LUISS

Degree Program in Management and Computer Science

Chair **Data Analysis for Business**

Bootstrap Inference for the Epidemic-type Aftershock Sequence (ETAS) Model

Ernesto Desideri - 281491

CANDIDATE

Prof. Alessia Caponera

SUPERVISOR

Academic Year **2024/2025**

Alla città de L'Aquila,
che ha avuto la forza di rinascere
e ai suoi cittadini,
che grazie alla loro resilienza
lo hanno reso possibile

A mio padre e a mia madre,
che mi hanno reso quello che sono
lasciandomi sempre libero
di trovare la mia strada

Contents

Abstract

This thesis investigates statistical inference for temporal point processes, focusing on advanced bootstrap methods tailored to the complexities of dependent event data. Point processes are essential for modeling phenomena characterized by the occurrence of discrete events over time. A particular emphasis is placed on the Epidemic-Type Aftershock Sequence (ETAS) model, a special case of marked Hawkes processes, widely used in seismology.

Traditional bootstrap techniques, commonly employed in statistical inference, assume independent and identically distributed observations, a condition that is violated in the context of point processes, where events are typically dependent and driven by a stochastic intensity function. To address this limitation, the thesis explores two advanced bootstrap procedures: the Fixed Intensity Bootstrap (FIB) and the Recursive Intensity Bootstrap (RIB), as developed by Cavaliere et al. (2023). The FIB assumes a fixed intensity over bootstrap replications, while the RIB updates the intensity recursively based on each newly simulated event. Both methods are complemented by their non-parametric variants, which allow for greater flexibility when model misspecification is suspected.

These methodologies are then applied to a real dataset of seismic events in L'Aquila, Italy, recorded during 2009, which includes over 3,000 elements. The ETAS model is calibrated using maximum likelihood, and bootstrap replications are generated to construct confidence intervals for key parameters. Comparative analysis of the FIB and RIB outputs reveals differences in the width and positioning of the resulting intervals, with RIB generally producing wider estimates due to the additional variability introduced by its recursive structure.

Chapter 1

Point Processes

In this chapter we introduce the mathematical foundation of point processes, a key tool in modeling events occurring randomly over time or space. We begin with classical probability models like the Binomial and Poisson distributions, and progressively build toward more complex and flexible formulations such as homogeneous and non-homogeneous Poisson point processes, Hawkes processes and the ETAS (Epidemic-Type Aftershock Sequence) model, which is widely used in seismology.

1.1 Binomial Distribution

Imagine rolling a dice n times and having two possible outcomes for every roll: even or odd. We call p the probability of obtaining an even result. If we want to calculate the probability P that the result is even x times out of n we obtain:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

We say that the dice has a **Binomial distribution** $\mathcal{B}(n, p)$.

$\binom{n}{x}$ is the **Binomial coefficient** and represent the number of combinations of x elements from a set of n independent elements and is calculated as:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

The mean and variance of the Binomial distribution are equal to:

$$E[X] = \sum_{i=1}^{n} E[A_i] = nE[A] = np.$$

$$Var[X] = \sum_{i=1}^{n} Var[A_i] = nVar[A] = np(1-p).$$

Where $A_1, A_2, \ldots, A_n$ are all the independent events (the rolls of the dice).

1.2 Poisson Distribution

*The **Poisson distribution** is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time, provided that these events happen with a known average rate λ and are independent of the time elapsed since the last event.*

1.2.1 Poisson Limit Theorem

The **Poisson Limit Theorem** asserts that:
Having a Binomial distribution $Y \sim \mathcal{B}(n, p)$ with the parameters $n \to \infty$ and $p \to 0$ such that $np = \lambda$ remains constant, then:

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y} = \frac{e^{-\lambda} \lambda^y}{y!}.$$

Which is the formula of the Poisson distribution. We can prove the Poisson Limit Theorem by following the following steps:

First, we use the Stirlings approximation to approximate n! and (n-y)! and we substitute p with $\frac{\lambda}{n}$:

$$\lim_{n \to \infty} \frac{1}{y!} \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{\sqrt{2\pi (n - y)} \left(\frac{n - y}{e}\right)^{n - y}} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n - y} =$$

$$= \lim_{n \to \infty} \frac{1}{y!} \frac{\sqrt{n}}{\sqrt{n - y}} \left(\frac{n}{e}\right)^n \frac{\left(\frac{n - y}{e}\right)^y}{\left(\frac{n - y}{e}\right)^n} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n - y} =$$

$$= \lim_{n \to \infty} \frac{1}{y!} \frac{1}{\sqrt{1 - \frac{y}{n}}} n^n \frac{(n - y)^y}{(n - y)^n} e^{-y} \frac{\lambda^y}{n^y} \left(1 - \frac{\lambda}{n}\right)^{n - y} =$$

$$= \lim_{n \to \infty} \frac{1}{y!} \frac{1}{\sqrt{1 - \frac{y}{n}}} \left(\frac{n}{n - y}\right)^{n - y} e^{-y} \lambda^y \left(1 - \frac{\lambda}{n}\right)^{n - y}.$$

The limit $\lim_{n \to \infty} \left(\frac{n}{n - y}\right)^{n - y}$ is the Taylor expansion of e^y. So, we get:

$$\lim_{n \to \infty} \frac{1}{y!} \frac{1}{\sqrt{1 - \frac{y}{n}}} \lambda^y \left(1 - \frac{\lambda}{n}\right)^{n - y}.$$

Secondly, we approximate $\left(1 - \frac{\lambda}{n}\right)^{n - y}$:

$$\ln\left(\left(1 - \frac{\lambda}{n}\right)^{n - y}\right) = (n - y) \ln\left(1 - \frac{\lambda}{n}\right).$$

