Departamento de Matemáticas, Facultad de Ciencias
Universidad Autónoma de Madrid

The Poisson Process and Applications to Spatial Data Analysis

Degree in Mathematics

Author: Gabriel López-Asiaín López
Tutor: José Ramón Berrendero Díaz

Course 2024-2025

Abstract

The Poisson process is a fundamental stochastic model used to describe the random occurrence of events in time or space, assuming no interaction between them. Its general formulation allows it to be defined over any measurable space, and many of its key properties are independent of the underlying domain. The particular case on the real line enables the analysis of temporal event structures and finds important applications in contexts such as queueing theory. In spatial settings, the Poisson process serves as an effective statistical tool for modeling and analyzing point patterns, enabling the identification of underlying structures from observed data.

Contents

1 Preliminaries **1**
- 1.1 Introduction . . . 1
- 1.2 The Poisson Distribution . . . 2

2 Poisson Processes in General Spaces **5**
- 2.1 Definition and Basic Properties . . . 5
- 2.2 Superposition of Poisson Processes . . . 7
- 2.3 Transformation of Poisson Processes . . . 9
- 2.4 Existence Theorem . . . 10
- 2.5 Sums over Points of a Poisson Process . . . 13

3 Poisson Processes on the Line **17**
- 3.1 Definition and Basic Properties . . . 17
- 3.2 Distribution of Inter-arrival Times and Waiting Times . . . 20
- 3.3 Conditional Distribution of Arrival Times . . . 21
- 3.4 Relation to Queueing Theory . . . 24
- 3.5 Non-homogeneous Poisson Process . . . 26

4 Poisson Models for Spatial Data Analysis **27**
- 4.1 Non-parametric Models of the Intensity Function of a Poisson Process 27
- 4.2 Parametric Models of Process Intensity . . . 29
- 4.3 Model Selection via AIC . . . 33

5 Conclusions **35**

CHAPTER 1
Preliminaries

1.1. Introduction

The objective of this work is to study the Poisson process, one of the most important and versatile stochastic models in the fields of statistics and probability. It is a process that appears naturally in the modeling of phenomena where events occur randomly in time or space, without apparent interaction between them. Typical examples include the arrival of customers at a queueing system, the spontaneous emission of radioactive particles, the location of trees in a natural environment, or the appearance of defects on a manufacturing surface.

In this chapter, in addition to introducing the topic, we summarize some of the most important properties of the Poisson distribution that will be necessary for the work.

In Chapter 2, we will analyze its general formulation from a theoretical perspective, emphasizing those properties that do not depend on the specific space on which the process is defined. This generality allows the same conceptual framework to be applied in very different contexts, from temporal processes to point distributions in more abstract spaces. For this part, we will mainly follow the approach presented in the book by Kingman [5].

Next, in Chapter 3, we will focus on the particular case of the Poisson process on the real line, which presents specific properties derived from the ordered structure of time. This case is particularly relevant in classical applications such as queueing theory or the analysis of events in time. Many of the results presented are based on the treatment found in the book by Ross [8].

Finally, in Chapter 4, we will address its direct application to spatial data modeling. We will study how the Poisson process can be used to describe the distribution of events in space and how, through different statistical techniques, it is possible to extract useful information from real data. This part will be illustrated using concrete examples. The main theoretical and computational reference in this section is the book by Baddeley et al. [3], along with the `R` package `spatstat`, developed in the same context.

1.2. The Poisson Distribution

The process takes its name from the Poisson distribution, with which it is closely related. A random variable X has a Poisson distribution with parameter μ if its possible values are positive integers and if

$$P\{X = n\} = \pi_n(\mu) = \frac{\mu^n e^{-\mu}}{n!}$$

for $n \geq 0$. We will say that $X \sim \mathcal{P}(\mu)$. Here μ can take any value $\mu > 0$, and the identity

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} n\pi_n(\mu) = \sum_{n=1}^{\infty} n\frac{\mu^n e^{-\mu}}{n!} = e^{-\mu}\mu \sum_{m=0}^{\infty} \frac{\mu^m}{m!} = \mu$$

shows that μ is the mean of the distribution $\mathcal{P}(\mu)$.

We can extend the definition of $\mathcal{P}(\mu)$ to include the limit cases 0 and ∞. Thus, $\mathcal{P}(0)$ will represent the distribution concentrated at 0, $P\{X = 0\} = 1$, and $\mathcal{P}(\infty)$ the distribution concentrated at $+\infty$, $P\{X = +\infty\} = 1$.

If z is any real or complex number with $|z| \leq 1$, the random variable z^X is bounded, so

$$\mathbb{E}[z^X] = \sum_{n=0}^{\infty} \pi_n(\mu) z^n = e^{-\mu} \sum_{n=0}^{\infty} \frac{(\mu z)^n}{n!} = e^{\mu(z-1)},$$

and this holds for $0 \leq \mu < \infty$. Identities for the moments of X can be obtained by differentiating and taking $z = 1$:

$$\begin{aligned} \mathbb{E}[X] &= \mu, \\ \mathbb{E}[X(X-1)] &= \mu^2, \\ \mathbb{E}[X(X-1)(X-2)] &= \mu^3, \quad \ldots \end{aligned}$$

From which we can deduce, for example, that $\mathbb{E}[X^2] = \mu + \mu^2$ and that $\mathrm{Var}[X] = \mu$.

One of the most important characteristics of the Poisson distribution is its additivity. If X and Y are independent random variables, with Poisson distributions $\mathcal{P}(\lambda)$ and $\mathcal{P}(\mu)$ respectively, then for $r, s \geq 0$, we have

$$P\{X = r, Y = s\} = P\{X = r\}P\{Y = s\} = \frac{\lambda^r e^{-\lambda}}{r!}\frac{\mu^s e^{-\mu}}{s!}. \tag{1.1}$$

We can find the distribution of the random variable $X + Y$ by summing (1.1) over values of r and s with $r + s = n$:

$$P\{X + Y = n\} = \sum_{r=0}^{n} P\{X = r, Y = n - r\}$$

$$= \sum_{r=0}^{n} \frac{\lambda^r e^{-\lambda}}{r!}\frac{\mu^{n-r} e^{-\mu}}{(n-r)!} = \frac{e^{-(\lambda+\mu)}}{n!} \sum_{r=0}^{n} \binom{n}{r} \lambda^r \mu^{n-r} = \frac{(\lambda+\mu)^n e^{-(\lambda+\mu)}}{n!}.$$

Therefore, $X + Y$ has distribution $\mathcal{P}(\lambda + \mu)$. This extends by induction to the sum of any finite number of independent random variables. However, we will need an even stronger result, which covers infinite sums and gives conditions for convergence.

Theorem 1.1 (Countable Additivity Theorem)**.** *Let X_j $(j = 1, 2, \dots)$ be independent random variables, where each X_j follows a distribution $\mathcal{P}(\mu_j)$ for each j. If $\sigma = \sum_{j=1}^{\infty} \mu_j < \infty$, then $S = \sum_{j=1}^{\infty} X_j$ converges with probability 1, and S has distribution $\mathcal{P}(\sigma)$. If, on the contrary, $\sigma = \infty$, then S diverges with probability 1.*

Proof. By induction on n, we define $S_n = \sum_{j=1}^{n} X_j$ which follows a Poisson distribution $\mathcal{P}(\sigma_n)$, where $\sigma_n = \sum_{j=1}^{n} \mu_j$. Therefore, for any r,

$$P\{S_n \leq r\} = \sum_{k=0}^{r} \pi_k(\sigma_n).$$

The events $\{S_n \leq r\}$ are decreasing as n increases for a fixed r, and their intersection is $\{S \leq r\}$. Thus, we have that

$$P\{S \leq r\} = \lim_{n\to\infty} P\{S_n \leq r\} = \lim_{n\to\infty} \sum_{k=0}^{r} \pi_k(\sigma_n).$$

If σ_n converges to a finite limit σ, the continuity of π_k gives us

$$P\{S \leq r\} = \sum_{k=0}^{r} \pi_k(\sigma),$$

which implies that $P\{S = r\} = \pi_r(\sigma)$. Therefore, S follows the distribution $\mathcal{P}(\sigma)$. However, if $\sigma_n \to \infty$, then

$$\sum_{k=0}^{r} \pi_k(\sigma_n) = e^{-\sigma_n} \sum_{k=0}^{r} \frac{\sigma_n^k}{k!} \to 0,$$

which implies that $P\{S > r\} = 1$. Given that this is valid for any r, we conclude that S diverges with probability 1, and the proof is complete. □

CHAPTER 2
Poisson Processes in General Spaces

2.1. Definition and Basic Properties

In general, the state space S in which a Poisson process Π is defined is usually the *Euclidean space* of dimension d. However, much of the theory does not depend on specific properties of this space, but solely on the existence of a family of "suitable" subsets that can be used as test sets $A \subset S$ to count the random points. That is, we need a collection of sets for which the counting function, which measures the number of points of Π in A,

$$N(A) = \#\{\Pi \cap A\} \tag{2.1}$$

is a well-defined random variable. Here we understand $\#A$ as the cardinality of the set A. The natural way to do this is to assume that S is a measure space. That is, we assume that there exist certain subsets of S called *measurable*, such that they form a σ-algebra in the sense that the empty set is measurable, the complement of a measurable set is measurable, and the union of a countable number of measurable sets is also measurable.

We also need there to be sufficient measurable sets to ensure that individual points can be distinguished. This can be achieved by assuming that the diagonal $D = \{(x, y); x = y\}$ is a measurable subset of the space $S \times S$. This has the immediate implication that any single-element subset is measurable, since, if D is measurable, any set of the form $\{(x, x\})$ can be expressed as $D \cap (\{x\} \times S)$, this being the intersection of two measurable sets in the product space $S \times S$ and, therefore, by Fubini's theorem (see [6]), its projection onto the first coordinate is measurable. With these assumptions, we can give a general definition of the Poisson process.

