propto \pi(\alpha)\overline{v}^{\alpha-1}\beta^{\alpha}\prod_{m=1}^{M}\frac{D_m^{\alpha_m-1}}{\Gamma(\alpha_m)}$$

Since we don't recognise any known distribution we will use a Random-walk Metropolis-Hastings algorithm.

Having computed all the full conditionals, the Metropolis within Gibbs Sampler is ready to be implemented.

3.4. The algorithm with more observations (time independent)

We can modify the the results above to sample from the posterior of (α, β) after we have seen multiple realizations of the same phenomenon. In the following it is supposed that the latent Gamma process doesn't change over time (so do not α and β) and each observation is independent from the others.

If T is the number of realizations, and N_t the number of each observations at time $t = 1, \ldots, T$, then the joint posterior becomes:

$$\mathcal{L}(\underline{X_1},\dots,\underline{X_T}|\alpha_1,\dots,\alpha_M,\beta)\pi(\alpha_1,\dots,\alpha_M,\beta) = \pi(\alpha_1,\dots,\alpha_M)\pi(\beta)\left(\prod_{t=1}^{T}\overline{v}_t\right)^{\alpha-1}$$
$$\times\prod_{t=1}^{T}\overline{v_t}^{N_t}\times\exp\left(-\beta\sum_{t=1}^{T}\overline{v_t}\right)\times\prod_{t=1}^{T}\exp\left(-\overline{v_t}\sum_{m=1}^{M}D_{m,t}k_\psi(E|\sigma_m)\right)$$
$$\times\beta^{\alpha T}\times\prod_{t=1}^{T}\frac{1}{N_t!}\times\prod_{t=1}^{T}\prod_{m=1}^{M}\frac{D_{m,t}^{\alpha_m-1}}{\Gamma(\alpha_m)}\times\prod_{t=1}^{T}\prod_{i=1}^{N_t}D_{i,t}k_\psi(X_{i,t}|S_{i,t}).$$

Therefore, the full conditionals of $S_{i,t}$, $D_{i,t}$, $\overline{v_t}$ $\forall i = 1,\dots,N_t$ are the same as above $\forall t = 1,\dots,T$; while, now:

1. $\beta|All$:
$$\propto \pi(\beta)\exp\left(-\beta\sum_{t=1}^{T}\overline{v_t}\right)\beta^{\alpha T}$$
Choosing again $\pi(\beta)\sim\mathcal{G}a(a_0,b_0)$ the model is still conjugate and
$$\beta|All\sim\mathcal{G}a\left(a_0+\alpha T, b_0+\sum_{t=1}^{T}\overline{v_t}\right)$$

2. Same as before, for simplicity we will suppose $\alpha_m=\frac{\alpha}{M}$ and we will sample from the posterior of α.
$\alpha|All$:
$$\propto\pi(\alpha)\left(\prod_{t=1}^{T}\overline{v}_t\right)^{\alpha-1}\beta^{\alpha T}\times\prod_{t=1}^{T}\prod_{m=1}^{M}\frac{D_{m,t}^{\alpha_m-1}}{\Gamma(\alpha_m)}$$
We don't recognise any known distribution, so we use a Random-walk Metropolis-Hastings algorithm. In order to evaluate the acceptance probability it is advisable to calculate the logarithm of the targets (see Algorithm 3.1).

Then, the algorithm to sample from the posterior of (α,β) is:

Algorithm 3.1 Metropolis within Gibbs Sampler for discrete Poisson-Gamma random measures

1: Initialisation:

- choose arbitrary starting values for: $\overline{v}_t^{(0)}, D_{m,t}^{(0)} \quad \forall t = 1, \dots, T \quad \forall m = 1, \dots, M$;
- choose arbitrary starting values for $\alpha^{(0)}, \beta^{(0)}$;
- choose wisely $\pi(\alpha)$, a prior for α, σ_{MC}^2, the variance for the proposal of α_{new}, and a_0, b_0 small $\in \mathbb{R}$, shape and rate for the prior $\pi(\beta) \sim \mathcal{G}a(a_0, b_0)$.

2: **for** $it \geq 1$ **do**

3: **for** $t = 1, \dots, T$ **do**

4: Assign $S_{i,t}^{(it)}$ with probability:

$$\mathbb{P}(S_{i,t}^{(it)} = \sigma_m) = \frac{D_{i,t}^{(it-1)} k_\psi(X_i|S_i)}{\sum_{m=1}^{M} D_{m,t}^{(it-1)} k_\psi(X_i|\sigma_m)} \quad \begin{matrix} \forall i = 1, \dots, N_t \\ \forall m = 1, \dots, M. \end{matrix}$$