Since $\frac{\lambda}{n} \to 0$, we use the first-order logarithm approximation:

$$\ln\left(\left(1 - \frac{\lambda}{n}\right)^{n - y}\right) \approx -\frac{\lambda}{n} (n - y).$$

$$\left(1 - \frac{\lambda}{n}\right)^{n - y} \approx e^{-\frac{\lambda}{n} (n - y)}.$$

Finally, we replace the obtained value in our function:

$$\lim_{n \to \infty} \frac{1}{y!} \frac{1}{\sqrt{1 - \frac{y}{n}}} \lambda^y e^{-\frac{\lambda}{n} (n - y)} =$$

$$= \lim_{n \to \infty} \frac{1}{y!} \frac{1}{\sqrt{1 - \frac{y}{n}}} \lambda^y e^{-\lambda \left(1 - \frac{y}{n}\right)} = \frac{\lambda^y e^{-\lambda}}{y!}.$$

1.2.2 Mean and Variance of Poisson Distribution

A particular characteristic of the Poisson distribution is that its mean and variance coincide:

$$E[Y] = \sum_{y = 0}^{\infty} y P(Y = y) = \sum_{y = 0}^{\infty} y \frac{e^{-\lambda} \lambda^y}{y!}.$$

Since $y \frac{e^{-\lambda} \lambda^y}{y!} = 0, \ y = 0$ we can rewrite $E[Y]$ as:

$$E[Y] = \sum_{y = 1}^{\infty} y \frac{e^{-\lambda} \lambda^y}{y!} = \sum_{y = 1}^{\infty} \lambda \frac{\lambda^{y - 1} e^{-\lambda}}{(y - 1)!} =$$

$$= \lambda e^{-\lambda} \sum_{j = 0}^{\infty} \frac{\lambda^j}{j!}, \quad j = y - 1.$$

The summation $\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$ is the Taylor series expansion for e^{λ}. So, we get:

$$E[Y] = \lambda \frac{e^{\lambda}}{e^{\lambda}} = \lambda.$$

$$Var(Y) = E\left[Y^2\right] - (E[Y])^2.$$

The second moment $E\left[Y^2\right]$ is calculated as:

$$E\left[Y^2\right] = E\left[Y^2 + Y - Y\right] = E[Y(Y-1) + Y].$$

Considering the linearity of the function $E\left[Y^2\right]$, we can rewrite it as:

$$E\left[Y^2\right] = E[Y] + E[Y(Y-1)].$$

$$E[Y(Y-1)] = \sum_{y=0}^{\infty} y(y-1) P(Y=y) = \sum_{y=2}^{\infty} y(y-1) \frac{e^{-\lambda}\lambda^y}{y!} =$$

$$= \lambda^2 e^{-\lambda} \sum_{y=2}^{\infty} \frac{\lambda^{y-2}}{(y-2)!} = \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!},\ j = y-2.$$

$$E[Y(Y-1)] = \lambda^2 \frac{e^{\lambda}}{e^{\lambda}} = \lambda^2.$$

$$E\left[Y^2\right] = \lambda + \lambda^2.$$

We substitute the values into the variance formula:

$$Var(Y) = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

1.3 Homogeneous Poisson Point Processes

*A **counting process** $\{N(t),\ t \geq 0\}$ must satisfy the following requirements:*

1. $N(t) \geq 0$
2. *$N(t)$ is integer valued.*
3. *If $t_i < t_j,\ \forall i, j \in \mathbb{N}$, then $N(t_i) \leq N(t_j)$.*
4. *For $t_i < t_j$, $N((t_i,\ t_j])$ counts the number of occurrences in the interval $(t_i,\ t_j]$.*

*A counting process is said to be a **Poisson Point Process** (PPP) with intensity $\lambda > 0$ if it satisfies the following requirements:*

1. $N(0) = 0$
2. *For every distinct time points $t_j < t_{j+1}$, $j = 0,\ 1,\ 2,\ \ldots$, the random variables $N((j,\ j+1])$ are independent.*
3. *If $t_i < t_j$, the number of events in any interval of length $t_j - t_i$ is Poisson distributed with mean $\lambda(t_j - t_i)$:*

$$P\{N((t_i,\ t_j]) = k\} = \frac{(\lambda(t_j - t_i))^k\ e^{-\lambda(t_j - t_i)}}{k!},\ k = 0,\ 1,\ \ldots$$

We denote the state space in which the points occur with $\mathcal{S}$, $\mathcal{S} \subseteq \mathbb{R}^m$, $m \geq 1$. This is not always true, but for our purposes we will take into consideration only this definition. With $\xi = (n,\ \{x_1,\ \ldots,\ x_n\})$ we indicate the unordered realizations of a PPP on a subset $\mathcal{R}$ of $\mathcal{S}$. $\xi = (0,\ \{\})$, $n = 0$. The event space consists of all possible finite subsets of R:

$$\varepsilon(\mathcal{R}) = \{(0,\ \{\})\} \bigcup_{n=1}^{\infty} \{(n,\ \{x_j\}) :\ x_j \in \mathcal{R},\ j = 1,\ \ldots,\ n\}.$$

