Marked Poisson cluster processes and applications

Žugec, Petra

Doctoral thesis / Disertacija

2019

University of Zagreb

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

Petra Žugec

MARKED POISSON CLUSTER PROCESSES AND APPLICATIONS

DOCTORAL THESIS

Zagreb, 2019.

Sveučilište u Zagrebu

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Petra Žugec

OZNAČENI POISSONOVI PROCESI S KLASTERIMA I PRIMJENE

DOKTORSKI RAD

Zagreb, 2019.

University of Zagreb

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

PETRA ŽUGEC

MARKED POISSON CLUSTER PROCESSES AND APPLICATIONS

DOCTORAL THESIS

Supervisor:

prof. dr. sc. Bojan Basrak

Zagreb, 2019.

Sveučilište u Zagrebu

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

PETRA ŽUGEC

OZNAČENI POISSONOVI PROCESI S KLASTERIMA I PRIMJENE

DOKTORSKI RAD

Mentor:

prof. dr. sc. Bojan Basrak

Zagreb, 2019.

Mojoj obitelji: Nikoli, Matildi i Šimunu

Summary

The theory of point processes constitutes an important part of modern stochastic process theory and is widely recognized as a useful and elegant tool for modelling. Point processes are well understood models and have applications in a wide range of applied probability areas, especially in risk theory which is important for understanding non-life insurance mathematics. It deals with the modelling of claims and gives answers on premium amount. Elegant mathematical analysis of the classical Cramér - Lundberg risk model has an important place in non-life insurance theory. The theory yields precise or approximate computations of the ruin probabilities, appropriate reserves, distribution of the total claim amount and other properties of an idealized insurance portfolio. In recent years, some special models have been proposed to account for the possibility of clustering of some events, for instance the Hawkes processes. We study asymptotic distribution of the total claim amount in the setting where Cramer - Lundberg risk model is augmented with a marked Poisson cluster structure. Marked Hawkes processes are then a special case and have an important role as the key example in our analysis. To make this more precise, we model arrival of claims in an insurance portfolio by a marked point process, say

$$N = \sum_{k=1}^{\infty} \delta_{\tau_k, A^k}$$

where τ_k's are non-negative random variables representing arrival times with some degree of clustering and A^k's represent corresponding marks in a rather general metric space $\mathbb{M}$. For each marked event, the claim size can be calculated using a measurable mapping of marks to non-negative real numbers, $f : \mathbb{M} \to \mathbb{R}_{\geq 0}$ say. So that the total claim amount in the time interval $[0, t]$ can be calculated as

$$S(t) = \sum_{\tau_k \leq t} f(A^k).$$

We determine the effect of the clustering on the quantity $S(t)$, as $t \to \infty$, even in the case when the distribution of the individual claims does not satisfy assumptions of the classical central limit theorem. Besides new results regarding the case when second moments do not exist, we use different approach based on the limit theory for two dimensional random walks which stems from the classical Anscombe's theorem and not on martingale central limit theorem which was commonly used. We present the central limit theorem for the total claim amount $S(t)$ in our setting under appropriate second moment conditions and prove a functional limit theorem concerning the sums of regularly varying non-negative random variables when subordinated to an independent renewal process. Based on this, we prove the limit theorem for the total claim amount $S(t)$ in cases when individual claims have infinite variance. Moreover, we apply these results to three special models. In particular, we give a detailed analysis of the marked Hawkes processes which are extensively studied in recent years. In the last chapter we move our attention to the maximal claim size and present our results regarding limiting behaviour of maximum when claims belong to the maximum domain of attraction of one of the three extreme value distributions (Fréchet, Weibull and Gumbel). We also apply those results to three special models which we studied in previous chapter. Besides that, we try to clarify the notion of stochastic intensity which can be described in several different ways. The understanding of the stochastic intensity is important because of it's usage in the implicit definition of Hawkes processes.

Key words

Point process, Poisson cluster processes, limit theorems, Hawkes process, total claim amount, maximal claim size

Sažetak

Teorija točkovnih procesa utemeljuje važan dio moderne teorije stohastičkih procesa i široko je prepoznata kao koristan i elegantan alat za modeliranje. Točkovni procesi su model koji se dobro razumije i koriste u mnogim područjima primijenjene vjerojatnosti, posebno u teoriji rizika (matematika neživotnih osiguranja). Teorija rizika se bavi modeliranjem zahtjeva za isplatom u svrhu određivanja visine premije. Elegantna matematička analiza Cramér-Lundbergovog modela rizika ima važnu ulogu u teoriji neživotnih osiguranja. Spomenuta teorija nam daje precizne ili aproksimativne izračune vjerojatnosti propasti, odgovarajućih rezervi, distribuciju sume zahtjeva za isplatama i druga svojstva idealiziranog portfelja osiguravatelja. Posljednjih godina predloženi su neki specijalni modeli koji uključuju mogućnost klasteriranja događaja, na primjer Hawkesovi procesi.

Proučavamo asimptotske distribucije ukupnog iznosa zahtjeva za isplatom u označenim Poissonovim procesima s klasterima u kojima oznake određuju visinu, ali i druge karakteristike pojedinih zahtjeva te potencijalno utječu na stopu dolazaka budućih zahtjeva. Označeni Hawkesovi procesi u tom slučaju postaju specijani slučaj općenitog modela označenih Poissonovih procesa s klasterima. Malo preciznije, dolaske zahtjeva za isplatom u promatranom portfelju modeliramo označenim točkovnim procesom, npr. oblika

$$N = \sum_{k=1}^{\infty} \delta_{\tau_k, A^k},$$

pri čemu su τ_k nenegativne slučajne varijable koje predstavljaju vremena dolazaka s nekim stupnjem klasteriranja, a A^k pripadne oznake u nekom metričkom prostoru $\mathbb{M}$. Visina zahtjeva za isplatom u svakom označenom događaju može se izračunati upotrebom izmjerivog preslikavanja $f(A^k)$ iz prostora oznaka u nenegativne realne brojeve. Tada se suma zahtjeva za isplatom u intervalu $[0, t]$ može izraziti kao

$$S(t) = \sum_{\tau_k \leq t} f(A^k).$$

Promatramo učinak klasteriranja na $S(t)$, kada $t \to \infty$ čak i u slučaju kada distribucija individualnih zahtjeva ne zadovoljava pretpostavke klasičnog centralnog graničnog teorema. Osim novih rezultata u slučaju kada drugi moment nije konačan, u izračunima koristimo drugačiji pristup koji se temelji na graničnim teoremima

za zaustavljene dvodimenzionalne slučajne šetnje (koji proizlaze iz Anscombeovog teorema), a ne na martingalnom centralnom graničnom teoremu koji je često korišten. Prezentirat ćemo dovoljne uvjete uz koje ukupan iznos zahtjeva zadovoljava centralni granični teorem ili alternativno teži po distribuciji stabilnoj slučajnoj varijabli s beskonačnom varijancom. Diskutirat ćemo nekoliko Poissonnovih modela s klasterima, pri čemu će označeni Hawkesovi procesi biti naš ključni primjer. U posljednjem poglavlju fokus prebacujemo na maksimalan iznos zahtjeva za isplatom u intervalu $[0, t]$. Prezentirat ćemo rezultate vezane uz granično ponašanje maksimuma u slučaju kada pojedinačni zahtjevi za isplatama pripadaju Fréchetovoj, Weibullovoj ili Gumbelovoj maksimalnoj domeni privlačnosti. Ponovo primjenjujemo dobivene rezultate na tri specijalna modela (s posebnim naglaskom na Hawkesove procese). Osim graničnog ponašanja suma i maksimuma, pokušali smo razjasniti pojam stohastičkog intenziteta, posebno jer se u literaturi može pronaći nekoliko različitih definicija spomenutog stohastičkog intenziteta. Razumijevanje stohastičkog intenziteta nam je važno jer se koristi prilikom definiranja Hawkesovih procesa.