5: Sample: $D_{new,t} = (D_{new,t,1}, \dots, D_{new,M,t}) \sim DIR(\alpha_1^{(it-1)} + n_{1,t}, \dots, \alpha_M^{(it-1)} + n_{M,t})$ and set $D_t^{(it)} = D_{new,t}$ with probability P_1, otherwise $D_t^{(it)} = D_t^{(it-1)}$, where:

$$P_1(D_{new,t}, D_t^{(it-1)}) = \min\left(1, \frac{e^{-\Lambda_{D_{new,t}}(E)}}{e^{-\Lambda_{D_t^{(it-1)}}(E)}}\right).$$

6: Sample: $\overline{v}_t^{(it)} \sim \mathcal{G}a\left(\alpha^{(it-1)} + N_t, \beta^{(it-1)} + \sum_{m=1}^{M} D_{m,t}^{(it)} k_\psi(E|\sigma_m)\right)$.

7: **end for**

8: Sample: $\beta^{(it)} \sim \mathcal{G}a\left(a_0 + \alpha^{(it-1)}T, b_0 + \sum_{t=1}^{T} \overline{v}_t^{(it)}\right)$.

9: Sample: $\alpha_{new} \sim \mathcal{N}(\alpha^{(it-1)}, \sigma_{MC}^2)$
and set $\alpha^{(it)} = \alpha_{new}$ with probability P_2, otherwise $\alpha^{(it)} = \alpha^{(it-1)}$, where:

$$\log P_2(\alpha_{new}, \alpha^{(it-1)}) = \min\left(0, \text{logt}(\alpha_{new}) - \text{logt}(\alpha^{(it-1)})\right)$$

and:

$$\text{logt}(\alpha) = \log(\pi(\alpha)) + (\alpha - 1)\sum_{t=1}^{T} \log\left(\overline{v}_t^{(it)}\right) + \alpha T \log\left(\beta^{(it)}\right)$$

$$+\frac{\alpha - M}{M}\sum_{t=1}^{T}\sum_{m=1}^{M} \log\left(D_{m,t}^{(it)}\right) - MT \log\left(\Gamma\left(\frac{\alpha}{M}\right)\right).$$

10: **end for**

3.5. The algorithm with more observations (time varying)

If we believe the latent process changes over time we can capture this aspect allowing α to evolve over time. In particular, if β is constant in time, then α_t is directly linked with the number of points observed at each time t; in fact, we can approximate an estimator of the number of observed points at time t with: $\hat{N}_t = \mathbb{E}(\Lambda_t(E)) \simeq \frac{\alpha_t}{\beta}$.

With α time-varying the full conditional becomes:

$$\alpha_1, \dots, \alpha_T | All \propto \pi(\alpha_1, \dots, \alpha_T) \left(\prod_{t=1}^{T} \overline{v}_t^{\alpha_t - 1} \right) \beta^{\sum_{t=1}^{T} \alpha_t} \times \prod_{t=1}^{T} \prod_{m=1}^{M} \frac{D_{m,t}^{\alpha_{t,m}-1}}{\Gamma(\alpha_{t,m})}$$

Having in mind the application in 4.2, one could model $\alpha_t = e^{\eta_0 + \eta_1 I(t) + \eta_2 A(t)}$ where:

- $I(t)$ is a dummy variable with values in $\{0, 1\}$ that allows to capture seasonality, such as dry vs wet season (if we model forest fires) or working days vs public holidays (if we model taxi pickups).
- $A(t)$ is a variable that allows to capture trend, such as increasing/decreasing number of forest fires, taxi pick-ups over weeks or years.

Now we want to sample from the posteriors of (η_0, η_1, η_2) and we do so by replacing Step 9 of Algorithm 3.1 with three steps (one for each η_i) of a RW-Metropolis sampler or one for the all three at the same time.

Step 9 of Algorithm 3.1 then becomes:

1. Sample: $(\eta_{0,new}, \eta_{1,new}, \eta_{2,new}) \sim \mathcal{N}\left(\left(\eta_0^{(it-1)}, \eta_1^{(it-1)}, \eta_2^{(it-1)} \right), \sigma_{MC}^2 \mathbb{I}_3 \right)$.
2. Compute: α_{new} and $\alpha^{(it-1)}$ using $\alpha_t = e^{\eta_0 + \eta_1 I(t) + \eta_2 A(t)}$.
3. Set $\eta_i^{(it)} = \eta_{i,new} \quad \forall i = 0, 1, 2$ with probability P_2, otherwise $\eta_i^{(it)} = \eta_i^{(it-1)}$, where:

$$\log P_2(\alpha_{new}, \alpha^{(it-1)}) = \min\left(0, \text{logt}(\alpha_{new}) - \text{logt}(\alpha^{(it)})\right)$$

and: $\text{logt}(\alpha) = \log(\pi(\eta_0)\pi(\eta_1)\pi(\eta_2)) + \sum_{t=1}^{T}(\alpha_t - 1)\log\left(\overline{v}_t^{(it)}\right) + \left(\sum_{t=1}^{T}\alpha_t\right)\log\left(\beta^{(it)}\right)$
$+ \sum_{t=1}^{T}\sum_{m=1}^{M}\left(\frac{\alpha_t}{M} - 1\right)\log\left(D_{m,t}^{(it)}\right) - M\sum_{t=1}^{T}\log\left(\Gamma\left(\frac{\alpha_t}{M}\right)\right)$.