A PPP $X(t)$ is homogeneous when $\lambda(s) \equiv \alpha, s \in \mathcal{S}$ for some constant $\alpha \geq 0$. If $\lambda(s) > 0$ then the mean and variance of $X(t)$ are:

$$E[X(t)] = \lambda(s)t.$$

$$Var[X(t)] = \lambda(s)t.$$

1.3.1 Memoryless property of PPP

We denote with $P(X > t_j)$ the probability that an event associated with X has not occurred by time t_j and with $P(X > t_i + t_j \mid X > t_i)$ the probability that an event associated with X has not occurred by time $t_i + t_j$, given that it has not occurred by time t_i:

$$P(X > t_j) = 1 - P(X < t_j) = 1 - \int_0^{t_j} \lambda e^{-\lambda x} dx = 1 - (1 - e^{-\lambda t_j}) = e^{-\lambda t_j}.$$

$$P(X > t_i + t_i \mid X > t_i) = \frac{P(X > t_i + t_j \cap X > t_i)}{P(X > t_i)} = \frac{P(X > t_i + t_j)}{P(X > t_i)} = \frac{1 - P(X < t_i + t_j)}{1 - P(X < t_i)} =$$

$$= 1 - \frac{\int_0^{t_i + t_j} \lambda e^{-\lambda x} dx}{\int_0^{t_i} \lambda e^{-\lambda x} dx} = \frac{1 - (1 - e^{-\lambda(t_i + t_j)})}{1 - (1 - e^{-\lambda t_i})} = \frac{e^{-\lambda t_i} e^{-\lambda t_j}}{e^{-\lambda t_i}} = e^{-\lambda t_j}.$$

As we can see the PPP has no memory, in other words, no matter how much time s has passed, the probability of observing a new event after an additional time t_j is always $e^{-\lambda t_j}$

1.4 Non-homogeneous Poisson Point Processes

Events in Homogeneous Poisson Point Processes happen with a fixed intensity λ. In order to describe more elaborate processes we need to calculate an intensity $\lambda(t \mid \mathcal{H}_t)$ that depends on the time and the history of the process up to time t.

*A **Non-Homogeneous Poisson Point Process** (NHPPP) is a generalization of the Homogeneous Poisson Point Process (HPPP), where the event intensity $\lambda(t)$ is no longer constant but varies over time or space.*
Given a process $N(t)$ and assuming the existence of the limit, the **conditional intensity function** is defined as:

$$\lambda(t \mid \mathcal{H}_t) = \lim_{h \to 0} \frac{P(N(t + h) - N(t) = 1 \mid \mathcal{H}_t)}{h}.$$

Where $\mathcal{H}_t$ represents the history of the process up to time t. In an NHPPP, the number of events in $\mathcal{R}$ follows a Poisson distribution with mean:

$$\Lambda(\mathcal{R}) = E[N(\mathcal{R})] = \int_{\mathcal{R}} \lambda(s) ds.$$

1.4.1 Likelihood Function

The conditional **cumulative distribution function** (cdf) of t given $\mathcal{H}_{\bar{t}}$, $F(t \mid \mathcal{H}_{\bar{t}})$ for $t \geq \bar{t}$ is defined as the probability that the next event will occur within time t, given the history of the process up to time $\bar{t}$:

$$F(t \mid \mathcal{H}_{\bar{t}}) = 1 - P(N(t) - N(\bar{t}) = 0) = 1 - \frac{e^{-\Lambda(\bar{t}, t)} \Lambda(\bar{t}, t)^0}{0!} = 1 - e^{-\int_{\bar{t}}^{t} \lambda(x) dx}.$$

We define the conditional **probability density function** (pdf) given $\mathcal{H}_{\bar{t}}$, $f(t \mid \mathcal{H}_{\bar{t}})$ as the first derivative of the conditional cdf:

$$f(t \mid \mathcal{H}_{\bar{t}}) = \frac{d}{dt} F(t \mid \mathcal{H}_{\bar{t}}).$$

$$f(t \mid \mathcal{H}_{\bar{t}}) = \frac{d}{dt} [1 - e^{-\int_{\bar{t}}^{t} \lambda(x) dx}].$$

$$f(t \mid \mathcal{H}_{\bar{t}}) = -\frac{d}{dt} e^{-\int_{\bar{t}}^{t} \lambda(x) dx}.$$

$$f(t \mid \mathcal{H}_{\bar{t}}) = -e^{-\int_{\bar{t}}^{t} \lambda(x) dx} (-\lambda(t)) = \lambda(t) e^{-\int_{\bar{t}}^{t} \lambda(x) dx}.$$

The **likelihood function** serves to measure how strongly the observed data, considered fixed, support the possible values of the unknown parameter λ, that are adjusted:

$$\mathcal{L}(\lambda(\cdot)) = \prod_{i=1}^{n_T} f(t_i \mid \mathcal{H}_{t_i - 1})(1 - F(T \mid \mathcal{H}_{t_n})) =$$

$$= \prod_{i=1}^{n_T} \lambda(t_i) e^{-\int_{t_{i-1}}^{t_i} \lambda(x)dx} e^{-\int_{t_n}^{T} \lambda(x)dx} = \prod_{i=1}^{n_T} \lambda(t_i) e^{-\int_{0}^{T} \lambda(x)dx}. \tag{1.1}$$

Where the product between the conditional pdfs indicates the probability of observing the data and the complement of the conditional cdf denotes the probability of not observing further events after the last one.

1.5 Hawkes Processes

Now consider a class of point processes called **Hawkes processes**. The intensity in Hawkes processes is dependent on past events and is defined as follows:

$$\lambda(t \mid \mathcal{H}_t) = \lambda_0 + \sum_{t_i < t} g(t - t_i). \tag{1.2}$$

Where λ_0 denotes the **base intensity**, which is independent of events, and $g(t - t_i)$ indicates the **memory kernel**, which express how past events influence the value of the intensity at time t.