Ključne riječi

Točkovni procesi, Poissonovi procesi s klasterima, granični teoremi, Hawkesovi procesi, suma zahtjeva za isplatom, maksimalan zahtjev za isplatom

Contents

1

Introduction

1.1 Motivation

The point process theory is very useful and is applied in order to understand and sometimes solve many problems coming from different areas, for example insurance, finance, biology, meteorology, genetics, seismology, epidemiology, geography, stochastic geometry and many others. We can conclude that the point processes represent significant part of the stochastic theory. The main idea behind them is counting points. As Mikosch says in [34], counting is bread and butter in non-life insurance. The modelling of claim numbers, claim amounts and calculating appropriate reserves are some of the most important tasks of the actuaries.

The most applied point processes are surely Poisson processes. They can be viewed as the building block or link to the almost every point process one can encounter in the literature. For example, in the Cramér–Lundberg model the claim number process $(N(t))$ (i.e. stochastic process representing the number of points which have appeared up until time t) is a homogenous Poisson process. For this model elegant mathematical analysis exists and takes acclaimed place in non-life insurance theory, see for instance [2] or [34].

We are interested in some special point process models which take into account the possibility of clustering of insurance events. So, we want to substitute Cramér–Lundberg model with model which can "capture" clusters. The leading role in our applications have linear marked Hawkes processes. Hawkes processes are one of the most important representatives of the so-called self-exciting processes. They were introduced in 1971. by Allan G. Hawkes ([22], [23]). The self-excitation property means that the potential of the future events depends on the whole past of the process which makes them non–Markovian. Hawkes processes are cluster processes and are rather popular, so one can find numerous examples of their application in many diverse areas such as finance (see [3], [10]), neuroscience (see [37]), criminology (see

[15]), seismology (see [36]), genetics (see [40]), and so on. Regardless of their diverse and rich application, theoretical results can be gathered in a relatively short list. For instance, some results on ruin probabilities can be found in [42] and [44]. Stability results can be found in [8] and in [9]. As to the limit theorems, Bacry [3] proved the central limit theorem for linear Hawkes process. Zhu [28] proved the functional limit theorem for non-linear Hawkes processes and the central limit theorem for linear but marked Hawkes processes.

We find this limiting behaviour especially interesting because if one has the information about the asymptotic behaviour of the total claim amount, one can recommend how much premium should be charged in order to avoid insolvency of the company. Also, we noticed that there are not so many results available in the literature for Hawkes processes in the cases when we do not have finiteness of second moments (in that case, central limit theorem–type result is not possible). We explore the possible limits even in that case, but in more general setting. Namely, we analyse marked Poisson cluster processes where the process of cluster centres is a homogenous Poisson process, and every cluster is also a point process generated in every point belonging to the ground process (intuitively, point process representing the cluster is superimposed on the each arriving point coming from the ground process).

1.2 Outline

In the next chapter we clarify the notion of the stochastic intensity which is important for understanding the Hawkes process and present the Poisson cluster representation of linear marked Hawkes processes. We give some existing definitions and results from Morariu–Patrichi and Pakkanen [35] and Massoulié [33] which will be used throughout the thesis.

Chapter 3 is dedicated to the study of asymptotic distribution of the total claim amount in the marked Poisson cluster model even in the case when the distribution of the individual claims does not satisfy assumptions of the classical central limit theorem. We apply our results to the three special models with marked Hawkes processes as the most significant example.

The last chapter is dedicated to limiting behaviour of the maximal claim size in the same model. We prove that under certain conditions on claim sizes and on the number of items in the so–called leftover effect, the maximal claim size, when properly normalized, converges weakly to an extreme value distribution. Again, we apply this result to the three special marked Poisson cluster models.

2

On stohastic intensity for linear marked Hawkes process

It is important to emphasize that the set up (with some adjustments) we used in this chapter can be found in Morariu–Patrichi and Pakkanen [35]. The aim of this chapter is to explain the stochastic intensity and there is no significant contribution to knowledge coming from the material in it, rather a collection of existing results which can be seen as preliminaries for Chapter 3 and 4.

2.1 Hawkes processes and stochastic intensity - intuition

Hawkes point processes are the key example in our analysis. In order to understand them, it is useful to present them intuitively. On the other hand, we need to defined them rigorously. The stochastic intensity which appears in the definition of the Hawkes process turns out to be the key notion for understanding Hawkes processes. To do so, our first step is to clarify several different approaches present in the literature when defining stochastic intensity. The largest part of the remaining of this section is based on the work of Morariu–Patrichi and Pakkanen [35] and Massoulié [33]. Other important references are [8], [12] and [9].

Intuitively, Hawkes processes are a random set of points which come in clusters. Little more precisely: a random measure on $\mathbb{R}$ (or $[0, \infty)$) which can be described through its stochastic intensity. Hawkes process N is a counting process, meaning that $N(t)$ represents the number of events that have occurred until time t. Intensity process (or stochastic intensity) is a notion describing the dynamics of N. Heuristically, when it exists, the intensity $\lambda(t)$ of N at time t is random and such that

$$\mathbb{E}\left[N(t+dt) - N(t)|\mathcal{F}_t^N\right] \approx \lambda(t)dt,$$

or

$$\mathbb{P}\left(dN(t) = 1|\mathcal{F}_{t-}^N\right) = \mathbb{P}\left(N(t) - N(t-dt) = 1|\mathcal{F}_{t-}^N\right) \approx \lambda(t)dt,$$

where $\mathbb{F}^N = (\mathcal{F}_t^N)_{t\geq 0}$ is the natural filtration (which corresponds to the internal history of N) and $\mathcal{F}_{t-}^N = \sigma(\cup_{s<t}\mathcal{F}_s^N)$. Intuitively, $\lambda(t)dt$ is the expected number of events in the infinitesimal time window $\langle t-dt, t\rangle$, given what has happen so far. Besides describing counting process, intensity process can be used to specify them implicitly. A key example where the intensity process is used in this way is the class of linear Hawkes processes [23] which are defined through the intensity

$$\lambda(t) = \nu + \int_{[0,t\rangle} h(t-s)N(ds), \quad t \geq 0$$

where $\nu > 0$ is fixed and $h : [0,\infty\rangle \to [0,\infty\rangle$ is an integrable function (usually called kernel, fertility or exciting function) such that $\kappa = \int_0^\infty h(s)ds < 1$.

Because of the fact that this type of processes depend on the whole past, it is common to name them as the main example of self-exciting point process: events (points) trigger new events later in time. We intent to study one possible generalization of Hawkes processes. Namely, marked linear Hawkes processes where marks A_s are i.i.d, independent of previous arrival times and live on some complete separabile metric space $(\mathbb{M}, \mathcal{B}(\mathbb{M}))$. Intuitvely, they can represent some additional information. In this case the intensity depends on both past events and marks, so it has the following form:

$$\lambda(t) = \nu + \int_{[0,t\rangle}\int_{\mathbb{M}} h(t-s,a)N(ds,da), \quad t \geq 0,$$

where the fertility function now depends on the marks (A_s) as well. Due to the implicit definition of the marked Hawkes processes via stochastic intensity which in turn depends on the history of the process we intend to define, we need to justify their existence and uniqueness.