Definition 2.1. A Poisson process on a measurable space S is a random subset Π of S that satisfies:

(i) for any collection of disjoint measurable sets $A_1, A_2, \ldots, A_n$ of S, the random variables $N(A_1), N(A_2), \ldots, N(A_n)$ are independent, and

(ii) $N(A)$ follows a Poisson distribution with parameter $\mu = \mu(A)$ with $0 \leq \mu \leq \infty$.

Given that $N(A) \sim \mathcal{P}(\mu(A))$ we have that $\mathbb{E}[N(A)] = \mu(A)$. Then, if $A_1, A_2, \ldots$ are disjoint and their union is A,

$$N(A) = \sum_{n=1}^{\infty} N(A_n),$$

and taking expectations, the monotone convergence theorem gives us

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n).$$

Therefore, μ is a measure on S, which we will call the *induced measure* or *mean measure* of the Poisson process Π.

It is important to highlight that not every measure can act as the induced measure of a Poisson process. Suppose that the measure μ on S has an atom at x, so that μ assigns non-zero measure to the singleton set $\{x\}$, that is, $\mu(\{x\}) > 0$. Then, applying (ii) with $A = \{x\}$ yields that a Poisson process with mean measure μ would verify that

$$P\{N(\{x\}) \geq 2\} = 1 - e^{-m} - me^{-m} > 0.$$

This clearly contradicts (2.1), which shows that μ cannot be a measure induced by a Poisson process. Therefore, an induced measure must be non-atomic in the sense that

$$\mu(\{x\}) = 0, \quad \text{for all } x \in S. \tag{2.2}$$

Naturally, this will be a necessary condition to construct a Poisson process with said induced measure; later we will see that this condition is almost sufficient for such a process to exist.

When $S = \mathbb{R}^d$, the induced measure in the most interesting cases is usually expressed in terms of a *rate* or *intensity*. This is a positive and measurable function λ on S, from which μ is defined by integrating λ with respect to the Lebesgue measure in dimension d:

$$\mu(A) = \int_A \lambda(x)\, dx, \quad dx = dx_1 dx_2 \ldots dx_d. \tag{2.3}$$

We recall that, given a measure space $(X, \mathcal{A}, \mu)$, μ is said to be σ-*finite* if $X = \bigcup_{n=1}^{\infty} A_n$, with $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ and $\mu(A_n) < \infty$ for all $n \geq 1$. On the other hand, a

measure ν is absolutely continuous with respect to μ ($\nu \ll \mu$) if whenever $\mu(A) = 0$, then $\nu(A) = 0$.

Note that for (2.3) to be possible, by the Radon-Nikodym theorem (see for example [6]), it suffices for the measure μ to be σ-finite and absolutely continuous with respect to the Lebesgue measure μ_L.

The term *rate* is used more frequently when $d = 1$ and S represents a timeline, while *intensity* is used more often when S has a spatial interpretation. If λ is continuous at x, then the previous equation implies that for small sets A around x,

$$\mu(A) \sim \lambda(x)|A|,$$

where $|A|$ denotes the measure of the set (length if $d = 1$, area if $d = 2$, volume if $d = 3$, etc.). Thus, $\lambda(x)|A|$ approximately represents the probability that a point of Π belongs to the small set A, and it is greater in regions where λ is large than in those where λ is small.

In the particular case where λ is constant, so that $\mu(A) = \lambda|A|$, the Poisson process is said to be *uniform* or *homogeneous*. A random set of this type has stochastic properties that remain invariant under translations or rotations of S, and it is the only Poisson process (finite on bounded sets) that possesses this property.

2.2. Superposition of Poisson Processes

The Poisson process has a series of special properties that make its use, and the calculation of associated probabilities, often surprisingly simple. Among these properties, the *superposition theorem* stands out. Its proof requires, as a preliminary step, establishing the following lemma.

Lemma 2.2. *Let Π_1 and Π_2 be independent Poisson processes on S, and let A be a measurable set with $\mu_1(A)$ and $\mu_2(A)$ finite. Then Π_1 and Π_2 are disjoint with probability 1 on A:*

$$P\{\Pi_1 \cap \Pi_2 \cap A = \emptyset\} = 1. \tag{2.4}$$

Proof. Let A^f be the set of all finite subsets Λ of A. We construct A^f as a measurable space by considering the smallest σ-algebra that makes the function

$$\Lambda \mapsto \#(\Lambda \cap B) \tag{2.5}$$

measurable for all measurable sets $B \subseteq A$. In this way, $\Pi_1 \cap A$ is a random element of A^f, with a probability distribution that we will denote as P_1. Similarly, $\Pi_2 \cap A$ has distribution P_2 on A^f. Since Π_1 and Π_2 are independent, the joint distribution of $\Pi_1 \cap A$ and $\Pi_2 \cap A$ is the product of measures $P_1 \times P_2$ on $A^f \times A^f$.

We define the mapping $\eta : A^f \times A^f \to (A \times A)^f$ as

$$\eta(\Lambda_1, \Lambda_2) = \Lambda_1 \times \Lambda_2.$$

We first check that η is measurable, for which it suffices to prove that $\eta^{-1}(\{\Lambda : \#(\Lambda \cap C) = n\})$ is measurable in $A^f \times A^f$ for any measurable set $C \subseteq A \times A$. This is verified for $C = B_1 \times B_2$, since the statement

$$\#(\eta(\Lambda_1, \Lambda_2) \cap (B_1 \times B_2)) = n$$

is equivalent to stating that there exist n_1, n_2 such that $n_1 + n_2 = n$ and

$$\#(\Lambda_1 \cap B_1) = n_1, \quad \#(\Lambda_2 \cap B_2) = n_2.$$

Furthermore, the class of all these sets C is closed under countable unions and monotone limits (see Theorem 1.5 of Kingman and Taylor [6]), and therefore contains all sets C that are measurable in the product space $A \times A$. In particular, since the diagonal D is measurable in $S \times S$, its restriction

$$D_A = D \cap (A \times A)$$

is measurable in $A \times A$, and therefore

$$J = \eta^{-1}\{\Lambda; \#(\Lambda \cap D_A) = 0\}$$

is measurable in $A^f \times A^f$. The statement (2.4) is then equivalent to saying that $(P_1 \times P_2)(J) = 1$ which, by Fubini's theorem, is true if, for all Λ_1 outside a null set of P_1,

$$P_2\{\Lambda_2 : (\Lambda_1, \Lambda_2) \in J\} = 1.$$

This is equivalent to stating that

$$P_2(N_2(\Lambda_1) = 0) = 1,$$

with Λ_1 being finite, so $\mu_2(\Lambda_1) = 0$ for all Λ_1, thus ending the proof. □

With this, we can now prove the following theorem.

Theorem 2.3 (Superposition Theorem)**.** *Let $\Pi_1, \Pi_2, \ldots$ be a countable collection of independent Poisson processes on S, such that Π_n has mean measure μ_n for each n. Then their superposition $\Pi = \bigcup_{n=1}^{\infty} \Pi_n$ is a Poisson process with induced measure $\mu = \sum_{n=1}^{\infty} \mu_n$.*

Proof. Let $N_n(A)$ be the number of points of Π_n in the measurable set A. If $\mu_n(A) < \infty$ for all n, the previous lemma shows that the random sets Π_n are disjoint on A, so that the number of points of Π in A is

$$N(A) = \sum_{n=1}^{\infty} N_n(A). \tag{2.6}$$

By the countable additivity theorem 1.1, $N(A)$ has distribution $\mathcal{P}(\mu)$, where $\mu = \mu(A)$. On the other hand, if $\mu_n(A) = \infty$ for some n, then $N_n(A) = N(A) = \infty$, and (2.6) is trivially satisfied.

To prove the theorem, we only need to prove that $N(A_1), N(A_2), \ldots, N(A_k)$ are independent if the sets $A_1, A_2, \ldots, A_k$ are disjoint. This is evident because the random variables

$$N_n(A_j), \quad n = 1, 2, \ldots, \quad j = 1, 2, \ldots, k,$$

are independent, and $N(A_j)$ is defined in terms of a subset of these variables, which is disjoint for those A_j with a distinct j. □

2.3. Transformation of Poisson Processes

The following fundamental property of the Poisson process allows us to determine that, under certain conditions, if a transformation is applied to the state space where the random set of points is located, the points form, again, a Poisson process. We will see that it is almost sufficient for the transformation not to map distinct points to the same image.

Theorem 2.4 (Transformation of a Poisson Process)**.** *Let Π be a Poisson process with σ-finite mean measure μ on the state space S, and let $f : S \to T$ be a measurable function such that the induced measure $\mu^* := \mu^*(B) = \mu\big(f^{-1}(B)\big)$ has no atoms. Then $f(\Pi)$ is a Poisson process on T with μ^* as its induced measure.*

Proof. Suppose first that $\mu(S) < \infty$. Let A be a measurable set and A^c its complement. We define Π_1 and Π_2 as the restriction of Π to A and A^c respectively, so that Π_1 and Π_2 are independent Poisson processes. The argument used in the proof of Lemma 2.2, now applied to the random set

$$f(\Pi_1) \times f(\Pi_2) \subseteq T \times T$$

shows that $f(\Pi_1)$ and $f(\Pi_2)$ are disjoint with probability 1.

Denote by M the measure on $S \times S$ defined as follows: for any measurable set $C \subseteq S \times S$, $M(C)$ is the expected number of pairs $(x, y) \in C$ such that

$$x \in \Pi, \quad y \in \Pi, \quad f(x) = f(y).$$

Since $f(\Pi_1)$ and $f(\Pi_2)$ are disjoint, this implies that $M(A \times A^c) = 0$, for any measurable set A.

For any pair of sets $A, B \subseteq S$, $A \times B$ can be written as the union of $(A \cap B) \times (A \cap B)$ and sets that are either in $A \times A^c$ or in $B^c \times B$, so that

$$M(A \times B) = m(A \cap B),$$

where $m(A) = M(A \times A)$.