4 | An application: inference on forest fires in the Amazon Forest

4.1. The time independent case

In this last section we propose an application of the algorithm above. Given the coordinates of forest fires in a portion of the Amazon Forest (Latitude in 30°S-20°S and Longitude in 65°W-55°W) over three different months of the dry season (August 2016, September 2017, September 2018) we would like to model the latent process generating the fires and sample from the posterior of (α, β). Figure 4.1 displays three plots, each corresponding to a different month, depicting the locations of forest fires as points on a plane based on their coordinates.

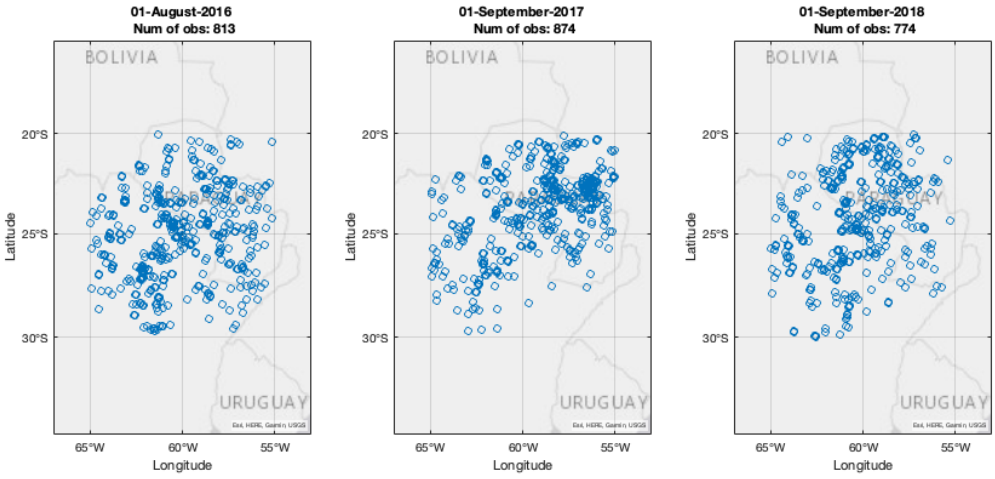

Figure 4.1: Locations of fires as points on a plane based on their coordinates, over three months.

For simplicity we decided not to take in consideration the curvature of the Earth, then we translated the observations in the square $E = [0, 10] \times [0, 10]$.

Assigning $\psi = 0.8$ a fair value for the spatial interaction, and choosing arbitrary values for the starting parameters (see the code for details), then running the sampler for 5000 iterations, we can estimate the posterior distribution of α and β (see Figure 4.2 and 4.3).

For every month $t = 1, \ldots, T$ (here $T = 3$), the algorithm also samples from the posterior of $D_t, \overline{v}_t$. We can then plot the results for every single month: see Figure 4.4 and 4.5 for an example regarding the third month.

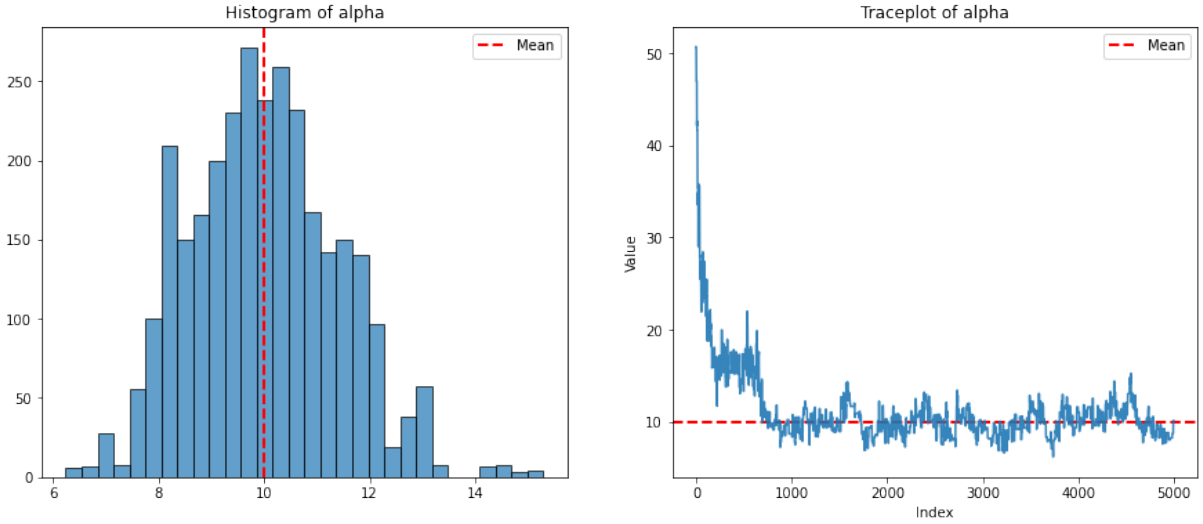

Figure 4.2: Histogram (Burn-in: 2000) and Traceplot of $\alpha^{(it)}$.