2.1.1 Remark. *It is known that a marked point process with intensity λ expressed in terms of an intensity functional ψ can be formulated as a solution to a Poisson driven stochastic differential equation (SDE for short). This was done by [9], [33]. In there, the intensity functional has to satisfy some Lipschitz-type condition. Morariu–Patrichi and Pakkanen [35] show that under certain integrability or decay condition it is enough for ψ to be dominated by either a Hawkes functional (i.e. functional bellonging to some Hawkes process) or an increasing function of the total number of past events in order to obtain the existence of a strong solution to the Poisson driven SDE.*

2.2 Framework

2.2.1 A framework for general point processes

Let U, $\mathbb{M}$ be a complete separable metric spaces and $\mathcal{B}(U)$, $\mathcal{B}(\mathbb{M})$ their Borel $\sigma-$algebras respectively. While $\mathbb{M}$ represents the mark space, U is usually $\mathbb{R} \times \mathbb{M}$ or $\mathbb{R}_{\geq 0} \times \mathbb{M}$.

Spaces of integer valued measures

Let ξ be a Borel measure on $\mathbb{R} \times \mathbb{M}$. If $\xi(A) < \infty$, for all bounded Borel set $A \in \mathcal{B}(\mathbb{R} \times \mathbb{M})$ then ξ is boundedly finite. Let $\mathcal{N}^{\infty}_{\mathbb{R}\times\mathbb{M}}$ be the space of Borel measures on $\mathbb{R} \times \mathbb{M}$ with values in $\mathbb{N} \cup \{\infty\}$. Let $\mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ be the set of all $\xi \in \mathcal{N}^{\infty}_{\mathbb{R}\times\mathbb{M}}$ such that ξ is boundedly finite. Let $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ be the set of all $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ such that their ground measure $\xi_g(\cdot) := \xi(\cdot \times \mathbb{M})$ satisfies:

- $\xi_g \in \mathcal{N}^{\#}_{\mathbb{R}}$ and
- $\xi_g(\{t\}) = 0$ or 1, $\forall t \in \mathbb{R}$.

The second condition means that the ground measure is simple, i.e. there can be at most one event at each time. It is a consequence of the first condition, $\xi_g \in \mathcal{N}^{\#}_{\mathbb{R}}$ that not all point processes on product spaces are marked point processes. For example, the bivariate Poisson process on $\mathbb{R} \times \mathbb{R}$ with parameter measure $\mu dx dy$ cannot be represented as an marked point process on $\mathbb{R} \times \mathbb{R}$ because such a Poisson process has $N(A \times \mathbb{R}) = \infty$ a.s. for Borel sets A of positive Lebesgue measure [12].

2.2.1 Remark. *Set $\mathcal{N}^{\infty}_{\mathbb{R}\times\mathbb{M}}$ contains realisations of potentially explosive point process and set $\mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ of non-explosive point process. Notice that $\mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ contains potentially explosive marked point process so we take $\xi \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ to represent a realization of a non-explosive marked point process. Observe: $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}} \subset \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}} \subset \mathcal{N}^{\infty}_{\mathbb{R}\times\mathbb{M}}$.*

2.2.2 Remark. *When $\xi \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ and $\xi(\{(t,a)\}) = 1$ for some $t \in \mathbb{R}$ and $a \in \mathbb{M}$, this means that an event is happening at time t with mark a.*

The so-called $w^{\#}$-distance ("vague hash") makes $\mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ a complete separable metric space (Theorem A2.6 III in [12], [4]). $\mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}})$ corresponds to the σ-algebra generated by all mappings $\xi \mapsto \xi(A)$, $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$, $A \in \mathcal{B}(\mathbb{R} \times \mathbb{M})$. Proposition A2.6 II in [12] characterises convergence in this topology ($w^{\#}$-topology). Lemma A.1.1 in [35] shows that $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ is a Borel set of $\mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ which is not completely trivial fact.

Non-explosive marked point processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The trace of $A \in \mathcal{S}$ on σ-algebra $\mathcal{S}$ is defined as $A \cap \mathcal{S} := \{A \cap S : S \in \mathcal{S}\}$.

2.2.1 Definition. *A **non-explosive point process** on U is a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathcal{N}_U^{\#}, \mathcal{B}(\mathcal{N}_U^{\#}))$.*

2.2.2 Definition. *A **non-explosive marked point process** N on $\mathbb{R} \times \mathbb{M}$ is a non-explosive point process N on $\mathbb{R} \times \mathbb{M}$ such that $N(\omega) \in \mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#g}$, $\forall \omega \in \Omega$.*

Lemma 1.6. in [27] implies that $\mathcal{B}(\mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#g}) = \mathcal{B}(\mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#}) \cap \mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#g}$, where $\mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#g}$ is also equipped with $w^{\#}$. So, $\mathcal{B}(\mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#g})$ is actually the trace of $\mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#g}$ on $\mathcal{B}(\mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#})$. Hence, Definition 2.2.2 is equivalent to saying that a non-explosive marked point process N on $\mathbb{R} \times \mathbb{M}$ is a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#g}, \mathcal{B}(\mathcal{N}_{\mathbb{R}\times\mathbb{M}}^{\#g}))$.

2.2.3 Definition. *Let N be a non-explosive point process on U. We define **induced probability measure** $\mathcal{P}^N$ on the measurable space $(\mathcal{N}_U^{\#}, \mathcal{B}(\mathcal{N}_U^{\#}))$ by the relation*

$$\mathcal{P}^N(A) := \mathbb{P}\left(N^{-1}(A)\right) = \mathbb{P}\left(\{\omega \in \Omega : N(\omega) \in A\}\right), \quad A \in \mathcal{B}(\mathcal{N}_U^{\#}).$$

Enumeration representation

It is common to define a marked point process on $\mathbb{R}_{\geq 0} \times \mathbb{M}$ as a sequence $(T_n, A_n)_{n\in\mathbb{N}}$ of random variables in $\langle 0, \infty] \times \mathbb{M}$ such that $(T_n)_{n\in\mathbb{N}}$ is a non-decreasing and $T_n < \infty$ implies $T_n < T_{n+1}$. This sequence is usually called **enumeration**. T_n represents the time when the n^{th} event occurs with mark A_n. $T_n < \infty$ with $T_{n+1} = \infty$ means that there are no more events after time T_n. To avoid an explosion (in the sense that $\lim_{n\to\infty} T_n < \infty$ is possible) we assume that $\lim_{n\to\infty} T_n = \infty$ a.s.. Morariu–Patrichi and Pakkanen [35] show that there is one to one correspondence between non-explosive marked point processes on $\mathbb{R}_{\geq 0} \times \mathbb{M}$ and non-explosive enumeration.