1.5.1 Memory kernel

When the function $g(\cdot)$ is positive, the process is said to be self-exciting. The two most common types of memory kernel function are:

- Exponential decay:
$$g(t) = \alpha e^{-\beta t}.$$
Where $\alpha \geq 0, \beta > 0, \alpha < \beta$. α controls the strength of excitation and β determines how quickly the influence of past events diminishes over time.

- Power-law decay:
$$g(t) = \frac{\alpha}{(t + \delta)^{\beta+1}}.$$
Where $\alpha \geq 0, \delta > 0, \beta > 0, \alpha < \beta^{\delta^{\beta}}$. α controls the strength of excitation, δ is a regularization parameter that ensures the function is well defined avoiding singularities at $t = 0$ and β determines how quickly the influence of past events diminishes over time.

Both functions are structured in such a way that the intensity is most affected by the latest events.

1.5.2 Cluster Process Representation of Hawkes Processes

In 1949, Kendall introduced an age-dependent birth and death process in which $N(x, t)$ specifies the age distribution of the population. $\int_{x_1}^{x_2} N(x, t)$ denotes the number of individuals in the age range (x_1, x_2). Kendall makes five assumptions in order to define the model:

1. *The sub-populations generated by two co-existing individuals develop in complete independence of one another.*

2. *For any individual of age x there is a probability*
$$\lambda(x)dt + o(dt)$$
of giving birth to another individual during the subsequent time interval of length dt. Extending the probability to the whole population, we obtain:
$$\Lambda = \int_0^{\infty} \lambda(x) dN(x, t).$$
It follows that the probability of having exactly r birth in the small time interval $(t, t + dt)$ is P_r:
$$P_0 = 1 - \Lambda dt + o(dt).$$
$$P_1 = \Lambda dt + o(dt).$$
$$P_r = o(dt).$$
Thus r is asymptotically a Poisson variable, its mean and variance being equal (to the first order in dt).

3. *The birth rate $\lambda(x)$ varies in any manner with the age x of the parent, but is independent of the epoch t.*

4. *For any individual of age x there is a probability*

$$\mu(x)dt + o(dt)$$

of dying during the subsequent time interval of length dt.

5. *The death rate $\mu(x)$ varies in any manner with the age x, but is independent of the epoch t.*

Now consider the death rate $\mu(x)$ equal to 0 and allow immigration at a rate λ_0 (on arrival, immigrants are considered to have age 0). We can now write the intensity function of the process, that is equal to (1.2), where t_i represents the moments when births or immigrations occur. In this cluster process, the cluster centers $N_C(t)$, also called immigrants, follow a Poisson process of intensity $\lambda_o(t)$. Each occurrence of $N_C(t)$ initiates a cluster of related events, also called offspring, consisting of all descendant events originating from the initial immigrants. The first generation of offspring follows a Poisson process with a rate determined by $g(t)$. Each offspring could generate further offspring, leading to a generalized branching process. We now define the **branching factor** ν, which denotes the expected number of offspring per event, and the **expected cluster size** $E[N_\infty]$:

$$\nu = \int_0^\infty g(t)dt.$$

$$N_i = N_{i-1}\nu = N_{i-2}\nu^2 = \cdots = N_0\nu^i, \quad N_0 = 1.$$

$$E[N_\infty] = \sum_{i=0}^{\infty} N_i = \sum_{i=0}^{\infty} \nu^i = \frac{1}{1-\nu}.$$

If $\nu < 1$ we define the process as **subcritical**. The number of events in every cluster is finite and the probability of extinction is 1. If, instead, $\nu > 1$ we define the process as **supercritical**. The number of events in every cluster may grow indefinitely, the probability of extinction is still nonzero but the process has a positive probability to survive perpetually.

1.6 ETAS Model

In 1988, Ogata proposed the **Epidemic-Type Aftershock Sequence** (ETAS) model in order to forecast and analyze the chances of earthquake occurrences. While in previous models there was the necessity to distinguish between triggering and derived events, in the ETAS model there is no need to make such distinction, since all events have a chance to generate their own offspring. As we can see, this is the same behavior of the birth process defined in section 1.5.2. Extending the conditional intensity rate (1.2) to a multivariate point process t_i^m, we obtain:

$$\lambda_j(t \mid \mathcal{H}_t) = \lambda_{0j}(t) + \sum_m \sum_{t_i^m < t} g_{jm}(t - t_i^m).$$

$$\lambda_j(t \mid \mathcal{H}_t) = \lambda_{0j}(t) + \sum_m \int_0^t g_{jm}(t-s)dN_m(s).$$

For the discrete magnitude values of j and m. Now let $g_{jm}(t) = \kappa(m_i)g_j(t)$ and consider the aggregate process $N(t) = \sum_m N_m(t)$ resulting from the combination of the individual point process components, then the conditional intensity of the ETAS model $\lambda(t \mid \mathcal{H}_t) = \sum_j \lambda_j(t \mid \mathcal{H}_t)$ is given by:

$$\lambda(t \mid \mathcal{H}_t) = \lambda_0 + \sum_{t_i < t} \kappa(m_i)g(t - t_i). \tag{1.3}$$

Where t_i is the occurrence time of the aggregate process $N(t)$, m_i is the corresponding magnitude of t_i, $\kappa(m) = e^{\alpha(m-m_0)}, m \geq m_0$ with m_0 being the magnitude threshold and α measuring the effect of magnitude in the production of descendants and $g(t) = \sum_j g_j(t)$.