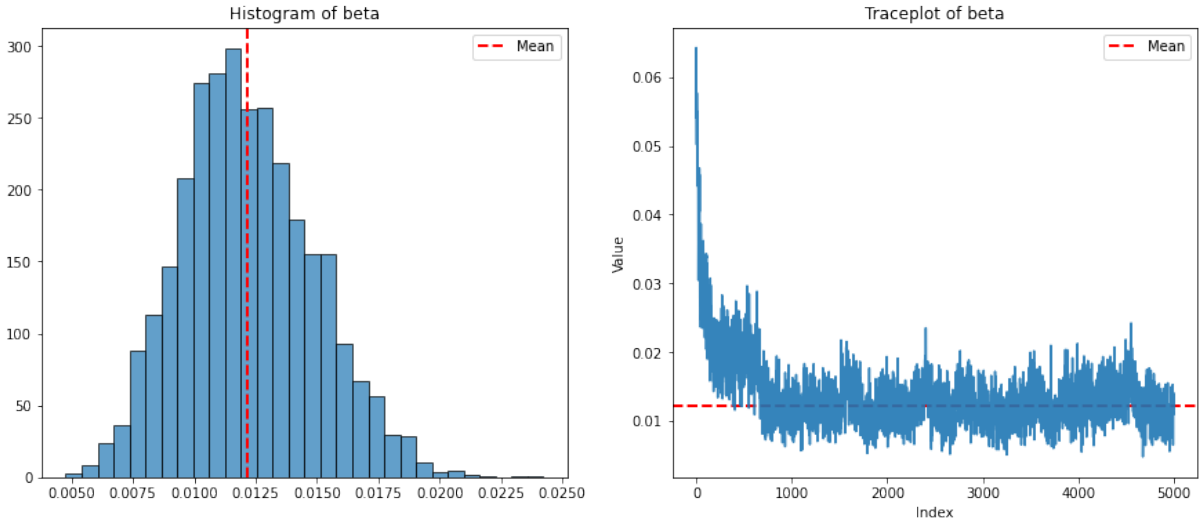

Figure 4.3: Histogram (Burn-in: 2000) and Traceplot of $\beta^{(it)}$.

Figure 4.4: On the left: plot of the coordinates of the fires (blue points) with estimates for the weights v_m (green dashed lines) of the latent process $W(dx)$; On the right: surface plot of $\mathbb{E}(\Lambda|\text{Data})$ on E.