Poisson process

Let ν be a boundedly finite measure on $(U, \mathcal{B}(U))$. We say that a non-explosive point process N on U is a Poisson point process or a Poisson random measure with parameter (or mean) measure ν (we write PRM(ν)) if $N(A_1), N(A_2), \ldots N(A_n)$ are mutually independent for all disjoint sets $A_1, A_2, \ldots, A_n \in \mathcal{B}(U)$, $n \in \mathbb{N}$ and $N(A)$ is a Poisson random variable with parameter $\nu(A)$, $\forall A \in \mathcal{B}(U)$, A bounded set.

Pathwise integration

Let N be a non-explosive point process on U. Let $H : \Omega \times U \to \mathbb{R}_{\geq 0}$ be an $\mathcal{F} \otimes \mathcal{B}(U)-$measurable mapping ($H$ is an $\mathbb{R}_{\geq}-$ valued stochastic process on U). The integral of H against N is

$$I(\omega) := \int_U H(\omega, u)N(\omega, du), \quad \omega \in \Omega.$$

By the monotone class theorem $\omega \mapsto I(\omega)$ is $\mathcal{F}-$measurable. When N is a non-explosive marked point process on $\mathbb{R}_{\geq 0} \times \mathbb{M}$, the integral can be written as

$$\int\int_{\mathbb{R}_{\geq 0}\times\mathbb{M}} H(t,a)N(dt,da) = \sum_{n\in\mathbb{N}} H(T_n, A_n)\mathbb{1}_{\{T_n<\infty\}} \quad a.s.$$

where (T_n, A_n) is the enumeration representation of N.

Shifts, restrictions, histories and predictability

2.2.4 Definition. *We define* ***the shift operator*** *for all $t \in \mathbb{R}$ as $\Theta_t : \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}} \to \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ by $\Theta_t\xi(A) := \xi(A + t)$, $A \in \mathcal{B}(\mathbb{R} \times \mathbb{M})$ where $A + t := \{(s + t, a) \in \mathbb{R} \times \mathbb{M} : (s, a) \in A\}$.*

Then, for any non-explosive point process N on $\mathbb{R} \times \mathbb{M}$ define $\Theta_t N$ as $(\Theta_t N)(\omega) := \Theta_t(N(\omega)), \omega \in \Omega$.

Lemma A.2.1. in [35] shows that $\Theta_t\xi$ is jointly continuous in t and ξ.

Let $\xi^{<0}$ be the restriction to the negative real line of any realisation $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ which is defined by $\xi^{<0}(A) := \xi(A \cap \mathbb{R}_{<0} \times \mathbb{M})$, $A \in \mathcal{B}(\mathbb{R} \times \mathbb{M})$. The restriction of any non-explosive point process N on $\mathbb{R} \times \mathbb{M}$ is simply defined by $N^{<0}(\omega) := (N(\omega))^{<0}$, $\omega \in \Omega$. In the same way we define:

$$\xi^{\leq 0}(A) := \xi(A \cap \mathbb{R}_{\leq 0} \times \mathbb{M}) \text{ and } N^{\leq 0}(\omega) := (N(\omega))^{\leq 0},$$

$$\xi^{\geq 0}(A) := \xi(A \cap \mathbb{R}_{\geq 0} \times \mathbb{M}) \text{ and } N^{\geq 0}(\omega) := (N(\omega))^{\geq 0},$$

$$\xi^{>0}(A) := \xi(A \cap \mathbb{R}_{>0} \times \mathbb{M}) \text{ and } N^{>0}(\omega) := (N(\omega))^{>0}.$$

2.2.3 Remark. *This notations will help to refer to the internal history of N. For example, for all $t \in \mathbb{R}$ $(\Theta_t N)^{<0}$ contains the history of the process up to time t, excluding t.*

Also, we will use $\Theta_t\xi^{<0} := (\Theta_t\xi)^{<0}$.

2.2.4 Remark. *Lemma A.2.2. and A.2.3. in [35] prove that these restrictions are measurable mappings and that $\Theta_t\xi^{<0}$ is left continuous as a function of $t \in \mathbb{R}$.*

2.2.5 Definition. *Let N be a non-explosive marked point process on $\mathbb{R} \times \mathbb{M}$. We can define the* ***filtration*** $\mathbb{F}^N = (\mathcal{F}_t^N)_{t\in\mathbb{R}}$ *that corresponds to the* ***internal history*** *of N by*

$$\mathcal{F}_t^N := \sigma\left\{N(B \times A) : B \in \mathcal{B}(\mathbb{R}), B \subset \langle -\infty, t\right], A \in \mathcal{B}(\mathbb{M})\}, \quad \forall t \in \mathbb{R}.$$

2.2.5 Remark. *Using Lemma 1.4 in [27] and the characterisation of $\mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}})$ given in Theorem A2.6. III in [12] one can check that $\mathcal{F}_t^N = \sigma(\Theta_t N^{\leq 0})$.*

2.2.6 Definition. ***History*** *is any filtration that contains the internal history of N, that is any $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ such that $\mathcal{F}_t^N \subset \mathcal{F}_t$, $t \in \mathbb{R}$. One says that N is $\mathbb{F}-$adapted.*

2.2.7 Definition. *The σ-algebra $\mathcal{F}^p$ is* ***the predictable*** $\sigma-$***algebra*** *on $\Omega \times \mathbb{R} \times \mathbb{M}$ corresponding to a history $\mathbb{F}$ if it is generated by all the sets of the form*

$$B \times \langle s, t] \times A, \quad s, t \in \mathbb{R}, \quad s < t, \quad A \in \mathcal{B}(\mathbb{M}), \quad B \in \mathcal{F}_s.$$

2.2.8 Definition. *Any mapping $H : \Omega \times \mathbb{R} \times \mathbb{M} \to \mathbb{R}$ that is $\mathcal{F}^p-$measurable is called an $\mathbb{F}-$* ***predictable process****. Any mapping $H : \Omega \times \mathbb{R}_{>0} \times \mathbb{M} \to \mathbb{R}$ that is $(\Omega \times \mathbb{R}_{>0} \times \mathbb{M}) \cap \mathcal{F}^p-$measurable is also called an $\mathbb{F}-$* ***predictable process****.*

2.2.9 Definition. *Given an $\mathbb{F}-$stopping time τ,* ***the strict past*** *$\mathcal{F}_{\tau-}$ is defined as the $\sigma-$algebra generated by all the classes $\{t < \tau\} \cap \mathcal{F}_t$, $t \in \mathbb{R}$.*

Stochastic intensity for non-marked processes

2.2.10 Definition. *Let N be a point process on $\mathbb{R}$ and $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ a history. Let $\lambda : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a non-negative $\mathbb{F}-$predictable process. We say that λ is* ***the*** $\mathbb{F}$ ***intensity*** *or* ***stochastic intensity*** *of N if for every non-negative, $\mathbb{F}-$predictable process $H : \Omega \times \mathbb{R}_{\geq 0} \times \mathbb{M} \to \mathbb{R}_{\geq 0}$,*

$$\mathbb{E}\left[\int_0^\infty H(t)N(dt)\right] = \mathbb{E}\left[\int_0^\infty H(t)\lambda(t)dt\right] \tag{2.1}$$

In the sequel of this subsection 2.2.1 two theorems from Brémaud [8] will be stated to show that in the unmarked case, the stochastic intensity process has a characterisation through a local martingale property.