Substituting $B = S$ in the previous equation we obtain that $m(A) = m(A \cap S) = M(A \times S)$ so that m is a measure. Let M_1 be the measure induced by m on the diagonal $D \subseteq S \times S$ via the mapping $x \mapsto (x, x)$. Then,

$$M_1(A \times B) = m(A \cap B) \quad \text{and} \quad M_1(A \times B) = M(A \times B)$$

for all measurable sets A, B.

By the uniqueness of measure extension, it follows that $M_1 = M$, and in particular that M is concentrated on the diagonal D. Thus, with probability 1, there are no $x, y \in \Pi$ with $f(x) = f(y)$, which implies that

$$\#\{x \in \Pi;\, f(x) \in B\} = N(f^{-1}(B)),$$

thus proving the result for $\mu(S) < \infty$.

Now we relax the assumption that $\mu(S) < \infty$, and assume instead that μ is σ-finite, that is, there exist disjoint sets S_n such that

$$S = \bigcup_{n=1}^{\infty} S_n, \quad \mu(S_n) < \infty.$$

If we restrict Π to each S_n, so that Π_n is the restriction of Π to S_n, then $\Pi_1, \Pi_2, \ldots$ are independent Poisson processes, each with its corresponding induced measure μ_n. Thus, $f(\Pi_n)$ is a Poisson process with induced measure $\mu_n^*(B)) = \mu_n(f^{-1}(B))$. Therefore, the disjoint union

$$f(\Pi) = f(\bigcup_{n=1}^{\infty} \Pi_n) = \bigcup_{n=1}^{\infty} f(\Pi_n)$$

is a Poisson process with measure

$$\mu^*(B) = \sum_{i=1}^{\infty} \mu_n^*(B) = \mu(f^{-1}(B)).$$

This concludes the proof. □

2.4. Existence Theorem

Now, we are interested in knowing under what conditions, given a measure μ, we can guarantee the existence of a Poisson process with said induced measure. That is, if we start from a given measure on the space S, can we construct a random set of points whose number of elements in each measurable set A follows a Poisson distribution with mean $\mu(A)$?

First, we will see what happens when conditioning a Poisson process with finite induced measure μ to contain exactly n points. Next, we will invert that argument to prove the existence theorem.

Let Π be a Poisson process on S with mean measure μ, and suppose that $\mu(S) < \infty$. Then Π is almost surely a finite subset of S, with $N(S)$ following the Poisson distribution $\mathcal{P}(\mu(S))$. We are interested in knowing what happens if we condition Π on the value of $N(S)$.

We denote by P_n the conditional probability measure:

$$P_n\{\cdot\} = P\{\cdot | N(S) = n\}.$$

Let $A_1, A_2, \ldots, A_k$ be disjoint subsets of S, A_0 the complement of $A_1 \cup A_2 \cup \cdots \cup A_k$ and $n_0 = n - \sum_i n_i$. Then

$$P_n\{N(A_1) = n_1, N(A_2) = n_2, \ldots, N(A_k) = n_k\}$$

$$= \frac{P\{N(A_0) = n_0, N(A_1) = n_1, \ldots, N(A_k) = n_k\}}{P\{N(S) = n\}}$$

where

$$P\{N(A_0) = n_0, N(A_1) = n_1, \ldots, N(A_k) = n_k\} = \prod_{j=0}^{k} \frac{e^{-\mu(A_j)}\mu(A_j)^{n_j}}{n_j!}.$$

Substituting and simplifying, we obtain

$$P_n\{N(A_1) = n_1, \ldots, N(A_k) = n_k\} = \frac{n!}{n_0!n_1!\ldots n_k!}\left(\frac{\mu(A_0)}{\mu(S)}\right)^{n_0}\prod_{i=1}^{k}\left(\frac{\mu(A_i)}{\mu(S)}\right)^{n_i}.$$

Let p be the probability measure given by

$$p(A) = \frac{\mu(A)}{\mu(S)}. \tag{2.7}$$

A finite random set $\Pi \subseteq S$ for which $\#\Pi = n$, which verifies

$$P\{N(A_1) = n_1, \ldots, N(A_k) = n_k\} = \frac{n!}{n_0!n_1!\ldots n_k!}p(A_0)^{n_0}\prod_{i=1}^{k}p(A_i)^{n_i} \tag{2.8}$$

for any collection $A_1, A_2, \ldots, A_k$ of disjoint sets, is called a *Bernoulli process*. Its mean measure is $\mathbb{E}[N(A)] = np(A)$, so p is a probability measure on S. With this terminology, we can say that conditioning $N(S) = n$ converts the Poisson process with finite mean measure μ into the Bernoulli process with parameters n and $p(\cdot)$.

There is a simpler way to construct a Bernoulli process. We consider a collection $X_1, X_2, \ldots, X_n$ of independent random variables, distributed in the space S according to the probability measure p. It can be shown that if p has no atoms, the X_r are

distinct with probability one, so the set $\{X_1, X_2, \dots, X_n\}$ is a random set with n elements. An immediate multinomial calculation shows that the counts

$$N(A) = \#\{r; X_r \in A\}$$

satisfy (2.8).

Thus, given $N(S)$, the points of a finite Poisson process are exactly like $N(S)$ independent random variables, with common distribution given by (2.7).

Now, by inverting this argument, we can demonstrate the existence of a Poisson process given a measure that meets certain regularity conditions. We will see that we only need this measure to be non-atomic and to verify a condition less restrictive than σ-finiteness.

Theorem 2.5 (Existence Theorem)**.** *Let μ be a non-atomic measure on S that can be expressed as*

$$\mu = \sum_{n=1}^{\infty} \mu_n, \quad \mu_n(S) < \infty. \tag{2.9}$$

Then, there exists a Poisson process on S that has μ as its induced measure.

Proof. Without loss of generality, assume that $\mu_n(S) > 0$ for all n. In a suitable probability space, we construct independent random variables

$$N_n, \quad X_{nr} \quad (n, r = 1, 2, 3, \dots)$$

such that the distribution of N_n is $\mathcal{P}(\mu_n(S))$ and the distribution of X_{nr} is

$$p_n(A) = \frac{\mu_n(A)}{\mu_n(S)}.$$

We define $\Pi_n = \{X_{n1}, X_{n2}, \dots, X_{nN_n}\}$ and

$$\Pi = \bigcup_{n=1}^{\infty} \Pi_n. \tag{2.10}$$

If we write

$$N_n(A) = \#\{\Pi_n \cap A\}$$

we have that, for disjoint subsets $A_1, A_2, \dots, A_k$, writing A_0 as the complement of their union,

$$P\{N_n(A_1) = m_1, N_n(A_2) = m_2, \dots, N_n(A_k) = m_k | N_n = m\}$$

$$= \frac{m!}{m_0! m_1! \dots m_k!} p_n(A_0)^{m_0} p_n(A_1)^{m_1} \dots p_n(A_k)^{m_k},$$

where $m_0 = m - m_1 - m_2 - \dots - m_k$. Therefore,

$$P\{N_n(A_1) = m_1, N_n(A_2) = m_2, \dots, N_n(A_k) = m_k\}$$

$$= \sum_{m=\sum m_j}^{\infty} \frac{e^{-\mu_n(S)}\mu_n(S)^m}{m!} \frac{m!}{m_0!m_1!\dots m_k!} p_n(A_0)^{m_0} p_n(A_1)^{m_1} \dots p_n(A_k)^{m_k}$$

$$= \sum_{m=\sum m_j}^{\infty} \pi_{m-\sum m_j}(\mu_n(A_0)) \prod_{j=0}^{k} \pi_{m_j}(\mu_n(A_j))$$

$$= \prod_{j=0}^{k} \pi_{m_j}(\mu_n(A_j)).$$

Thus, the random variables $N_n(A_j)$ are independent and follow the distribution $\mathcal{P}(\mu_n(A_j))$, which implies that the Π_n are independent Poisson processes with mean measures μ_n. The superposition theorem 2.3 establishes that the union in (2.10) defines a Poisson process with mean measure μ, thus completing the proof. □

We observe that the σ-finiteness of the measure μ directly implies the existence of a Poisson process on S with that measure. Indeed, if there exist sets $S_1, S_2, \dots$ such that $\mu(S_n) < \infty$ for all n, we can define μ_n as the measure μ restricted to each set S_n, and condition (2.9) is trivially satisfied. Therefore, σ-finiteness constitutes a sufficient condition, although not strictly necessary, to guarantee the existence of the process.

2.5. Sums over Points of a Poisson Process

The objective of this section is to study sums of the form

$$\Sigma = \sum_{x \in \Pi} f(X)$$

where f is a real function defined on the state space S of the process Π. There are a series of important results that offer conditions for the absolute convergence of such sums as well as expressions for their expectation and variance.

We will start working on a function $f : S \to \mathbb{R}$ that takes only a finite number of distinct values $f_1, f_2, \dots f_k$ and is zero outside a set of finite measure with respect to the mean measure μ of the Poisson process Π. These functions are called *simple*. Subsequently, we will extend these results to other functions.