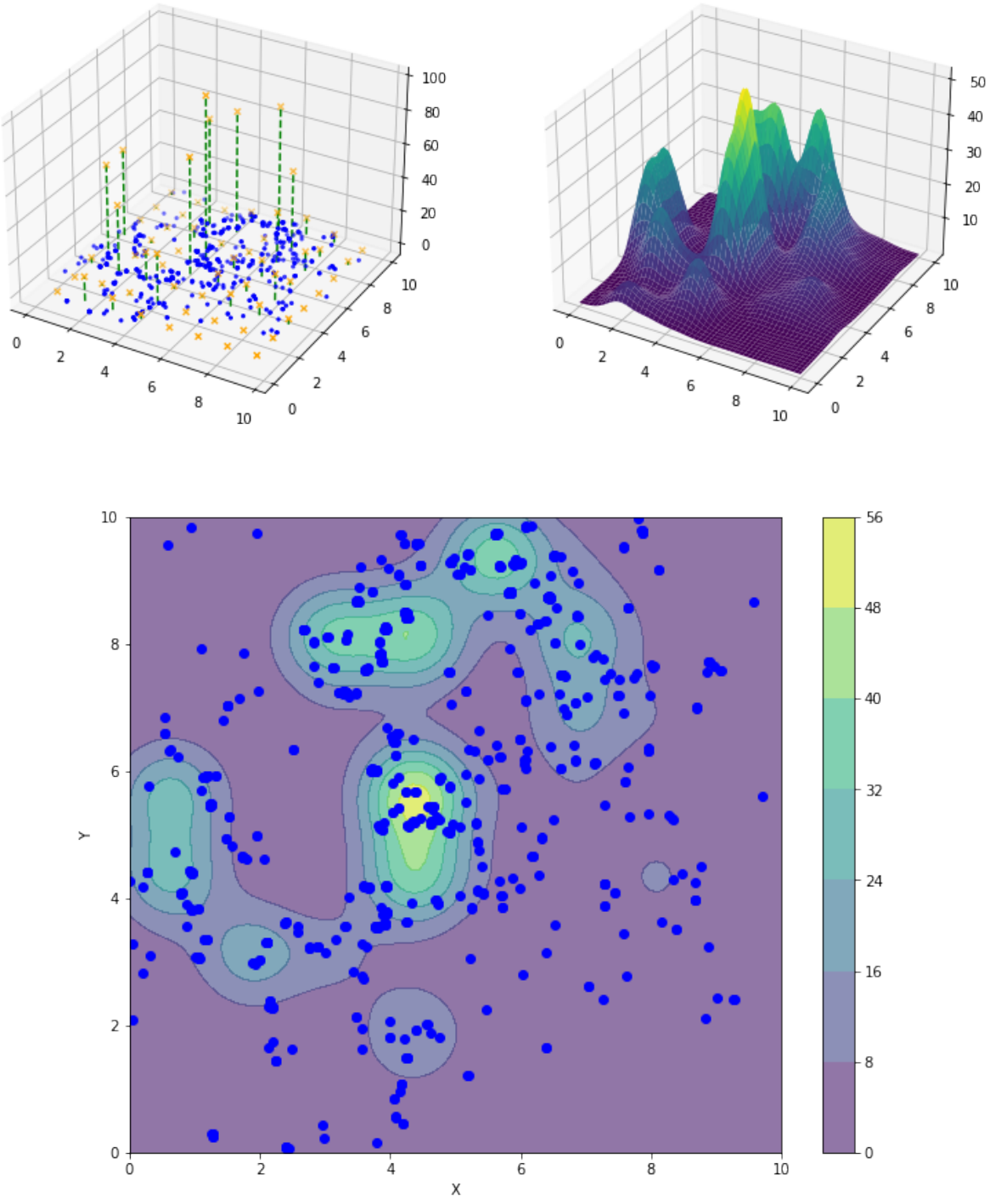

Figure 4.5: Plot of the locations (blue points) of the fires over the contour plot of $\mathbb{E}(\Lambda|\text{Data})$ on E.

4.2. The time varying case

Now, we want to analyse the same dataset on a different time-frame and take in consideration 33 months, from January 2018 to September 2020 (the coordinates for each month are available here: **Project's Github**). Plotting the number of fires by month (see Figure 4.6) we can hypothesise that latent process changes over time and therefore model α as:

$$\alpha_t = e^{\eta_0 + \eta_1 I(t) + \eta_2 A(t)}$$

where:

- $I(t) = 1$ if the t^{th} month belongs to the dry season (which spans from August to November) and $I(t) = 0$ otherwise.
- $A(t) = 0, 1, 2$ if the t^{th} month belongs to the first, second, third year (2018, 2019, 2020); however, we consider December part of the previous year as it is the first month of the wet season.

Note that this way of modelling α_t reflects the idea that in the same season of the same year the latent process has the same distribution.