2.2.11 Definition. *Let $(X_t)_{t\in\mathbb{R}}$ be a real-valued stochastic process adapted to a history $\mathbb{F}$ and let $(T_n)_{n\geq 1}$ be an increasing family of $\mathbb{F}-$stopping times such that $\lim_{n\to\infty} T_n = \infty$ and for each $n \geq 1$, $X_{t\wedge T_n}$ is a $\mathbb{F}-$martingale. Then (X_t) is called an $\mathbb{F}-$***local martingale****.*

2.2.1 Theorem (Theorem 8 (a) in [8])**.** *Let N be a non-explosive point process adapted to $\mathbb{F}$. If N admits the $\mathbb{F}$-intensity λ (where for every $t \geq 0$, $\int_0^t \lambda(s)ds < \infty$, a.s), then*

$$M(t) = N(t) - \int_0^t \lambda(s)ds$$

is an $\mathbb{F}$-local martingale.

2.2.12 Definition. *Let N be a non-explosive point process on $\mathbb{R}_{>0}$ adapted to $\mathbb{F}$. N is $\mathbb{F}-$**progressive process** if for each $t \geq$, the map*

$$[0,t] \times \Omega \to \mathbb{R}$$

$$(s,\omega) \mapsto X_s(\omega)$$

is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$–measurable.

2.2.2 Theorem (Theorem 9 in [8]). *Let N be a non-explosive point process adapted to $\mathbb{F}$ and suppose that for some non-negative $\mathbb{F}-$progressive process λ and for all $n \geq 1$*

$$N(t \wedge T_n) - \int_0^{t \wedge T_n} \lambda(s) ds \tag{2.2}$$

is an $\mathbb{F}$-martingale, where (T_n) represents the enumeration of N. Then λ is the $\mathbb{F}$-intensity of N.

In the literature one can find several different definitions of stochastic intensity process for non-marked processes. Namely, some other definitions one can find are:

- Zhu [44] and Brémaud and Massoulié [9] define the $\mathbb{F}-$intensity as an $\mathbb{F}-$progressively measurable process λ with

$$\mathbb{E}\left[N\langle a,b] | \mathcal{F}_a^N\right] = \mathbb{E}\left[\int_a^b \lambda(s) ds | \mathcal{F}_a^N\right] \tag{2.3}$$

  a.s., for all intervals $\langle a, b]$.

- Brémaud [8] defines the $\mathbb{F}$-intensity as an $\mathbb{F}-$progressively measurable process λ by Equation (2.1) for every non-negative, $\mathbb{F}-$predictable process $H : \Omega \times \mathbb{R}_{\geq 0} \times \mathbb{M} \to \mathbb{R}_{\geq 0}$.

- Daley and Vere Jones [12] define the $\mathbb{F}-$intensity as any $\mathbb{F}-$adapted process λ such that a.s. for all t,

$$A(t) = \int_0^t \lambda(s) da,$$

  where A represents the $\mathbb{F}$ compensator of N (which means that A is a non-decreasing right–continuous predictable process such that (2.2) is satisfied).

- Kirchner [30] and Laub et al. [31] simply define intensity as a limit (when it exists)

$$\lambda(t) := \lim_{\delta \to 0} \frac{\mathbb{E}(N(t+\delta) - N(t) | \mathcal{F}_t)}{\delta}.$$

Let us now clarify their connection using Theorem 2.2.1 and Theorem 2.2.2.

Firstly, Theorem T13 in [8] says that when the progressive intensity exists, then we can always find an $\mathbb{F}$-intensity $\lambda'(t)$ which is $\mathbb{F}-$predictable. By Theorem T12 in [8] this predictable version is almost everywhere unique. So, if we put the requirement on stochastic intensity to be a progressive process, uniqueness is omitted.

Assume (2.1). Then, using Theorem (2.2.1) we have

$$\begin{aligned}
&\mathbb{E}[M_{t\wedge T_n}|\mathcal{F}_s] = M_{s\wedge T_n}\\
&\mathbb{E}[M_{t\wedge T_n}|\mathcal{F}_s] - \mathbb{E}[M_{s\wedge T_n}|\mathcal{F}_s] = 0\\
&\mathbb{E}[N_{t\wedge T_n} - N_{s\wedge T_n}|\mathcal{F}_s] = \mathbb{E}\left[\int_{s\wedge T_n}^{t\wedge T_n} \lambda(u)du|\mathcal{F}_s\right].
\end{aligned}$$

Letting $n \to \infty$,

$$\begin{aligned}
\mathbb{E}[N(t) - N(s)|\mathcal{F}_s] &= \mathbb{E}\left[\int_s^t \lambda(u)du|\mathcal{F}_s\right],\\
\mathbb{E}[N\langle s,t]|\mathcal{F}_s] &= \mathbb{E}\left[\int_s^t \lambda(u)du|\mathcal{F}_s\right]
\end{aligned}$$

which is the formula appearing in [28] and [9]. Assume now (2.3). Substituting in (2.3) s with $s \wedge T_n$ and t with $t \wedge T_n$ we get (similarly as above) that $N_{t\wedge T_n} - \int_0^{t\wedge T_n} \lambda(u)du$ is a martingale. Using Theorem (2.2.2) we can conclude that λ is the $\mathbb{F}-$intensity of N. If λ is right-continuous and bounded, it follows from (2.3), by application of the Lebesgue averaging theorem and the Lebesgue dominated convergence theorem successively, that a.s.

$$\lim_{b\to a} \frac{\mathbb{E}[(N(b) - N(a))|\mathcal{F}_a]}{b-a} = \lambda(a)$$

which is the expression one can find in [30].

Stochastic intensity for marked point processes

2.2.13 Definition. *Let $(A)_n$ be i.i.d random variables in $\mathbb{M}$ representing the marks. Let $Q(da)$ be the distribution of the marks, i.e. a probability measure on $(\mathbb{M}, \mathcal{B}(\mathbb{M}))$ defined by $Q(C) = \mathbb{P}(A_n \in C)$, $C \in \mathcal{B}(\mathbb{M})$. Let N be a marked point process on $\mathbb{R} \times \mathbb{M}$ and $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ a history. Let $\lambda : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a non-negative $\mathbb{F}-$predictable process. We say that λ is **the** $\mathbb{F}$ **intensity** or **stochastic intensity** of N if for every non-negative, $\mathbb{F}-$predictable process $H : \Omega \times \mathbb{R}_{\geq 0} \times \mathbb{M} \to \mathbb{R}_{\geq 0}$,*

$$\mathbb{E}\left[\int_0^\infty \int_{\mathbb{M}} H(\omega,t,a)N(\omega,dt,da)\right] = \mathbb{E}\left[\int_0^\infty \int_{\mathbb{M}} H(\omega,t,a)\lambda(\omega,t)Q(da)dt\right]. \quad (2.4)$$

2.2.6 Remark. *If an intensity process exists, it is then unique up to $\mathbb{P}(d\omega)Q(da)dt$-null sets thanks to the predictability requirement ([8], Section II.4, [12], p. 391).*

Intensity can be expressed in terms of a functional applied to the point process.

2.2.14 Definition. *Let $\Psi : \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be a measurable functional. We say that a non-explosive marked point process $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ admits Ψ as its* ***intensity functional*** *if N admits an $\mathbb{F}$-intensity $\lambda : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that*

$$\lambda(\omega, t) = \Psi(\theta_t N(\omega)^{<0}), \quad \mathbb{P}(d\omega)Q(da)dt - a.e.$$

Recall, almost everywhere convergence is a weakened version of point-wise convergence hypothesis which states that, for X a measure space, $f_n(x) \to f(x)$ for all $x \in Y$, where Y is a measurable subset of X such that $\mu(X\backslash Y) = 0$.