1.6.1 ETAS Model's memory kernel

In 1894, Omori analyzed the aftershocks of the 1891 M_s8.0 Nobi earthquake and derived a formula that describes how the frequency of aftershocks decreases over time, following a mainshock:

$$n(t) = K(t+c)^{-1}.$$

Where t is the time passed since the occurrence of the mainshock and K and c are constants. We call this the **Omori formula**. In 1961, Utsu understood that the decay of the aftershock numbers could vary, and demonstrated that:

$$n(t) = K(t+c)^{-p}. \tag{1.4}$$

produces better fitting results. We call this the **Omori-Utsu formula** or modified Omori formula. In 1995, Utsu et al. analyzed p-values from over 200 aftershock sequences, finding that they range from 0.6 to 2.5, with a median of 1.1. They observed no clear correlation between p-value estimates and mainshock magnitudes. We will use (1.4) as the memory kernel $g(t)$ for our model.

1.6.2 Maximum Likelihood Estimator

We define $\theta \in \Theta \subseteq R^d, \quad d := dim\theta$ as a finite-dimensional vector of unknown parameters. For the ETAS Model $\theta = (\lambda_0, K, \alpha, c, p)$. Let θ_0 denote the **true value** of the parameter vector. Applying θ in the Likelihood function (1.1), we obtain:

$$\mathcal{L}(\lambda(\cdot);\theta) = \prod_{i=1}^{n_T} \lambda(t_i;\theta) e^{-\int_0^T \lambda(x;\theta)dx}. \tag{1.5}$$

We now consider the logarithmic form of the likelihood function in (1.6):

$$\ell_T(\theta) = \sum_{i=1}^{n_T} \log \lambda(t_i;\theta) - \int_0^T \lambda(x;\theta)dx. \tag{1.6}$$

The **maximum likelihood estimator** (MLE) $\hat{\theta}_T$ is given by the parameter value that maximizes the log-likelihood function:

$$\hat{\theta}_T := \arg\max_{\theta\in\Theta} \ell_T(\theta). \tag{1.7}$$

Under suitable regularity conditions, both the unrestricted MLE $\hat{\theta}_T$ and the restricted MLE $\tilde{\theta}_T$ under the null hypothesis consistently estimate the true parameter vector θ_0: that is, as the sample size increases, $\hat{\theta}_T$ and $\tilde{\theta}_T$ converge in probability to θ_0.

Chapter 2

Bootstrap Inference For Point Processes

Traditional bootstrap methods, commonly used in classical statistical settings, rely on the assumption of independent and identically distributed observations. However, such assumptions do not hold for point processes, where the timing and structure of events are inherently dependent and governed by a potentially time-varying intensity function. As a result, classical bootstrap techniques are inadequate in this context. To address these limitations, specialized bootstrap methods have been developed for point processes. In this chapter, we focus on two such approaches: the Fixed Intensity Bootstrap (FIB) and the Recursive Intensity Bootstrap (RIB).

2.1 The Bootstrap

Cavaliere et al. (2023) introduce and compare two bootstrap procedures. The first, termed the **fixed intensity bootstrap** (FIB), is a novel method whose most important feature is that the bootstrap intensity function remains constant across all repetitions. The second approach, referred to as the **recursive intensity bootstrap** (RIB), has been previously applied in the literature. The RIB incorporates a recursive structure that renders the bootstrap intensity a stochastic process, conditioned on the observed data.

2.1.1 Waiting Times

We denote with w_i the **waiting time** between two consecutive events, defined as:

$$w_i = t_i - t_{i-1}.$$

Cavaliere et al. (2023) exploit the fact that, by definition, the transformation the original event times $\{t_i\}$ using the integrated intensity function at the true value θ_0 standardizes the process: the resulting times $\{s_i\}$ form a homogeneous Poisson process with unit intensity. We can then transform the original waiting times $\{w_i\}$ into new, i.i.d. $\mathcal{E}(1)$ waiting times $\{v_i\}$, defined as:

$$v_i = s_i - s_{i-1}.$$

We obtain the transformed event times $\{s_i\}$ by:

$$s_i(\theta_0) := \Lambda(t_i; \theta_0)$$

consequently, the transformed waiting times are given by:

$$v_i(\theta_0) = s_i(\theta_0) - s_{i-1}(\theta_0) = \Lambda(t_i, t_{i-1}; \theta_0).$$

For the ETAS model in (1.3), we obtain:

$$v_i(\theta_0) = \int_{t_{i-1}}^{t_i} (\lambda_0 + \sum_{t_j<t} \kappa(m_j) g(t - t_j)) dt = \lambda_0 w_i + \int_0^{w_i} (\sum_{j<i} \kappa(m_j) g(w + t_{i-1} + t_j)) dw =$$

$$= \lambda_0 w_i + \sum_{j<i} \kappa(m_j) (\frac{K}{(t_i - t_j + c)^{p-1}(1-p)} - \frac{K}{(t_{i-1} - t_j + c)^{p-1}(1-p)}) =$$

$$= \lambda_0 w_i + \frac{K}{p-1} \sum_{j<i} \kappa(m_j)((t_{i-1} - t_j + c)^{1-p} - (t_i - t_j + c)^{1-p}). \tag{2.1}$$

The reverse time transformation $s_i \mapsto t_i$ plays a crucial role in the bootstrap implementation. Under the true model, we can numerically invert the mapping to generate the i-th waiting time w_i recursively. The recursion begins by generating the first waiting time $w_1(\theta_0) = \Lambda(v_1; \theta_0)$, where v_1 is the first transformed waiting time.