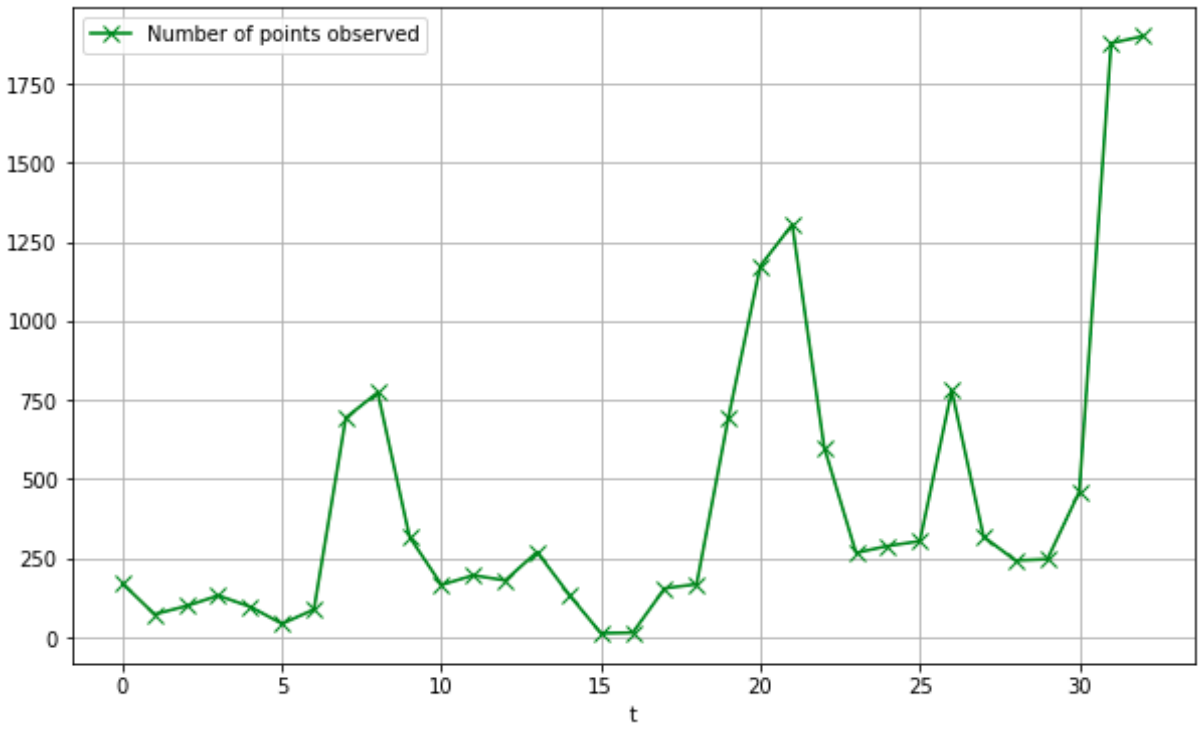

Figure 4.6: Number of forest fires (on the y-axis) observed in the t^{th} month.

Applying the same transformations as in the stationary case and assigning $\psi = 0.8$ we can now run the Metropolis within Gibbs sampler proposed in Algorithm 3.1 with the modification proposed in section 3.5 and sample from the posteriors of $\eta_0, \eta_1, \eta_2, \beta$. We run it for 10000 iterations to ensure reaching stationarity.