Examples

1. ([12], p.358) Consider a one-point process consisting of a single point whose location is defined by a positive, continuous random variable X with distribution function F (and density f). The associated counting process is defined by

   $$N(t, \omega) = \delta_{\langle 0,t\rangle}(X(\{\omega\})), \quad t > 0, \quad \omega \in \Omega.$$

   N is non-decreasing, right-continuous, and uniformly bounded so there is no problem about the existence of moments. Next, because $N(t, \omega) = 1$ implies $N(t', \omega) = 1$ for all $t' \geq t$. On the other hand, if $N(t, \omega) = 0$, then we know that $X(\omega) > t$, so the stochastic intensity is

   $$\lambda(t) = \frac{f(t)}{1 - F(t)} \mathbb{1}_{N[0,t\rangle = 0}.$$

2. Let N be a Poisson point process such that $N\langle a, b] \sim \text{Poisson}\left(\int_a^b l(s)ds\right)$ for some deterministic, locally integrable function $l : [0, \infty\rangle \to [0, \infty\rangle$. The Poisson process exist, see for example Theorem 9.2.X in [12], Example 9.2(b). (Stochastic) intensity is $\lambda(t) = l(t)$ because of the so-called Partial result in [8], page 24 (reversal of Watanabe's theorem which says that if $N_t - \int_0^t \lambda(s)ds$ is a martingale, then point process N is Poisson point process).

Intensity of (linear marked) Hawkes processes

2.2.15 Definition. *We call a non-explosive point process $N : \Omega \to \mathcal{N}^{\#}_{\mathbb{R}}$ a linear (unmarked) Hawkes process if it admits a stochastic intensity $\lambda : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ of the form*

$$\lambda(\omega, t) = \nu + \int_0^t h(t - s)N(\omega, ds),$$

where $\nu > 0$ *is fixed and* $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ *measurable function such that* $\int_0^\infty h(s)ds < 1$.

Some possible generalizations are:

- non-linear Hawkes processes (where intensity is not represented as a linear function), see [28],
- Hawkes processes defined on the whole real line (intensity is then called the complete intensity process), see [12],
- **marked Hawkes processes**, see [28], [5],
- multi-type Hawkes processes, see [3] and [35] and
- graph (and skeleton) Hawkes processes, see [30].

Stability results (convergence towards stationary version of the process) can be found in [12], [23], [9], [8], [33], strong law of large numbers for $\frac{N_t}{t}$ in [12] and central limit theorem in [3] and [28].

We are interested in linear marked Hawkes processes, so we have to include marks in their intensity. Namely,

2.2.16 Definition. *We say that a non-explosive point process* $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ *is a* ***linear marked Hawkes process*** *if it admits a stochastic intensity* $\lambda : \Omega \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ *of the form*

$$\lambda(\omega, t) = \nu + \int_0^t \int_{\mathbb{M}} h(t-s, a) N(\omega, ds, da), \tag{2.5}$$

where $\nu > 0$ *is fixed and* $h : \mathbb{R}_{\geq 0} \times \mathbb{M} \to \mathbb{R}_{\geq 0}$ *a measurable function such that* $\int_0^\infty \int_{\mathbb{M}} h(s,a) Q(da) ds < 1$.

2.2.7 Remark. *Notice that we can express intensity (because of the enumeration representation) in terms of sum; namely*

$$\begin{aligned} \lambda(\omega, t) &= \nu + \int_0^t \int_{\mathbb{M}} h(t-s, a) N(\omega, ds, da) \\ &= \nu + \sum_{T_i(\omega) \leq t} h(t - T_i(\omega), A_i(\omega)) \end{aligned}$$

Initial condition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space endowed with a right-continuous filtration $(\mathcal{F}^d_t)$ that will correspond to the driving Poisson process in the SDE (see bellow). Let $N_{\leq 0}$ be a given marked point process on $\mathbb{R} \times \mathbb{M}$ such that $N_{\leq 0}(\omega)^{\leq 0} = N_{\leq 0}(\omega)$, for all $\omega \in \Omega$ (there are no events on $\mathbb{R}_{>0}$). We reserve the notation $N_{\leq 0}$ to refer to the initial condition.

2.2.17 Definition. *Let $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ be a non-explosive marked point process on $\mathbb{R} \times \mathbb{M}$. We say that N satisfies a **strong initial condition** $N_{\leq 0}$ if $N(\omega)^{\leq 0} = N_{\leq 0}(\omega)$ a.s., where $\omega \in \Omega$.*

2.2.18 Definition. *Let $(\Omega', \mathcal{F}', \mathbb{P}')$ be another probability space potentially different from $(\Omega, \mathcal{F}, \mathbb{P})$. Let $N' : \Omega' \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ be a non-explosive marked point process on $\mathbb{R} \times \mathbb{M}$. We say that N' satisfies a **weak initial condition** $N_{\leq 0}$ if the induced probability $\mathcal{P}^{N'^{\leq 0}}$ coincides with $\mathcal{P}^{N_{\leq 0}}$.*

2.2.2 Existence and uniqueness of linear marked Hawkes processes

Existence and uniqueness problem

The linear marked Hawkes process is defined implicitly via its intensity process, which, in turn, depends on the history of the process. Clearly, it is not clear that such a point process exists. More generally, given an initial condition $N_{\leq 0}$ and an $\mathbb{F}$- intensity $(\lambda(t))$ (or measurable intensity functional $\Psi : \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$), one can ask if there exists a unique non-explosive marked point process N that satisfies the initial condition $N_{\leq 0}$ on $\mathbb{R}_{\leq 0}$ and admits λ as its stochastic intensity (or, equivalently, Ψ as its intensity functional) on $\mathbb{R}_{>0}$. [33] deals with this problem by reformulating the existence problem as a Poisson driven SDE, extending the work of [9].

In those papers strong existence and uniqueness are obtained by imposing a Lipschitz type condition on the intensity functional Ψ. To be more precise, they assume that there exists a function $h : \mathbb{R}_{\geq 0} \times \mathbb{M} \to \mathbb{R}_{\geq 0}$ such that the (Lipschitz type) condition

$$\left|\Psi(\zeta) - \Psi(\zeta')\right| \leq \int_{\langle -\infty, 0\rangle} \int_{\mathbb{M}} h(-s, a)|\zeta - \zeta'|(ds, da), \quad a \in \mathbb{M}, \zeta, \zeta' \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}} \tag{2.6}$$

holds. In the work of [35] one of the goals was to construct strong solution to a Poisson driven SDE without imposing the Lipschitz condition on the intensity functional Ψ. They did that for the so called hybrid marked point processes by imposing a weaker sub-linearity condition on Ψ. Namely, they demand for Ψ to be bounded by an intensity functional coming from Hawkes process. For linear marked Hawkes processes the Lipschitz condition (2.6) is trivially satisfied so we will present a version of the existence and uniqueness result coming from [33].

The Poisson-driven SDE

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be given and let $M : \Omega \to \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}\times\mathbb{R}}$ be a Poisson process on $\mathbb{R} \times \mathbb{M} \times \mathbb{R}$ with a mean measure $dt \times Q(da) \times dz$. Denote by $(\mathcal{F}^M_t)_{t\in\mathbb{R}}$ the internal history of M on Ω.