2.1.2 Fixed Intensity Bootstrap

Let $\{t_i\}_{i=1}^{n_T}$ denote the sample of event times observed in $[0, T]$, we fix a bootstrap true value parameter θ_T^*, which can be set to the unrestricted MLE $\hat{\theta}_T$, or to the restricted MLE $\tilde{\theta}_T$ under the null hypothesis. Under the FIB, the intensity process remains fixed across bootstrap replicates. Define the FIB intensity

$$\hat{\lambda}(t) := \lambda(t; \theta_T^*), \tag{2.2}$$

and its integrated form as

$$\hat{\Lambda}(t) := \int_0^t \lambda(u; \theta_T^*)\, du. \tag{2.3}$$

Both $\hat{\lambda}(t)$ and $\hat{\Lambda}(t)$ depend on the observed data $\{t_i\}_{i=1}^{n_T}$ and the fixed parameter θ_T^*. Hence, given the data, they are are known and fixed.

FIB Algorithm

(i) Generate a sample of i.i.d. bootstrap transformed waiting times $\{v_j^*\}$ from the $\mathcal{E}(1)$ distribution, conditionally on the observed data. The bootstrap transformed event times are given by $\{s_i^*\}$, where

$$s_i^* = \sum_{j=1}^{i} v_j^*.$$

(ii) Construct the bootstrap event times in the original time scale:

$$t_i^* = \hat{\Lambda}^{-1}(s_i^*), \quad i = 1, \ldots, n_T^*,$$

where $\hat{\Lambda}$ is defined as in (2.3) and

$$n_T^* = \max\{k : s_k^* \leq \hat{\Lambda}(T)\} = \max\{k : t_k^* \leq T\}.$$

The associated bootstrap counting process is defined as:

$$N^*(t) := \sum_{i \geq 1} \mathbb{I}(t_i^* \leq t), \quad t \in [0, T].$$

(iii) Estimate the bootstrap MLE:

$$\hat{\theta}_T^* := \arg\max_{\theta \in \Theta} \ell_T^*(\theta),$$

where the bootstrap log-likelihood is:

$$\ell_T^*(\theta) := \int_0^T \log \lambda(t; \theta) dN^*(t) - \int_0^T \lambda(t; \theta) dt = \sum_{i=1}^{n_T^*} \log \lambda(t_i^*; \theta) - \int_0^T \lambda(t; \theta) dt. \tag{2.4}$$

For implementation details and theoretical justification, see Remark 3.1 in Cavaliere et al. (2023), which outlines additional considerations on the FIB.

2.1.3 Recursive Intensity Bootstrap

Unlike the FIB, the Recursive Intensity Bootstrap (RIB) has a conditional intensity, denoted by $\lambda^*(t;\theta)$, based on the functional form of the original intensity $\lambda(t;\theta)$ evaluated on recursively generated bootstrap event times t_i^*. As a result, for any $\theta \in \Theta$, $\lambda^*(t;\theta)$ is a random process, even being conditioned on the original data. Therefore, the RIB intensity differs from the FIB one, that is fixed across bootstrap replications. Notably, $\lambda^*(t;\theta)$ retains the same properties, such as differentiability with respect to θ, inherited from $\lambda(t;\theta)$. Define the RIB intensity

$$\lambda^*(t) := \lambda^*(t;\theta_T^*), \tag{2.5}$$

and its integrated form as

$$\Lambda^*(t) := \int_0^t \lambda^*(u;\theta_T^*)\, du \tag{2.6}$$

RIB Algorithm

(i) As in Algorithm 1.

(ii) For $i = 1, \ldots, n_T^*$, recursively construct the bootstrap event times t_i^* in the original time scale as:

$$t_i^* = (\Lambda^*)^{-1}(s_i^*),$$

where $\Lambda^*(t)$ is defined as in (2.6) and

$$n_T^* := \max\{k : s_k^* \le \Lambda^*(T)\} = \max\{k : t_k^* \le T\}.$$

The associated bootstrap counting process is defined as:

$$N^*(t) := \sum_{i \ge 1} \mathbb{I}(t_i^* \le t), \quad t \in [0,T].$$

(iii) Estimate the bootstrap MLE:

$$\hat{\theta}_T^* = \arg\max_{\theta \in \Theta} \ell_T^*(\theta),$$

where the bootstrap log-likelihood is:

$$\ell_T^*(\theta) := \int_0^T \log \lambda^*(t;\theta)\, dN^*(t) - \Lambda^*(T;\theta) = \sum_{i=1}^{n_T^*} \log \lambda^*(t_i^*;\theta) - \int_0^T \lambda^*(t;\theta)\, dt.$$

For implementation details and theoretical justification, see Remark 3.2 in Cavaliere et al. (2023), which outlines additional considerations on the RIB.

2.1.4 Non-parametric RIB and FIB

In cases where the model is misspecified, the transformed waiting times $\hat{v}_i$ may not follow an exponential distribution (asymptotically). Thus, we consider a nonparametric alternative based on residual-based i.i.d. bootstrapping, similar to techniques used in discrete-time time series analysis. After fitting the point process model to the data, the residuals, interpreted as transformed scale waiting times $\hat{v}_i$, $i = 1, \ldots, n_T$, can be treated as an i.i.d. sample. This forms the basis of the **non-parametric bootstrap**, applicable to both FIB and RIB schemes. For conditional mean and variance time series models, residuals are typically centered and/or scaled before bootstrapping. Similarly, in our case, the waiting times $\hat{v}_i$ must be standardized so that the bootstrap times v_i^* at least match the mean of the $\mathcal{E}(1)$ distribution. This is accomplished by defining the rescaled residuals:

$$\hat{v}_i^c := \frac{\hat{v}_i}{\bar{v}_T}, \quad i = 1, \ldots, n_T, \tag{2.7}$$

where $\bar{v}_T := n_T^{-1} \sum_{j=1}^{n_T} \hat{v}_j$. Note that $\hat{v}_i^c > 0$ for all i, and drawing randomly from the set $\{\hat{v}_i^c\}_{i=1}^{n_T}$ yields values that have, conditionally on the observed data, mean $\mathbb{E}^*(v_i^*) = n_T^{-1} \sum_{i=1}^{n_T} \hat{v}_i^c = 1$.