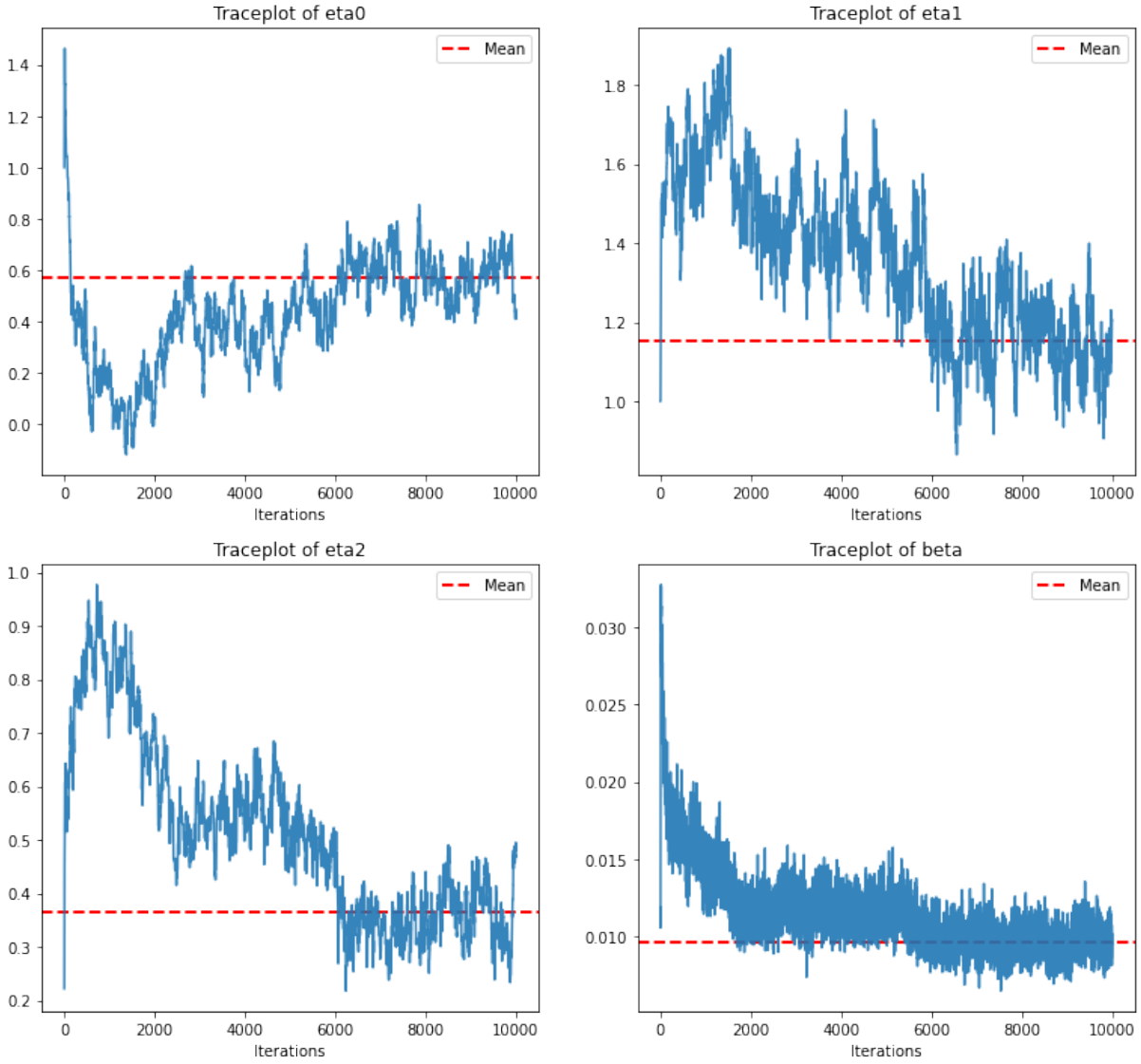

Figure 4.7: Traceplots of $\eta_0, \eta_1, \eta_2, \beta$ and relative means (with a Burn-in of 8000 iterations).

We can approximate an estimator for the number of points observed in each month with the empirical mean of $\Lambda(E)$, i.e $\hat{N}_t = \mathbb{E}(\Lambda_t(E)|\text{Data}) \simeq \frac{1}{L}\sum_{it=1}^{L}\overline{\Lambda_t^{(it)}(E)}$ (see upper graph in Figure 4.7).

A similar way to do so is: $\tilde{N}_t = \mathbb{E}(\Lambda_t(E)|\text{Data}) \simeq \frac{\alpha_t}{\beta} \simeq \frac{1}{L}\sum_{it=1}^{L}\frac{\alpha_t^{(it)}}{\beta^{(it)}}$ which is an estimator of the number of points within a fixed year and season (see lower graph in Figure 4.7).