Consider the sets $A_j = \{x \in S : f(x) = f_j\}$, which are measurable with measure $m_j = \mu(A_j) < \infty$. Furthermore, the distinct A_j are disjoint and therefore the random variables, $N_j = N(A_j)$ are independent with respective distributions $\mathcal{P}(m_j)$, and we have that

$$\Sigma = \sum_{x \in \Pi} f(X) = \sum_{j=1}^{k} f_j N_j.$$

This determines the distribution of Σ, which is most conveniently described through its characteristic function. For any real or complex value θ, one has

$$\mathbb{E}[e^{\theta\Sigma}] = \prod_{j=1}^{k} \mathbb{E}[e^{\theta f_j N_j}] = \prod_{j=1}^{k} \exp\left(\mu_j(e^{\theta f_j} - 1)\right)$$
$$= \exp\left(\sum_{j=1}^{k} \int_{A_j} \left(e^{\theta f(x)} - 1\right) \mu(dx)\right) = \exp\left(\int_S (e^{\theta f(x)} - 1)\mu(dx)\right),$$

since $f = 0$ outside the union of the sets A_j. This gives rise to the master equation

$$\mathbb{E}[e^{\theta\Sigma}] = \exp\left(\int_S (e^{\theta f(x)} - 1)\mu(dx)\right), \tag{2.11}$$

from which everything follows. If f takes positive and negative values, it is convenient to take θ purely imaginary, so that (2.11) becomes

$$\mathbb{E}[e^{it\Sigma}] = \exp\left(\int_S (e^{itf(x)} - 1)\mu(dx)\right) \tag{2.12}$$

for all real t. The left side of (2.12) corresponds to the characteristic function of Σ, which entirely determines its distribution. If f takes only positive values, we have that (2.11) becomes

$$\mathbb{E}[e^{-u\Sigma}] = \exp\left(-\int_S (1 - e^{-uf(x)})\,\mu(dx)\right) \tag{2.13}$$

for all $u \geq 0$. In this form, the equation is valid without restriction: if the integral diverges, then $\Sigma = \infty$ with probability 1, and (2.13) trivially becomes $0 = 0$.

Expanding (2.11) in a power series of θ and equating the coefficients of θ and θ^2 we obtain the identities:

$$\mathbb{E}[\Sigma] = \int_S f(x)\,\mu(dx)$$

and

$$\mathrm{Var}[\Sigma] = \int_S f(x)^2\,\mu(dx).$$

Now let us look at the general result, which is collected in the following theorem.

Theorem 2.6 (Campbell's Theorem)**.** *Let Π be a Poisson process on S with mean measure μ, and let $f : S \to \mathbb{R}$ be a measurable function. Then the sum*

$$\Sigma = \sum_{x\in\Pi} f(x) \tag{2.14}$$

is absolutely convergent with probability 1 if and only if

$$\int_S \min(|f(x)|, 1)\,\mu(dx) < \infty. \tag{2.15}$$

If this condition is met, then

$$\mathbb{E}[e^{\theta\Sigma}] = \exp\left(\int_S (e^{\theta f(x)} - 1)\,\mu(dx)\right) \tag{2.16}$$

for any $\theta \in \mathbb{C}$ for which the integral on the right converges, and in particular for purely imaginary θ. Furthermore,

$$\mathbb{E}[\Sigma] = \int_S f(x)\,\mu(dx), \tag{2.17}$$

in the sense that the expectation exists if and only if the integral in (2.17) converges. And if the integral in (2.17) converges, then

$$\mathrm{Var}[\Sigma] = \int_S f(x)^2\,\mu(dx), \tag{2.18}$$

whether finite or infinite.

Proof. We already know that equation (2.16) is valid for all $\theta \in \mathbb{C}$, if f is a simple function.

Any positive measurable function f can be expressed as the increasing limit of a sequence of simple functions f_j. Then taking $\theta = -u$ real and negative, and defining $\Sigma = \sum_{x\in\Pi} f(x)$, we have:

$$\begin{aligned}\mathbb{E}[e^{-u\Sigma}] &= \lim_{j\to\infty} \mathbb{E}[e^{-u\Sigma_j}] \\ &= \lim_{j\to\infty} \exp\left(-\int_S (1 - e^{-uf_j(x)})\mu(dx)\right) \\ &= \exp\left(-\int_S (1 - e^{-uf(x)})\mu(dx)\right),\end{aligned}$$

for all $u > 0$, by monotone convergence. If condition (2.15) is met, then the integral converges and tends to 0 when $u \to 0$, showing that Σ is a finite random variable.

Both sides of equation (2.16) are analytic functions of θ in a neighborhood of 0 in $\mathbb{C}$, therefore equation (2.16) is valid also for θ in a complex neighborhood of 0. On the other hand, if (2.15) is not met, then the integral diverges for all $u > 0$, and $\mathbb{E}(e^{-u\Sigma}) = 0$, which implies that $\Sigma = \infty$ with probability 1.

An argument analogous to the case of simple functions establishes the formulas for the expectation (2.17) and the variance (2.18). This proves the theorem for positive functions f. For the general case, the theorem is applied to the positive functions:

$$f^* = \max(f, 0), \quad f^- = \max(-f, 0).$$

The sum Σ converges absolutely if and only if

$$\Sigma_+ = \sum_{x\in\Pi} f^*(x) = \sum_{x\in S_+} f(x)$$

and

$$\Sigma_- = \sum_{x\in\Pi} f^-(x) = -\sum_{x\in S_-} f(x)$$

converge, where Π_+ and Π_- are the restrictions of the process Π to the subsets:

$$S_+ = \{x : f(x) > 0\}, \quad S_- = \{x : f(x) < 0\}.$$

This shows that condition (2.15) is necessary and sufficient for convergence. If it is met, and θ is purely imaginary, then

$$\begin{aligned}
\mathbb{E}[e^{\theta\Sigma}] &= \mathbb{E}[e^{\theta\Sigma_+}] \cdot \mathbb{E}[e^{-\theta\Sigma_-}] \\
&= \exp\left(\int_{S_+} (e^{\theta f} - 1)\, d\mu\right) \cdot \exp\left(\int_{S_-} (e^{-\theta f} - 1)\, d\mu\right) \\
&= \exp\left(\int_{S_+} (e^{\theta f} - 1) + \int_{S_-} (e^{-\theta f} - 1)\, d\mu\right) \\
&= \exp\left(\int_{S} (e^{\theta f} - 1)\, d\mu\right).
\end{aligned}$$

Here we have used the fact that Π_+ and Π_- are independent Poisson processes, as they are restrictions of Π to disjoint sets.

Therefore, (2.16) is valid at least for purely imaginary θ, and its extension to all $\theta \in \mathbb{C}$ for which the integral converges is a matter of analytic continuation. Again, the proofs of (2.17) and (2.18) follow analogously. □

CHAPTER 3
Poisson Processes on the Line

3.1. Definition and Basic Properties

The Poisson process on the line $S = \mathbb{R}$ is a case of great importance. Although we have already seen that the most important general results do not depend on the particular structure of $\mathbb{R}$, its ordered nature gives rise to new questions that would not make sense to ask in higher dimensional spaces, such as the distribution of the distance between successive occurrences.

Furthermore, the Poisson process on $\mathbb{R}$ provides a precise theoretical framework for modeling the random occurrence of events in time. It has a mathematical structure that allows obtaining explicit results on the distribution of times between events related to fundamental distributions, such as the exponential distribution and the gamma distribution.

Due to this temporal interpretation, generally only $t \in \mathbb{R}$, with $t > 0$ are considered, over which the counting process $N(t) := N([0, t])$ is defined, which measures the number of occurrences of the process Π between 0 and t. For simplicity, from here on, if Π is a Poisson process, we will also say that $N(t)$ is, identifying $N(t)$ with $\Pi \cap [0, t]$.

In addition, we will say that the process $N(t)$ has *independent increments* if, for any collection of disjoint intervals, the number of occurrences in each interval is independent of each other, and that it has *stationary increments* if the distribution of the number of events in an interval of length t is invariant in time, that is, for all $t, s, h > 0$

$$P\{N(t+s) - N(s) = n\} = P\{N(t+h) - N(h) = n\}, \quad n = 0, 1, 2, \ldots.$$

In the previous chapter we already saw that a *homogeneous Poisson process* with rate λ, with $\lambda > 0$, is one whose induced measure verifies $\mu(A) = \lambda|A|$. Translating this to the line, we have that $\mu([s, s+t]) = \lambda t$ or, equivalently:

$$P\{N(t+s) - N(s) = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots.$$

In this chapter, we will focus on the study of the homogeneous Poisson process. Later, we will see that most non-homogeneous processes can be transformed into homogeneous ones via a simple transformation.

As an introduction to a second definition of a homogeneous Poisson process, we recall that a function f is $o(h)$ if $\lim_{h\to 0} \frac{f(h)}{h} = 0$. With this, we can give the following definition:

Definition 3.1. The process $\{N(t), t \geq 0\}$ is said to be a *Poisson process* with rate λ, $\lambda > 0$, if it meets the following conditions:

(i) $N(0) = 0$, with probability 1.

(ii) The process has stationary and independent increments.

(iii) $P\{N(h) = 1\} = \lambda h + o(h)$.

(iv) $P\{N(h) \geq 2\} = o(h)$.