Non-parametric bootstrap Algorithm

(i) Draw a sample $\{v_i^*\}$ of bootstrap transformed waiting times by resampling with replacement from $\{\hat{v}_i^c\}_{i=1}^{n_T}$, such that:

$$v_i^* = \hat{v}_{u_i^*}^c, \quad \text{for } i = 1, 2, \ldots, \tag{2.8}$$

where $\{u_i^*\}$ is an i.i.d. discrete uniformly distributed sequence over $\{1, \ldots, n_T\}$. The corresponding bootstrap transformed event times are given by $s_i^* = \sum_{j=1}^{i} v_j^*$.

(ii) Proceed with steps (ii)–(iii) of FIB Algorithm or RIB Algorithm depending on whether a fixed or recursive intensity bootstrap is employed.

For implementation details and theoretical justification, see Remark 5.1 in Cavaliere et al. (2023), which outlines additional considerations on the Non-parametric bootstrap.

Chapter 3

Implementation and Results

In this chapter, we present the practical implementation of the ETAS (Epidemic-Type Aftershock Sequence) model and apply bootstrap procedures to real seismic event data. We describe the fitting process for both the model and the bootstrap algorithms. Our focus is on two parametric bootstrap methods introduced earlier: the Fixed Intensity Bootstrap (FIB) and the Recursive Intensity Bootstrap (RIB). These approaches are implemented and evaluated with respect to the characteristics of the confidence intervals they produce.

3.1 Data description

The dataset comprises information on seismic events in the province of L'Aquila between 01/01/2009 and 31/12/2009. For each event, it provides details such as the geographical coordinates of the epicenter (latitude and longitude), the magnitude of the event, and the time of occurrence.

Dataset characteristics	
Time period	01/01/2009 - 31/12/2009
Longitude interval considered	13.114 - 14.072
Latitude interval considered	41.751 - 42.586
Magnitude interval considered	2 - 6.1
Total number of records	3229

Table 3.1: Dataset Characteristics

The data can be visualized on a map (Figure 3.1), and a plot of event times versus magnitudes (Figure 3.2) can be generated. The latter graph illustrates the progression of the seismic sequence that occurred in L'Aquila in 2009.

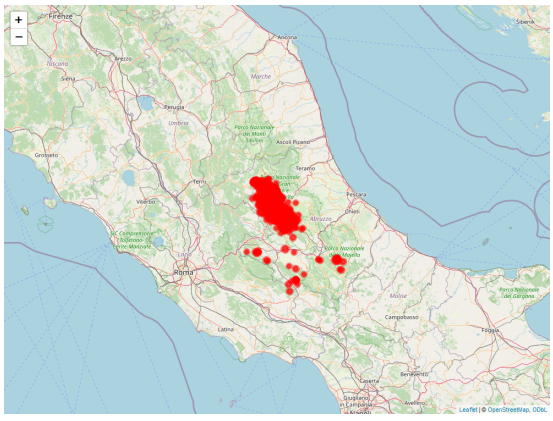

Figure 3.1: Seismic Events Map

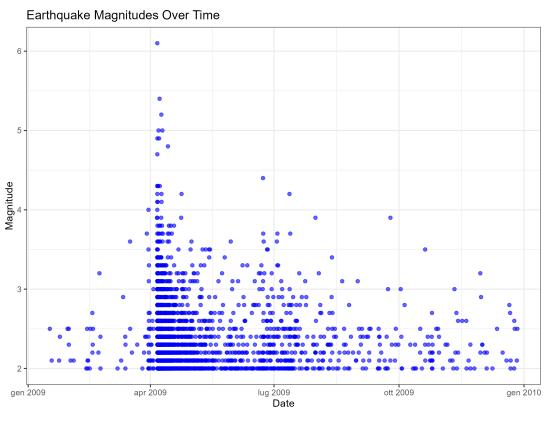

Figure 3.2: Earthquake Magnitudes Over Time

3.2 Data Preparation and ETAS Model Implementation

A new dataset named `Data_for_ETAS` is first created, consisting of two columns. The first column, `ts`, contains the numeric values representing the time difference between each event's date and time and a fixed reference point, defined as 01/01/2009 at 00:00:00. The second column, `magnitudes`, stores the magnitude associated with each seismic event.

3.2.1 `fit_etas` function

```
fit_etas <- function(data, m0) {
    data <- data[order(data$ts), ]

    T_max <- max(data$ts)

    init_par <- c(lambda_0 = 0.01, K = 0.1, c = 0.01, p = 1.1, alpha = 1.0)

    fit <- optim(par = init_par,
           fn = neg_loglik,
           data = data,
           T_max = T_max,
           m0 = m0,
           method = "L-BFGS-B",
           lower = c(1e-5, 1e-5, 1e-5, 1.01, 0),
           upper = c(10, 10, 1, 5, 10))

    return(list(par = fit$par, loglik = -fit$value))
}
```

This function fits the ETAS Model to a given dataset of seismic events using maximum likelihood estimation. The function takes as input a dataset and a magnitude threshold `m0`. Firstly the dataset is sorted chronologically to ensure temporal consistency. The observation window `T_max` is defined as the time of the latest event in the dataset. Initial guesses for the unknown parameters λ_0, K, c, p and α are specified in the vector `init_par`, and box constraints are set through the `lower` and `upper` bounds to ensure numerical stability and adherence to theoretical restrictions. The `optim()` function is used with the `L-BFGS-B` method to minimize the negative log-likelihood. The function returns a list containing the values of the unknown parameters that correspond to the unrestricted MLE $\hat{\theta}_T$ and the value of the maximized log-likelihood.