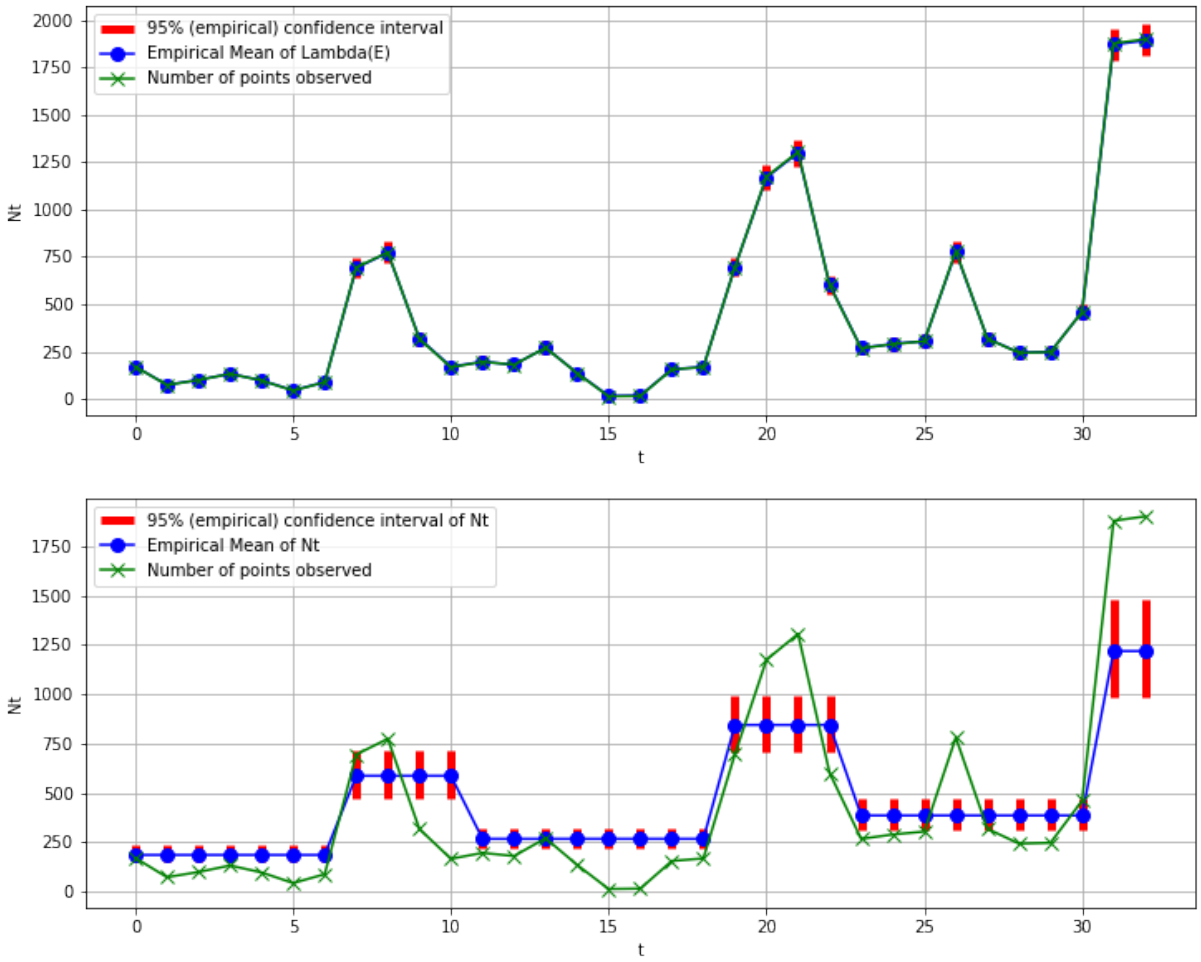

Figure 4.8: Each graph plots an estimator of number of fires for each month (in blue) against the real number of fires during the same month (in green). In the first graph the estimator is $\hat{N}_t$ while in the second is $\tilde{N}_t$ as explained before the figure.

5 | Further work

We believe the following areas present promising directions for future research:

1. **Convergence Analysis**: Prove a stronger result for the convergence of the Poisson-Gamma Random measure when the underlying discrete Gamma process converges to a non discrete one.

2. **Algorithmic Refinement**: Focus on optimizing the existing code and implementing an adaptive Metropolis-Hastings algorithm to enhance efficiency and accuracy.

3. **Exploration of Diverse Datasets**: Extend the study to analyze various datasets with the aim of improving the predictive performance (with a focus on extrapolation) of the algorithm. Examples from spatial statistics, such as taxi pickups in urban areas or instances of violent crime in neighborhoods, offer promising avenues for investigation. Utilize available online datasets to conduct comprehensive analyses and validate the algorithm's effectiveness across different scenarios.

Bibliography

[1] G. Last, M. Penrose. *Lectures on the Poisson process*. Cambridge University Press, 2018.

[2] D. J. Daley, D. Vere-Jones. *An introduction to the theory of point processes, volume I. Elementary Theory and Methods*. Springer, 2003.

[3] D. J. Daley D. Vere-Jones. *An introduction to the theory of point processes, volume II. General Theory and Structure*. Springer, 2008.

[4] R. L. Wolpert, K. Ickstadt. *Poisson/Gamma random field models for spatial statistics*. Biometrika 85(2):251–267, 1998.

[5] M. Abramowitz, I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Applied Mathematics Series, 55. Washington, DC: National Bureau of Standards, 1964.

[6] J. M. Marin, C. P. Robert. *A Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer, 2007.

[7] F.Bassetti, R. Casarin, M. Iacopini. *A Spatiotemporal Gamma Shot Noise Cox Process*. ArXiv eprint arXiv:2308.08481, 2023.

[8] O. Kallenberg. *Foundations of Modern Probability*. Springer, 2021.

Ringraziamenti

Per prima cosa ringrazio il mio relatore, il professor Federico Bassetti, che, con tanta pazienza e disponibilità, mi ha aiutato nella comprensione degli argomenti di questa tesi e nella stesura della stessa. Chiaramente il suo supporto è stato fondamentale per la conclusione di questo lavoro.

Mi è tuttavia doveroso dedicare questo spazio della mia tesi a tutte le persone che mi hanno supportato nel mio percorso di crescita accademica e soprattutto personale. Una dedica speciale va ai miei genitori, Giacinto e Simona, che hanno sempre creduto in me e continuano a farlo e che sponsorizzano questa fantastica avventura. Nonostante abbia cercato di mostrarmi sempre autonomo e indipendente, il raggiungimento di questo traguardo non sarebbe stato possibile senza di voi. Ringrazio anche mio fratello Riccardo, compagno di molte avventure, a cui auguro il meglio per la sua imminente carriera universitaria qui al Politecnico, anche se non ho dubbi che sarà di successo.

Un ringraziamento di cuore va alla mia ragazza Chiara, che, con tanta pazienza, ha condiviso con me questi ultimi due (bellissimi) anni. La sua costante presenza, il suo affetto e la sua comprensione hanno reso ogni momento speciale e mi hanno dato la forza di superare le sfide più difficili. La sua dolcezza e il suo sorriso hanno per me un valore inestimabile.

Grazie anche ai miei migliori amici, Emanuele, Marco e Simone (in ordine alfabetico) con i quali sono amico da una vita e con cui ho condiviso anche le gioie e le difficoltà di questo percorso.

Concludo con un ringraziamento speciale per i miei nonni: Angelo e Virginia. Nonostante la sopraggiunta vecchiaia, siete sempre stati e sarete per me un punto di riferimento. Mi avete insegnato il valore della dedizione, della pazienza e dell'amore incondizionato, e per questo vi sarò eternamente grato. Il momento è difficile ma so che con l'aiuto dei vostri cari supererete anche queste asperità.