Theorem 3.2. *Definitions 2.1 (in the case $S = \mathbb{R}$) and 3.1 are equivalent.*

Proof. Let us see that 3.1 implies 2.1. To do this, we define

$$P_0(t) = P(N(t) = 0).$$

We derive a differential equation for $P_0(t)$ as follows:

$$\begin{aligned} P_0(t+h) &= P(N(t+h) = 0) = P(N(t) = 0, N(t+h) - N(t) = 0) \\ &= P(N(t) = 0)P(N(t+h) - N(t) = 0) = P_0(t)[1 - \lambda h + o(h)]. \end{aligned}$$

Dividing by h and taking the limit when $h \to 0$, we obtain

$$\frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) + \frac{o(h)}{h}.$$

Therefore,

$$P_0'(t) = -\lambda P_0(t).$$

Integrating this differential equation, we obtain

$$\log P_0(t) = -\lambda t + C.$$

Given that $P_0(0) = P(N(0) = 0) = 1$, we have $C = 0$, and therefore,

$$P_0(t) = e^{-\lambda t}.$$

Similarly, for $n \geq 1$, we have

$$\begin{aligned} P_n(t+h) &= P(N(t+h) = n) \\ &= P(N(t) = n, N(t+h) - N(t) = 0) \\ &+ P(N(t) = n-1, N(t+h) - N(t) = 1) \end{aligned}$$

$$+P(N(t)=n, N(t+h)-N(t)\geq 2).$$

The last term is $o(h)$, so using the previous equation we obtain

$$P_n(t+h)=P_n(t)P_0(h)+P_{n-1}(t)P_1(h)+o(h).$$

Dividing by h and taking the limit when $h\to 0$, we obtain

$$\frac{P_n(t+h)-P_n(t)}{h}=-\lambda P_n(t)+\lambda P_{n-1}(t)+\frac{o(h)}{h}.$$

Therefore,

$$P_n'(t)=-\lambda P_n(t)+\lambda P_{n-1}(t).$$

Rewriting in terms of exponents,

$$e^{\lambda t}[P_n'(t)+\lambda P_n(t)]=\lambda e^{\lambda t}P_{n-1}(t).$$

Thus, we can write

$$\frac{d}{dt}(e^{\lambda t}P_n(t))=\lambda e^{\lambda t}P_{n-1}(t).$$

Using the case $n=1$, we obtain

$$\frac{d}{dt}(e^{\lambda t}P_1(t))=\lambda.$$

Integrating,

$$P_1(t)=(A+C)e^{-\lambda t}.$$

Given that $P_1(0)=0$, we have $A=\lambda$, which gives us

$$P_1(t)=\lambda te^{-\lambda t}.$$

To prove that

$$P_n(t)=e^{-\lambda t}\frac{(\lambda t)^n}{n!},$$

we use mathematical induction. We assume the expression is valid for $n-1$. Then, applying the previous differential equation,

$$\frac{d}{dt}(e^{\lambda t}P_n(t))=\frac{\lambda(\lambda t)^{n-1}}{(n-1)!}.$$

Integrating,

$$e^{\lambda t}P_n(t)=\frac{(\lambda t)^n}{n!}+C.$$

Since $P_n(0) = 0$, we conclude that

$$P_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}.$$

$N([s, s+t])$ follows a Poisson distribution with parameter $\mu = \lambda t$, thus proving the implication. Now let us see the opposite implication. Suppose that

(i) For any collection of disjoint intervals, the number of occurrences in each of them is independent and

(ii) $N([s, s+t])$ follows a Poisson distribution with parameter $\mu = \lambda t$.

Since the distribution of $N([s, s+t])$ does not depend on s, we can also state that the increments are stationary, as well as independent, by (i). On the other hand, clearly

$$N(0) = N([s, s]) = 0$$

and, finally using Taylor's theorem, we see that

$$P\{N(h) = 1\} = \lambda h e^{-\lambda h} = \lambda h + o(h)$$

and

$$P\{N(h) > 1\} = 1 - P\{N(h) = 1\} = 1 - \lambda h + o(h)$$

thus ending the proof.

□

This alternative formulation highlights the usefulness of the Poisson process: starting from general and reasonably weak hypotheses, one naturally arrives at a characterization that allows understanding the process not as a particular model, but as an inevitable consequence under certain minimal conditions.

3.2. Distribution of Inter-arrival Times and Waiting Times

As we have already advanced in the previous section, the structure of $\mathbb{R}$ allows us to study the time or distance between two successive occurrences of the process. For $n \geq 1$, we define X_n as the random variable corresponding to the time between the $(n-1)$-th and the n-th event. The sequence $\{X_n, n \geq 1\}$ is called the *sequence of inter-arrival times*.

We will now determine the distribution of the variables X_n. To do this, we first observe that the event $\{X_1 > t\}$ occurs if, and only if, no event of the Poisson process occurs in the interval $[0, t]$, so

$$P\{X_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}.$$

Therefore, X_1 follows an exponential distribution with mean $1/\lambda$. To obtain the distribution of X_2 conditioned on X_1, we have

$$\begin{aligned} P\{X_2 > t \mid X_1 = s\} &= P\{0 \text{ events in } (s, s+t] \mid X_1 = s\} \\ &= P\{0 \text{ events in } (s, s+t]\} \quad \text{(by independent increments)} \\ &= e^{-\lambda t} \quad \text{(by stationary increments).} \end{aligned}$$

With this, we conclude that X_2 is also an exponential random variable with mean $1/\lambda$, and furthermore, that X_2 is independent of X_1. Repeating the same argument, we obtain the following result:

Proposition 3.3. *The variables $X_n, n = 1, 2, \ldots$ are independent and identically distributed exponential random variables with mean $1/\lambda$.*

Another variable of interest is S_n, which represents the arrival time of the n-th event, (or *waiting time until the n-th event*).

Proposition 3.4. *The random variable $S_n, n = 1, 2, \ldots$ follows a gamma distribution with parameters n and λ.*

Proof. Note that the n-th event occurs before or at time t if, and only if, the number of events occurring up to time t is at least n. That is,

$$N(t) \geq n \iff S_n \leq t.$$

Therefore,

$$P(S_n \leq t) = P(N(t) \geq n) = \sum_{j=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!}.$$

Differentiating this function, we obtain the density function of S_n:

$$\begin{aligned} f(t) &= -\sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^j}{j!} + \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^{j-1}}{(j-1)!} \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}. \end{aligned}$$

□

3.3. Conditional Distribution of Arrival Times

Suppose we know that exactly one event of a Poisson process has occurred up to time t, and we want to determine the distribution of the time at which the event occurred. Since a Poisson process has stationary and independent increments, it is reasonable that each interval in $[0, t]$ of equal length has the same probability of containing the

event. In other words, the time at which the event occurred should be uniformly distributed in $[0,t]$. This is easily verified considering that, for $s \leq t$,

$$\begin{aligned} P(X_1 < s \mid N(t)=1) &= \frac{P(X_1 \leq s, N(t)=1)}{P(N(t)=1)} \\ &= \frac{P(1 \text{ event in } [0,s]), 0 \text{ events in } (s,t])}{P(N(t)=1)} \\ &= \frac{P(1 \text{ event in } [0,s])P(0 \text{ events in } (s,t])}{P(N(t)=1)} \\ &= \frac{\lambda s e^{-\lambda s} e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} = \frac{s}{t}. \end{aligned}$$

Given n variables $Y_1, \ldots, Y_n$, we denote $Y_{(k)}$ as the order statistic k, that is, the k-th smallest value among $Y_1, Y_2, \ldots, Y_n$.

Proposition 3.5. *Let $Y_1, \ldots, Y_n$ be independent and identically distributed continuous random variables and absolutely continuous with density function $f(y)$, then the joint density function of the order statistics $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ is given by*

$$f(y_1, y_2, \ldots, y_n) = n! \prod_{i=1}^{n} f(y_i), \quad 0 < y_1 < y_2 < \ldots < y_n < t.$$

Proof. The previous result follows from the following two statements:

(i) $(Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})$ will take the value $(y_1, y_2, \ldots, y_n)$ if $(Y_1, Y_2, \ldots, Y_n)$ is equal to any of the $n!$ permutations of $(y_1, y_2, \ldots, y_n)$.

(ii) The joint density function of $(Y_1, Y_2, \ldots, Y_n)$ evaluated at $(y_1, y_2, \ldots, y_n)$ is obtained as the product of the marginal densities, that is:

$$f(y_1)f(y_2)\cdots f(y_n) = \prod f(y_i),$$

when $(y_{i_1}, y_{i_2}, \ldots, y_{i_n})$ is a permutation of $(y_1, y_2, \ldots, y_n)$.

Then the joint density of $(Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})$ will be the sum of the joint densities $(Y_1, Y_2, \ldots, Y_n)$ evaluated at each of the $n!$ possible permutations. □

If the random variables Y_i, with $i = 1, \ldots, n$, are uniformly distributed in the interval $(0,t)$, it follows from proposition 3.5 that the joint density function of the order statistics $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ is

$$f(y_1, y_2, \ldots, y_n) = \frac{n!}{t^n}, \quad 0 < y_1 < y_2 < \ldots < y_n < t.$$

We are now ready for the following useful theorem.

Theorem 3.6. *Suppose that $N(t) = n$, the n arrival times $S_1, S_2, \ldots, S_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed in the interval $(0,t)$.*

Proof. We first calculate the conditional density function of $S_1, S_2, \ldots, S_n$ given that $N(t) = n$. Let $0 < t_1 < t_2 < \ldots < t_{n+1} = t$, and let h_i be small enough such that $t_i + h_i < t_{i+1}, i = 1, \ldots, n$, we define $A_i = [t_i, t_i + h_i]$.

In the previous chapter we saw that conditioning a Poisson process such that $N(S) = n$ converts the process into a Bernoulli process. Applying (2.8) we obtain that:

$$\begin{aligned} &P(t_i \le S_i \le t_i + h_i,\, i = 1, 2, \ldots, n \mid N(t) = n) \\ &= P(N(A_i) = 1,\, i = 1, 2, \ldots, n \mid N(t) = n) \\ &= n! \prod_{i=1}^{n} \frac{h_i}{t} = \frac{n!}{t^n} h_1 h_2 \cdots h_n. \end{aligned}$$

Therefore,

$$\frac{P(t_i \le S_i \le t_i + h_i, i = 1, 2, \ldots, n \mid N(t) = n)}{h_1 h_2 \cdots h_n} = \frac{n!}{t^n}.$$

Taking limits when $h_i \to 0$ for $i = 1, \ldots, n$, we obtain the density function of $S_1, \ldots, S_n$ conditioned on $N(t) = n$:

$$f_{S_1,\ldots,S_n}(t) = \frac{n!}{t^n}, \quad 0 < S_1 < S_2 < \ldots < S_n < t.$$

□

Let us see an example of the utility of theorem 3.6:

Example 3.7 (Waiting Times)**.** Suppose that passengers arrive at a train station according to a Poisson process with rate λ. If the train departs at time t, we want to calculate the expected sum of the waiting times of the passengers who arrive in $(0, t]$. That is, we want to calculate

$$\mathbb{E}\left[\sum_{i=1}^{N(t)} (t - S_i)\right],$$

where S_i is the arrival time of the i-th passenger. Conditioning on $N(t)$ we obtain:

$$\mathbb{E}\left[\sum_{i=1}^{N(t)} (t - S_i) \mid N(t) = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} (t - S_i) \mid N(t) = n\right]$$

$$= nt - \mathbb{E}\left[\sum_{i=1}^{n} S_i \mid N(t) = n\right].$$

If we define $U_1, \ldots, U_n$ as n independent random variables with uniform distribution in $(0, t)$, then

$$\mathbb{E}\left[\sum_{i=1}^{n} S_i \mid N(t) = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} U_i\right] \quad \text{(by theorem 3.6)}$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} U_{(i)}\right] = \frac{nt}{2}. \quad \text{(given that } \sum_{i=1}^{n} U_{(i)} = \sum_{i=1}^{n} U_i\text{)}$$

Therefore,

$$\mathbb{E}\left[\sum_{i=1}^{N(t)} (t - S_i) \mid N(t) = n\right] = nt - \frac{nt}{2} = \frac{nt}{2}.$$

Finally, taking the expectation with respect to $N(t)$,

$$\mathbb{E}\left[\sum_{i=1}^{N(t)} (t - S_i)\right] = \frac{t}{2}\mathbb{E}[N(t)] = \frac{\lambda t^2}{2}.$$

3.4. Relation to Queueing Theory

The structure of the Poisson process makes it the most classic and fundamental model for describing random arrival phenomena of customers to service systems, making it a recurring tool in the study of problems related to queueing theory. In this section we will present an important application of theorem 3.6, which we will subsequently use for solving a concrete queueing theory problem.

Suppose that each event of a Poisson process with rate λ is classified as a type I or type II event, and that the probability that an event is classified as type I depends on the time at which it occurs. Specifically, suppose that if an event occurs at time s, then, independently of all other events, it is classified as a type I event with probability $P(s)$ and as a type II event with probability $1 - P(s)$. From theorem 3.6 we can demonstrate the following proposition:

Proposition 3.8. *If $N_1(t)$ and $N_2(t)$ represent the number of type I and type II events that occur up to time t, then $N_1(t)$ and $N_2(t)$ are independent Poisson random variables with respective means λp and $\lambda(1-p)$, where*

$$p = p(t) = \frac{1}{t}\int_0^t P(s)ds.$$

Proof. We calculate the joint distribution of $N_1(t)$ and $N_2(t)$ conditioning on $N(t)$ and applying the law of total probability:

$$\begin{aligned}
&P(N_1(t) = n, N_2(t) = m) \\
&= \sum_{k=0}^{\infty} P(N_1(t) = n, N_2(t) = m \mid N(t) = k) P(N(t) = k) \\
&= P(N_1(t) = n, N_2(t) = m \mid N(t) = n + m) P(N(t) = n + m).
\end{aligned}$$

Let us consider an arbitrary event that occurred in the interval $(0, t]$. If it had occurred at time s, then the probability of it being a type I event would be $P(s)$. Since, by Theorem 3.6, this event will have occurred at some instant uniformly

distributed in $(0, t)$, it follows that the probability of it being a type I event is

$$p(t) = \frac{1}{t}\int_0^t P(s)ds.$$

Independently of the other events, the joint distribution of $N_1(t)$ and $N_2(t)$ given $N(t) = n + m$ reduces to the probability of obtaining n successes and m failures in $n + m$ independent trials where $p = p(t)$ is the probability of success in each trial. That is,

$$P(N_1(t) = n, N_2(t) = m \mid N(t) = n + m) = \binom{n+m}{n} p^n (1-p)^m.$$

Therefore, the joint distribution of $N_1(t)$ and $N_2(t)$ is

$$P(N_1(t) = n, N_2(t) = m) = \frac{(n+m)!}{n!m!} p^n (1-p)^m e^{-\lambda t} \frac{(\lambda t)^{n+m}}{(n+m)!}.$$

$$= e^{-\lambda p}\frac{(\lambda p)^n}{n!} e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!}.$$

which completes the proof. □

The importance of the proposition is illustrated in the following example:

Example 3.9 (Infinite Server Queue)**.** Suppose that customers arrive at a service station according to a Poisson process with rate λ. Upon arrival, the customer is immediately served by one of an infinite number of possible servers, and service times are assumed to be independent with a common distribution G.

To calculate the joint distribution of the number of customers who have completed their service and the number of customers who are in service at time t, we call an incoming customer a type-I customer if they complete their service before t and a type-II customer if they do not complete it before t. Then, if a customer enters at time s, with $s \le t$, then they will be a type-I customer if their service time is less than $t - s$, and given that the distribution of service times is G, the probability of this is $G(t - s)$. Therefore,

$$P(s) = G(t - s), \quad s \le t.$$

From Proposition 3.8, we obtain that the distribution of $N_1(t)$, the number of customers who have completed their service at time t, follows a Poisson process with mean

$$\mathbb{E}[N_1(t)] = \lambda \int_0^t G(t-s)\, ds = \lambda \int_0^t G(y)\, dy.$$

Similarly, the number of customers who are in service at time t, denoted by $N_2(t)$, follows a Poisson distribution with mean

$$\mathbb{E}[N_2(t)] = \lambda \int_0^t (1 - G(y))\, dy.$$

Furthermore, $N_1(t)$ and $N_2(t)$ are independent.

3.5. Non-homogeneous Poisson Process

As we have already mentioned, the reason for concentrating on homogeneous processes is that a non-homogeneous process on the line can be converted into a homogeneous one by a monotone transformation. More precisely, consider a Poisson process Π on $\mathbb{R}$ whose induced measure μ is finite on bounded intervals (it does not need to be given by a rate $\lambda(t)$). Then, the measure μ is uniquely determined by its value on intervals $(a, b]$. We define $M(t)$ as

$$M(t) = \mu((0, t]) = \mathbb{E}\{N((0, t]\}, \quad t \geq 0,$$

such that M is an increasing function and satisfies

$$\mu(a, b] = M(b) - M(a).$$

We thus conclude that μ is given entirely by the function M, and is called the *Stieltjes measure* associated with M.

Consider now the non-homogeneous process Π, and its transformation $\Pi_1 = M(\Pi)$. By theorem 2.4 of transformation of Poisson processes, we have that if $x = M(t)$, then

$$\mu^*((0, x]) = \mu^*((0, M(t)]) = \mu((0, t]) = M(t) = x,$$

so $M(\Pi)$ is a homogeneous Poisson process with rate $\lambda = 1$.

All results of this chapter apply to Π_1 and, therefore, imply properties of Π by inverting M. The only possible ambiguity arises if M has flat sections for which $M^{-1}\{y\}$ is an interval instead of a single point. However, this can occur at most in a countable set of Π_1.

Unfortunately, this nonlinear transformation of the line destroys some of the most useful properties. In particular, although the times between successive occurrences of Π_1 are independent, those of Π are not necessarily so. For this reason, it is usually easier to transform the process to homogeneity before performing any serious analysis.

CHAPTER 4
Poisson Models for Spatial Data Analysis

One of the most relevant applications of the Poisson process is the modeling of point processes in space, with the aim of analyzing phenomena that manifest as dispersed locations in a geographical region. In this chapter, we will study different methods to extract significant information from spatial point processes. The objective is to understand how the distribution of events in space can reveal underlying patterns or relationships with other variables.

4.1. Non-parametric Models of the Intensity Function of a Poisson Process

As we have already seen, it is common to describe Poisson processes using an intensity function. Our goal is, given a dataset, to assume that it follows a Poisson process and, based on that assumption, to estimate the intensity function based on the available information.

We will start working with an example, the `gorillas` dataset [2] from the `spatstat` package, which collects the locations of 647 gorilla nests observed in the Kagwene Gorilla Sanctuary, in Cameroon, as well as many other geographical variables. Our goal is to find a way to model and analyze the spatial distribution of these nests, identifying potential patterns or relationships with other variables.

The first approach we propose is to use a kernel density estimator. Let $\mathbf{x} = (x_1, \ldots, x_d)$ and $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n$ be a dataset with $\mathbf{x_i} = (x_{i1}, x_{i2}, \ldots, x_{id})$ for $i = 1, 2, \ldots n$. The kernel estimator of $\lambda(\mathbf{x})$ is given by

$$\hat{\lambda}(\mathbf{x}) = \sum_{i=1}^{n} K_h(x - x_i) \tag{4.1}$$

where $K_h(\cdot)$ is the kernel function, which is typically an isotropic multivariate Gaussian kernel with standard deviation h. This parameter h is called the *bandwidth* or *smoothing parameter*, which controls the degree of smoothing. A small value of h

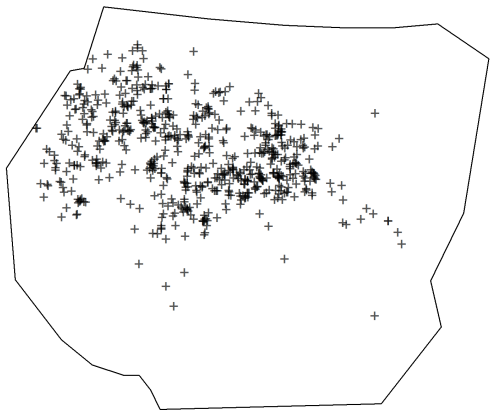

Figure 4.1: Locations of gorilla nests in the sanctuary

produces a tighter estimator but with greater variability, while a large value smooths more but introduces greater bias. We can write (4.1) explicitly as a function of h as:

$$\hat{\lambda}(\mathbf{x}) = \frac{1}{h^d} \sum_{i=1}^{n} K_1 \left(\frac{\mathbf{x} - \mathbf{x}_i}{h} \right), \tag{4.2}$$

which, for the case $d = 2$, can be written as

$$\hat{\lambda}(\mathbf{x}) = \frac{1}{h^d} \sum_{i=1}^{n} \prod_{j=1}^{d} k_1 \left(\frac{x_j - x_{ij}}{h} \right)$$

where k_1 is the univariate version of the same Gaussian kernel.