3.3 FIB and RIB Implementation

3.3.1 `fib_bootstrap_ETAS`

```
fib_bootstrap_etas <- function(Data_for_ETAS, m0, n_boot) {
    theta_hat <- fit$par
    T_max <- max(Data_for_ETAS$ts)
    Lambda_vals <- compute_cumulative_intensity(Data_for_ETAS, theta_hat, m0)

    boot_params <- matrix(NA, nrow = n_boot, ncol = length(theta_hat))
    colnames(boot_params) <- c("lambda_0", "K", "c", "p", "alpha")

    for (b in 1:n_boot) {
        cat("\n- FIB BOOTSTRAP", b, "\n")
        data_star <- generate_fib_events(Lambda_vals, T_max)
        boot_params[b, ] <- fit_etas_fib(data_star, m0)
    }

return(boot_params)
}
```

The function performs the Fixed Intensity Bootstrap (FIB) algorithm. Firstly it assigns to `theta_hat` the ETAS parameters $\hat{\theta}_T$ calculated on the original dataset. Then the cumulative intensity $\Lambda(t)$ is computed using `compute_cumulative_intensity()`. For each of the `n_boot` bootstrap iterations the function generates a synthetic dataset using `generate_fib_events()`. The ETAS model is re-fitted to this synthetic dataset using `fit_etas_fib()` and the estimated parameters are stored in a matrix. The function returns a matrix of dimension `n_boot` $\times$ 5, containing the bootstrap estimates of the ETAS parameters `lambda_0, K, p, alpha`.

3.3.2 rib_bootstrap_ETAS

```
rib_bootstrap_etas <- function(Data_for_ETAS, m0, n_boot) {
    theta_hat <- fit$par
    T_max <- max(Data_for_ETAS$ts)

    boot_params <- matrix(NA, nrow = n_boot, ncol = length(theta_hat))
    colnames(boot_params) <- c("lambda_0", "K", "c", "p", "alpha")

    for (b in 1:n_boot) {
        cat("- RIB BOOTSTRAP", b, "\n")
        data_star <- generate_rib_events(theta_hat, T_max, m0)
        boot_params[b, ] <- fit_etas_rib(data_star, m0)
    }

return(boot_params)
}
```

The function performs the Recursive Intensity Bootstrap (RIB) for the ETAS model. Firstly it assigns to `theta_hat` the ETAS parameters $\hat{\theta}_T$ calculated on the original dataset. For each of the `n_boot` bootstrap iterations the function generates a synthetic dataset using `generate_rib_events()` with the times calculated recursively. The ETAS model is re-fitted to this synthetic dataset using `fit_etas_rib()` and the estimated parameters are stored in a matrix. The function returns a matrix of dimension `n_boot` $\times$ 5, containing the bootstrap estimates of the ETAS parameters `lambda_0, K, p, alpha`.

3.4 Results

Parameter	$\hat{\theta}_T$	**FIB**	**RIB**
λ_0	0.2218	[0.1881; 0.3177]	[0.1584; 0.3329]
K	0.0559	[0.0380; 0.1040]	[0.0397; 0.1965]
c	0.2206	[0.0157; 0.4514]	[0.1154; 0.7897]
p	1.4741	[1.2621; 1.6360]	[1.2748; 2.2107]
α	1.8598	[0.7074; 2.1667]	[1.5785; 2.1141]

Table 3.2: ETAS Parameter Estimates and Bootstrap Confidence Intervals

We compare the ETAS parameter estimates using confidence intervals obtained via the Fixed Intensity Bootstrap (FIB) and Recursive Intensity Bootstrap (RIB) algorithms and contrast them with the $\hat{\theta}_T$ point estimates. The results show significant differences in the width and position of the confidence intervals. In particular, the intervals derived from the FIB are generally narrower than those from the RIB, especially for parameters like K and c. The RIB method also results in wider intervals for p. This differences likely reflect the additional randomness introduced in the RIB procedure, where bootstrap times are recursively generated and thus more variable. Interestingly, FIB gives a much wider confidence interval for α, whereas RIB tightens around the $\hat{\theta}_T$ estimates. This suggests that RIB may be better at preserving the magnitude dependence seen in the original data.

Bibliography

[1] Giuseppe Cavaliere, Ye Lu, Anders Rahbek, and Jacob Strk-stergaard. Bootstrap inference for hawkes and general point processes. *Journal of Econometrics*, 235(1):133–165, 2023.

[2] Alan G. Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493503, 1974.

[3] David G Kendall. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):230–282, 1949.

[4] Yosihiko Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, (30):243–261, 1978.

[5] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[6] Sheldon M Ross. *Stochastic processes*. John Wiley & Sons, 1995.

[7] Roy L. Streit. *The Poisson Point Process*, pages 11–55. Springer US, Boston, MA, 2010.

[8] H.M. Taylor and S. Karlin. *An Introduction to Stochastic Modeling*. Academic Press, 2014.

[9] J Zhuang. Stochastic simulation of earthquake catalogs. *Community online resource for statistical seismicity analysis*, 2015.

[10] Jiancang Zhuang, David S Harte, Maximilian J Werner, Sebastian Hainzl, and Shiyong Zhou. Basic models of seismicity: Temporal models, August 2012.