2.2.19 Definition. *Let M be a Poisson process on $\mathbb{R} \times \mathbb{M} \times \mathbb{R}$ and $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ be a filtration. We say that M is Poisson* ***relative*** *to $\mathbb{F}$ (or $\mathbb{F}-$Poisson process) if for all $t \in \mathbb{R}$, the point process $\theta_t M^{\leq 0}$ is $\mathcal{F}_t-$measurable and $\sigma(\theta_t M^{>0})$ is independent of $\mathcal{F}_t$.*

Naturally, M is Poisson relative to $(\mathcal{F}^M_t)_{t\in\mathbb{R}}$.

Let $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ be the filtration on Ω such that, for all $t \in \mathbb{R}$, $\mathcal{F}_t$ is the $\mathbb{P}-$completion of $\mathcal{F}^{N_{\leq 0}}_t \otimes \mathcal{F}^M_t$ in $\mathcal{F}$. The filtration $\mathbb{F}$ is complete ([27], p. 123). Morariu-Patrichi and Pakkanen [35] proved that M is Poisson relative to $\mathbb{F}$. We want to solve the following Poisson-driven SDE.

2.2.20 Definition. *Let $\Psi : \mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be a given measurable functional and M a Poisson process on $\mathbb{R} \times \mathbb{M} \times \mathbb{R}$ relative to $\mathbb{F}$. By a solution to the* ***Poisson–driven SDE*** *we mean an $\mathbb{F}$-adapted non-explosive marked point process $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$, which admits λ as its stochastic intensity process, that solves*

$$\begin{cases} N(dt, da) = M(dt, da, [0, \lambda(t)]), & t \in \mathbb{R}_{>0}, a.s., \\ \lambda(\omega, t) = \Psi(\theta_t N(\omega)^{<0}), & t \in \mathbb{R}_{>0}, \omega \in \Omega, \\ N^{\leq 0}(\omega) = N_{\leq 0}(\omega), & \omega \in \Omega, a.s., \end{cases}$$

where $N_{\leq 0}$ is a given initial condition.

We will need the notion of point processes stochastic intensity kernel ([8], [12]).

2.2.21 Definition. *Let $\mu(\omega, t, da)$ be a non-negative measure on $(\mathbb{M}, \mathcal{B}(\mathbb{M}))$ indexed by $(\omega, t) \in \Omega \times \mathbb{R}$. $\{\mu(\omega, t, .)\}$ is an $\mathbb{F}-$(stochastic) intensity kernel of N if for all $B \in \mathcal{B}(\mathbb{M})$ $\{\mu(\omega, t, B)\}$ is an $\mathbb{F}$-intensity of the process N_B ($N_B(A) = N(A \times B)$, $A \in \mathcal{B}(\mathbb{R})$).*

For any non-negative function $h(\omega, t, a)$, $\mathcal{F}^p-$measurable, where $\mathcal{F}^p$ is the $\mathbb{F}-$predictable σ-algebra on $\Omega \times \mathbb{R} \times \mathbb{M}$, the following integration formula then holds [8]:

$$\mathbb{E} \int_{\mathbb{R}\times\mathbb{M}} h(\omega, t, a) N(dt, da) = \mathbb{E} \int_{\mathbb{R}\times\mathbb{M}} h(\omega, t, a) dt \mu(\omega, t, da). \tag{2.7}$$

The next lemma is actually an easy extension of the method proposed by Lewis and Shedler [32] for the simulation of non-homogeneous Poisson processes.

2.2.1 Lemma. *Let M be an $\mathbb{F}-$Poisson process on $\mathbb{R} \times \mathbb{M} \times \mathbb{R}_{\geq 0}$, with intensity measure $dt \times Q(da) \times ds$. Let f and g be two non-negative $\mathcal{F}^p-$measurable functions on $\Omega \times \mathbb{R} \times \mathbb{M}$. Define the point process N on $\mathbb{R} \times \mathbb{M}$ by*

$$N(dt, da) = M(dt, da, [f(t,a) \wedge g(t,a), f(t,a) \vee g(t,a)]), \quad t \in \mathbb{R}, a \in \mathbb{M}. \tag{2.8}$$

Then N admits an $\mathcal{F}_t-$intensity kernel $\{|f(t,a) - g(t,a)|Q(da)\}$.

By Lemma 2.2.1, if N is a solution to the Poisson–driven SDE defined in 2.2.20, it admits the $\mathcal{F}_t$-intensity kernel $\{\Psi(\theta_t N^{<0})Q(da)\}$.

2.2.3 Theorem. *Assume that there exists a function $h : \mathbb{R}_{\geq 0} \times \mathbb{M} \to \mathbb{R}_{\geq 0}$ such that the Lipschitz condition (2.6) holds. Assume further that*

$$\kappa := \int_0^\infty \int_{\mathbb{M}} h(t,a) dt Q(da) < 1 \tag{2.9}$$

and

$$\alpha := \Psi(0) < \infty \tag{2.10}$$

hold. Consider a strong initial condition $N_{\leq 0}$ as given in the Definition 2.2.17 which satisfies

$$\sup_{t>0} \epsilon(t) < \infty \text{ and } \lim_{t\to\infty} \epsilon(t) = 0 \tag{2.11}$$

where

$$\epsilon(t) = \mathbb{E} \int_{\mathbb{R}_{\leq 0} \times \mathbb{M}} h(t-s, a) N_{\leq 0}(ds, da). \tag{2.12}$$

Then there exists a solution N to the Poisson–driven SDE defined in 2.2.20.

Proof. We use Picard's method for solving differential equation. We construct recursively point processes N^n and functions $\lambda^n(t)$ on $\mathbb{R}_{\geq 0}$ by letting every N^n coincide with $N_{\leq 0}$ on $\mathbb{R}_{\leq 0} \times \mathbb{M}$ and

$$\lambda^{n+1}(t) = \Psi(\theta_t N^{n<0}), \quad t > 0,$$
$$N^n(dt, da) = M(dt, da, [0, \lambda^n(t)]), \quad t > 0, \quad a \in \mathbb{M},$$

the procedure being initialized by taking $\lambda^0(t) \equiv 0$ and $N^0 \equiv N_{\leq 0}$. It can be shown by induction that the N^n are $\mathcal{F}_t^d-$adapted and that $(t, \omega) \to \lambda^n(t)$ is $\mathcal{F}^p$-measurable. For detailed proof, see Proposition 4.13. in [35]. Next we will show that the sequence N^n converges. By (2.6) and Lemma 2.2.1, one obtains for all $n > 0$

$$\begin{aligned}
\sup_{t\geq 0} \mathbb{E} \left|\lambda^{n+1}(t) - \lambda^n(t)\right| &= \sup_{t\geq 0} \mathbb{E} \left|\Psi\left(\theta_t N^{n<0}\right) - \Psi\left(\theta_t N^{n-1<0}\right)\right| \\
&\leq \sup_{t\geq 0} \mathbb{E} \int_{\langle -\infty, 0\rangle} \int_{\mathbb{M}} h(-s,a) \left|\theta_t N^{n<0} - \theta_t N^{n-1<0}\right| (ds, da) \\
&= \sup_{t\geq 0} \mathbb{E} \int_0^t \int_{\mathbb{M}} h(t-s,a) \left|\lambda^n(s) - \lambda^{n-1}(s)\right| ds Q(da) \\
&\leq \kappa \sup_{s\geq 0} \mathbb{E} \left|\lambda^n(s) - \lambda^{n-1}(s)\right|
\end{aligned}$$

where the last inequality follows from (2.9). So, we have

$$\sup_{t\geq 0} \mathbb{E}\left|\lambda^{n+1}(t)-\lambda^{n}(t)\right| \leq \kappa^{n} \sup_{s\geq 0} \mathbb{E}\left|\lambda^{1}(s)\right|. \tag{2.13}$$