It is important to highlight that the estimator $\hat{\lambda}(\mathbf{x})$ is not unbiased, not even in the homogeneous case. Using Campbell's Theorem 2.6, we have that

$$\mathbb{E}\left[\hat{\lambda}(\mathbf{x})\right] = \mathbb{E}\left[\sum_{i=1}^{n} K_h(\mathbf{x} - \mathbf{x}_i)\right] = \int_W K(\mathbf{x} - u)\lambda(u)du,$$

which for the homogeneous case $\lambda(\mathbf{x}) = \beta$ is

$$\mathbb{E}\left[\hat{\lambda}(\mathbf{x})\right] = \beta \int_W K(\mathbf{x} - u)du.$$

This motivates the definition of a new corrected estimator. Calling $e(\mathbf{x})$ to $\int_W K(\mathbf{x} - u)du$ we define the *uniformly corrected kernel estimator* (Baddeley et al. [3] Ch 6.5) as:

$$\hat{\lambda}^{(U)}(\mathbf{x}) = \frac{1}{e(\mathbf{x})} \sum_{i=1}^{n} K_h(x - x_i). \tag{4.3}$$

It can be verified using a calculation analogous to the previous one that, in the case of the homogeneous process, it is an unbiased estimator.

The `spatstat` package offers the possibility of calculating this kernel estimator using the `density` function. The choice of the smoothing parameter is not trivial and there are multiple ways to calculate it with very different results. In this case, the `bw.diggle` method from Berman and Diggle [4] has been used, which minimizes the mean squared error via cross-validation. Figure 4.2 shows the contour lines of the estimator 4.1 for the `gorillas` point set with $h = 0.38$.

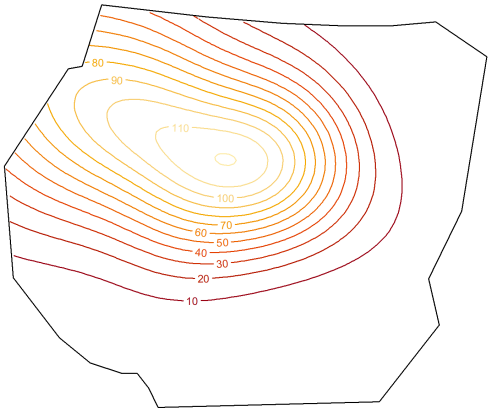

Figure 4.2: Contour lines of the kernel intensity estimator

As a first analysis, the intensity estimator via kernel smoothing proves to be a very useful tool, as it provides an intuitive visualization of the density of events in space. However, it is a purely descriptive approach that, while capturing the local structure of the process, does not incorporate additional information about possible explanatory factors of the pattern, nor does it allow making inferences about the underlying generating mechanism.

4.2. Parametric Models of Process Intensity

Another idea to approximate the density function of the process is to consider a parametric model of the intensity function and adjust the parameters of said function to the dataset. We recall that any non-negative and measurable function can act as an intensity function of a Poisson process. In this section, we will study a series of models called *log-linear* proposed in the book by Baddeley et al. [3] which have the following form:

$$\lambda_{\boldsymbol{\theta}}(\mathbf{x}) = \exp\left(B(\mathbf{x}) + \boldsymbol{\theta}^\top \mathbf{X}(\mathbf{x})\right) = \exp\left(B(\mathbf{x}) + \theta_1 X_1(\mathbf{x}) + \cdots + \theta_p X_p(\mathbf{x})\right) \tag{4.4}$$

where $B(\mathbf{x})$ and $\mathbf{X}(\mathbf{x}) = (X_1(\mathbf{x}), X_2(\mathbf{x}), \ldots, X_p(\mathbf{x}))$ are known functions and $\boldsymbol{\theta}^\top = (\theta_1, \theta_2, \ldots, \theta_p)$ are the parameters to be estimated.

The parameters $\theta_1, \ldots, \theta_p$ can be estimated using the maximum likelihood method. The likelihood function for a Poisson process, governed by an intensity function $\lambda_{\boldsymbol{\theta}}(\mathbf{x})$,

for a dataset $\mathbf{x}_1, \ldots, \mathbf{x}_n$ observed in a subset W is the following:

$$\begin{aligned} L(\boldsymbol{\theta}) &= f(\mathbf{x}_1, \ldots, \mathbf{x}_n | N(W) = n) \cdot P(N(W) = n) \\ &= \prod_{i=1}^{n} \frac{\lambda_{\boldsymbol{\theta}}(\mathbf{x}_i)}{\int_W \lambda_{\boldsymbol{\theta}}(\mathbf{x})} \cdot \frac{\left(\int_W \lambda_{\boldsymbol{\theta}}(\mathbf{x})\right)^n \exp\left(-\int_W \lambda_{\boldsymbol{\theta}}(\mathbf{x})\, d\mathbf{x}\right)}{n!} \\ &= \frac{1}{n!} \prod_{i=1}^{n} \lambda_{\boldsymbol{\theta}}(\mathbf{x}_i) \cdot \exp\left(-\int_W \lambda_{\boldsymbol{\theta}}(\mathbf{x})\, d\mathbf{x}\right). \end{aligned}$$

Taking logarithms and discarding constants, we obtain the log-likelihood function:

$$\log L(\theta) = \sum_{i=1}^{n} \log \lambda_\theta(\mathbf{x}_i) - \int_W \lambda_\theta(\mathbf{x})\, d\mathbf{x}.$$

And finally, substituting (4.4):

$$\log L(\theta) = \sum_{i=1}^{n} \left(B(\mathbf{x}_i) + \theta^\top \mathbf{X}(\mathbf{x}_i)\right) - \int_W \exp\left(B(\mathbf{x}) + \theta^\top \mathbf{X}(\mathbf{x})\right) d\mathbf{x}. \tag{4.5}$$

The log-likelihood (4.5) is a concave function of the parameter θ and is differentiable with respect to θ, even if the functions B and X_j are not continuous. If the matrix

$$M = \int_W \mathbf{X}(\mathbf{x})\mathbf{X}(\mathbf{x})^\top\, d\mathbf{x}$$

is positive definite, then the model is identifiable (Lehmann and Casella [7], example 6.3, p. 470). If the data are such that $\sum_i X_j(\mathbf{x}_i) \neq 0$ for all j, the maximum likelihood estimator (MLE) exists and is unique. Unless there are additional restrictions on θ, the MLE is the solution of the equations obtained by differentiating and equating the likelihood function to 0:

$$\sum_{i=1}^{n} \mathbf{X}(\mathbf{x}_i) - \int_W \mathbf{X}(\mathbf{x})\lambda_\theta(\mathbf{x})\, d\mathbf{x} = 0, \tag{4.6}$$

or equivalently

$$\sum_{i=1}^{n} X_j(\mathbf{x}_i) - \int_W X_j(\mathbf{x})\lambda_\theta(\mathbf{x})\, d\mathbf{x} = 0$$

for $j = 1, \ldots, p$.

The integral in (4.6) cannot be solved, in general, analytically, so in most cases equations (4.6) can only be solved numerically.

Furthermore, we know that, under regularity conditions, the MLE $\hat{\theta}$ is a consistent estimator of θ and that for sufficiently large samples, $\hat{\theta}$ follows approximately a multivariate normal distribution with mean θ and covariance matrix $\mathbf{I}_\theta^{-1}$, where $\mathbf{I}_\theta$ is the Fisher information matrix, which is defined as the covariance matrix of the gradient of the log-likelihood:

$$\mathbf{I}_\theta = \operatorname{Var}\left[\frac{\partial \log L(\theta)}{\partial \theta}\right] = \mathbb{E}_\theta\left[\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^\top\right] \tag{4.7}$$

where $\mathbb{E}_\theta$ and Var_θ denote the mean and variance when the true value of the parameter is θ. Under certain regularity conditions, (4.7) can be expressed as the expected value of the negative Hessian matrix:

$$\mathbf{I}_\theta = \mathbb{E}_{\boldsymbol{\theta}} \left[-\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^\top} \right].$$

For the particular case (4.4) we have that the value of the Hessian is the following:

$$H(\theta) = \int_W \mathbf{X}(\mathbf{x})\mathbf{X}(\mathbf{x})^\top \lambda_\theta(\mathbf{x})\, d\mathbf{x}.$$

As it does not depend on the dataset, this coincides with the Fisher information matrix

$$(I_\theta)_{ij} = \int_W X_i(\mathbf{x}) X_j(\mathbf{x}) \lambda_\theta(\mathbf{x})\, d\mathbf{x}.$$

This is the basis behind the calculations of standard errors and confidence intervals for Poisson models.

Let us return to the example of the gorilla nests. One of the most interesting spatial covariates collected in the dataset is the vegetation type of the territory (Figure 4.3).

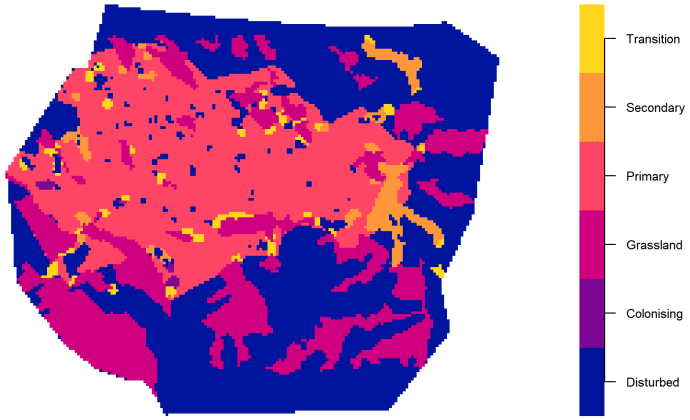

Figure 4.3: *Gorillas.* Vegetation types

The `R` package `spatstat` also allows fitting parametric Poisson models to datasets. This is done using the package's `ppm` (point process model) function. The format is as follows: `ppm(X ~ trend, data)`. Here `X` is the point process and `trend` is an expression containing the various spatial variables on which the intensity function depends, while `data` is a list containing all the information necessary to fit the model.

A relatively simple model is to consider that the intensity is constant within a given type of region (piecewise constant). This is equivalent to assuming the following model:

$$\lambda_\theta(\mathbf{x}) = \exp(\theta_0 + \sum_{i=1}^{k} \theta_i X_i(\mathbf{x}))$$

where k is the number of distinct region types and $V_i(\mathbf{x})$ is the dichotomous covariate that equals 1 when $\mathbf{x}$ is in a given region type and 0 otherwise. In the case of gorillas, we can assume that the X_i correspond to the different vegetation types. Figure 4.4 shows the output of the `ppm` command evaluating the proposed model.

```
> model <- ppm(gor ~ vegetation, data=gex)
> model
Nonstationary Poisson process
Fitted to point pattern dataset 'gor'

Log intensity:  ~vege

Fitted trend coefficients:
(Intercept)  vegeColo  vegeGras  vegePrim  vegeSeco  vegeTran
     2.3238   2.0817    -0.7721    2.1148    1.1341    1.6151

             Estimate   S.E. CI95.lo CI95.hi Ztest   Zval
(Intercept)   2.3238 0.1060  2.1161  2.5316   *** 21.923
vegeColo      2.0817 0.5870  0.9312  3.2322   *** 3.546
vegeGras     -0.7721 0.2475 -1.2571 -0.2871    ** -3.120
vegePrim      2.1148 0.1151  1.8892  2.3403   *** 18.373
vegeSeco      1.1341 0.2426  0.6586  1.6096   *** 4.675
vegeTran      1.6151 0.2647  1.0964  2.1339   *** 6.102
```

Figure 4.4: Output of the `ppm` command from `spatstat`

The `Estimate` column contains the value of the estimators θ_i. The table also includes the standard error of each of the estimators as well as a confidence interval for them. Table 4.1 shows the estimated intensity within each region.

Vegetation Type	**Log-intensity**	**Estimated intensity**
Disturbed	2.324	10.22
Colonising	4.406	81.96
Grassland	1.552	4.72
Primary	4.439	84.79
Secondary	3.458	31.80
Transition	3.939	51.33

Table 4.1: Estimated intensity of the Poisson process according to vegetation type.

In addition, the command performs a hypothesis test based on the asymptotic normality of the maximum likelihood estimator. In particular, the null hypothesis $H_0 : \theta_j = 0$ is evaluated against the alternative $H_1 : \theta_j \neq 0$, using Z-type statistics whose variance is estimated via the inverse of the Fisher information matrix. The results appear in the `Zval` and `Ztest` columns, where it is observed that all p-values are less than 0.01, allowing the rejection of the nullity of the coefficients at a significance level $\alpha = 0.01$.

4.3. Model Selection via AIC

We have already seen how a model is fitted to a dataset. However, for datasets with multiple covariates, there is a very large number of possible models to evaluate. Therefore, it is necessary to have a criterion that helps us select a model that is simple enough, but at the same time fits the data adequately. This is possible using the *Akaike Information Criterion* (AIC), defined as:

$$\text{AIC} = -2 \log L_{\max} + 2p \tag{4.8}$$

where $L_{\max} = L(\hat{\boldsymbol{\theta}})$ is the maximized likelihood of the considered model, and p is the number of estimated parameters. The model that minimizes the AIC will be preferred, since this criterion seeks a compromise between the quality of the fit and the simplicity of the model, penalizing those excessively complex models that could overfit the data.

We are going to see this with an example. The `clmfires` dataset [1] collects the location of all forest fires recorded in Castilla-La Mancha between 1998 and 2007, as well as their main cause. For this experiment, we will consider only those caused by lightning strikes (Figure 4.5). The dataset also collects information on a series of spatial covariates which are elevation, slope, orientation, and land use.

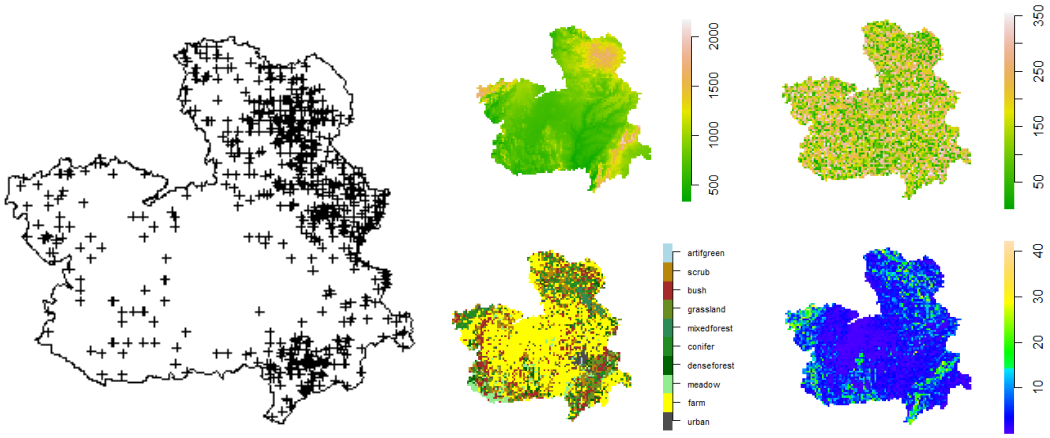

Figure 4.5: Spatial distribution of lightning-caused fires (left) along with associated covariates (center and right). From top to bottom and left to right: elevation, land use, orientation, and slope.

Since we have 4 covariates, we have a total of 16 possible (linear) models, including the homogeneous case. To select the best model, we calculate (4.8) for each of them using the `AIC` function of `spatstat` and select the one that obtains the lowest result. Table 4.2 shows the models that present the best results in this regard.

We observe that, in terms of AIC, the optimal model is the one that uses elevation and orientation, as well as land use. However, all models that use both elevation and land use obtain very similar results. On the other hand, the simple model that obtains

Model	AIC
elevation + orientation + landuse	12342.16
elevation + slope + orientation + landuse	12342.19
elevation + slope + landuse	12342.77
elevation + landuse	12342.79
elevation + slope + orientation	12447.51
elevation + slope	12448.61
elevation + orientation	12450.88
elevation	12452.12
⋮	⋮

Table 4.2: Best models for explaining fires according to the Akaike Information Criterion

the best results is elevation, this being the variable that would best explain the fires on its own.

Using the `predict(model)` command from `spatstat` we can obtain the fitted intensity estimator. Figure 4.6 shows the intensity estimator for the case of the model `X` $\sim$ `elevation + orientation + landuse`.

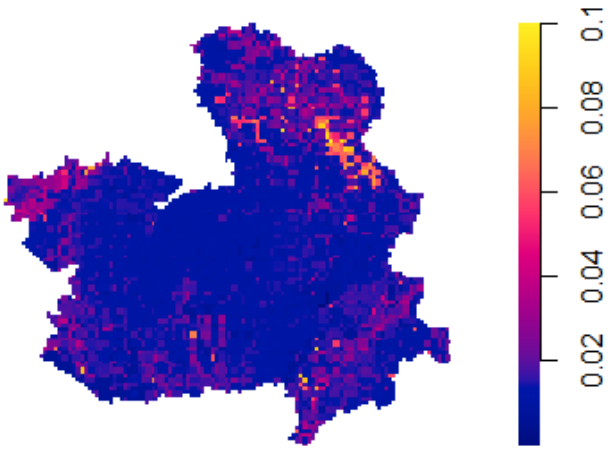

Figure 4.6: Estimated fire intensity.

CHAPTER 5
Conclusions

The Poisson process is one of the fundamental models in the theory of stochastic processes and occupies a central place in applied statistics and probability. Its importance lies in its ability to model the random occurrence of events in a given domain, whether temporal, spatial, or more abstract. Despite its formal simplicity, the Poisson process is capable of effectively describing a wide variety of real phenomena, which makes it an extremely valuable tool in both theoretical contexts and practical applications.

We began by studying its general formulation, and we were able to observe that many of its fundamental properties do not depend on the specific space in which the process is defined, allowing the same probabilistic framework to be applied to very different situations without the need to modify the model structure. This abstraction makes the Poisson process a key piece in the development of point process theory and in the construction of more complex models.

Subsequently, we delved into the particular case of the Poisson process on the real line, which presents interesting particularities due to the ordered structure of said space. This version of the model allows interpreting the points of the process as time instants, which opens the door to the study of inter-event times, waiting times, and other temporal properties that are essential in applications such as queueing theory.

Finally, we explored its utility as a statistical tool in the analysis of spatial data. In this context, the Poisson process offers a natural framework to represent the distribution of observed events in space, allowing the extraction of valuable information about the generating phenomenon. This type of analysis is especially relevant in disciplines where data have a clear spatial component, such as ecology, epidemiology, or environmental sciences, where the Poisson process can serve as a reference model.

Bibliography

[1] Dataset `clmfires` from the `spatstat.data` package, 2025. URL https://search.r-project.org/CRAN/refmans/spatstat.data/html/clmfires.html. Accessed: 2025-04-30.

[2] Dataset `gorillas` from the `spatstat.data` package, 2025. URL https://search.r-project.org/CRAN/refmans/spatstat.data/html/gorillas.html. Accessed: 2025-04-30.

[3] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Interdisciplinary Statistics. Chapman and Hall/CRC Press, 2015.

[4] M. Berman and P. J. Diggle. Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(1):81–92, 1989.

[5] J. F. C. Kingman. *Poisson Processes*, volume 3 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 1993.

[6] J. F. C. Kingman and S. J. Taylor. *Introduction to Measure and Probability*. Cambridge University Press, Cambridge, 1966.

[7] E. L. Lehmann and George Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, New York, 2nd edition, 1998.

[8] Sheldon M. Ross. *Stochastic Processes*. John Wiley & Sons, United States, 2nd edition, 1996.