Hence

$$\forall \epsilon > 0, \quad \sum_{n=1}^{\infty} \mathbb{P}\left(\left|\lambda^{n+1}(t)-\lambda^{n}(t)\right| > \epsilon\right)$$

converges because the associated sequence of partial sums is bounded by (2.13) $(\kappa < 1)$. Now we can apply Borel-Cantelli's lemma [14] and conclude that

$$\mathbb{P}\left(\left|\lambda^{n+1}(t)-\lambda^{n}(t)\right| > \epsilon \text{ i.o.}\right) = 0,$$

i.e. that $(\lambda^n(t))$ converges almost surely to some limit $\lambda(t)$. So, for every $t \geq 0$ we have

$$\lambda^{n}(t) \xrightarrow{as} \lambda(t), \quad n \to \infty. \tag{2.14}$$

Moreover, we have the estimate

$$\begin{aligned}
\sup_{t\geq 0} \mathbb{E}\lambda(t) &\leq \frac{1}{1-\kappa} \sup_{t\geq 0} \mathbb{E}\lambda^{1}(t) \\
&= \frac{1}{1-\kappa} \sup_{t\geq 0} \mathbb{E}\Psi\left(\theta_t N^{0^{<0}}\right) \\
&\leq \frac{1}{1-\kappa} \sup_{t\geq 0} \mathbb{E}\left[\epsilon(t)+\Psi(0)\right] \\
&< +\infty
\end{aligned}$$

where $\epsilon(t)$ is defined in (2.12). The above expression is finite because of (2.11) and (2.10). Moreover, in view of the calculations

$$\begin{aligned}
&\sum_{n\geq 0} \mathbb{P}\left(\left|N^{n+1}-N^{n}\right|(\langle 0,T]\times\mathbb{M}) \neq 0\right) \\
&\quad \leq \sum_{n\geq 0} \mathbb{E}\left|N^{n+1}-N^{n}\right|(\langle 0,T]\times\mathbb{M}) \\
&\quad = \sum_{n\geq 0} \int_{\langle 0,T]\times\mathbb{M}} \mathbb{E}\left|\lambda^{n+1}(t)-\lambda^{n}(t)\right| dtQ(da) \\
&\quad \leq T \sum_{n\geq 0} \sup_{t\geq 0} \mathbb{E}\left|\lambda^{n+1}(t)-\lambda^{n}(t)\right| \\
&\quad < +\infty,
\end{aligned}$$

for all $T > 0$. Using Borel-Cantelli lemma again we can conclude that the processes N^n are constant on $\langle 0,T]\times\mathbb{M}$ for large n. In this sense, they converge to a limiting point process N as $n \to \infty$. The limiting process N is, as the limit of the N^n, $\mathcal{F}_t$−adapted (see [35], page 27). Next, we have to see that N satisfies

$$N(dt, da) = M(dt, da, [0, \lambda(t)]).$$

Indeed, by Fatou's lemma and Lemma (2.2.1)

$$\begin{aligned}
&\mathbb{E}\int_{\langle 0,T]\times\mathbb{M}} |N(dt, da) - M(dt, da, [0, \lambda(t)])| \\
&\leq \liminf_{n\to\infty} \mathbb{E}\int_{\langle 0,T]\times D} |N^n(dt, da) - M(dt, da, [0, \lambda(t)])| \\
&= \liminf_{n\to\infty} \mathbb{E}\int_{\langle 0,T]\times D} |M(dt, da, [0, \lambda^n(t)]) - M(dt, da, [0, \lambda(t)])| \\
&\leq T \liminf_{n\to\infty} \sup_{t\geq 0} \mathbb{E}\,|\lambda^n(t) - \lambda(t)| = 0.
\end{aligned}$$

The last equality is due to (2.14). Let us check that

$$\lambda(t) = \Psi\left(\theta_t N^{<0}\right).$$

This will enable us to conclude that N is a solution to the Poisson–driven SDE defined in 2.2.20.

$$\begin{aligned}
&\mathbb{E}\left|\lambda(t) - \Psi\left(\theta_t N^{<0}\right)\right| \\
&\leq \mathbb{E}\,|\lambda(t) - \lambda^n(t)| + \mathbb{E}\int_{\langle 0,t]\times\mathbb{M}} h(t-s, a)|N - N^{n-1}|(ds, da) \\
&\leq \mathbb{E}\,|\lambda(t) - \lambda^n(t)| + \int_{\langle 0,t]\times\mathbb{M}} h(s, a) ds Q(da) \mathbb{E}|\lambda(s) - \lambda^{n-1}(s)|.
\end{aligned}$$

The right-hand side of this inequality goes to 0 uniformly in $t > 0$, hence N is a solution.

□

Uniqueness

Since Massoulié in [33] considers point processes on $\mathbb{R} \times \mathbb{M}$ which are not necessarily non-explosive marked point processes, he uses the Lipschitz condition to obtain strong uniqueness in a space of regular point processes. Morariu–Patrichi and Pakkanen in [35] restricted themselves to non-explosive marked point processes, the enumeration representation allows us to prove strong uniqueness without any specific assumptions.

2.2.4 Theorem (Theorem 2.20 in [35]). *Let $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ and $N' : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathbb{M}}$ be two non-explosive marked point processes solving the Poisson driven SDE (2.2.20). Then $N = N'$ a.s.*

Proof. Let $\Omega' \in \mathcal{F}$ be the almost sure event that both N and N' solve SDE 2.2.20. Let $(T_n, A_n)_{n\in\mathbb{N}}$ and $(T'_n, A'_n)_{n\in\mathbb{N}}$ be the enumerations in $\langle 0,\infty] \times \mathbb{M}$ to which N and N' are respectively equivalent. Now fix arbitrary $\omega \in \Omega'$. We show by strong induction that $T_n(\omega) = T'_n(\omega)$ and $A_n(\omega) = A'_n(\omega)$ for all $n \in \mathbb{N}$.
Let $n \in \mathbb{N}$ and assume that $T_i(\omega) = T'_i(\omega)$ and $A_i(\omega) = A'_i(\omega)$ for all $i = 1,\dots,n-1$. By contradiction, assume that $T_n(\omega) \neq T'_n(\omega)$ and, without loss of generality, that $T_n(\omega) < T'_n(\omega)$. This implies that

$$N\left(\omega, \langle 0, T_n(\omega)] \times \mathbb{M}\right) = \int_{\langle 0,T_n(\omega)]}\int_{\mathbb{M}}\int_{\langle 0,\lambda_n(\omega,t,a)]} M(\omega,dt,da,dz) = n,$$

$$N'\left(\omega, \langle 0, T_n(\omega)] \times \mathbb{M}\right) = \int_{\langle 0,T_n(\omega)]}\int_{\mathbb{M}}\int_{\langle 0,\lambda'_n(\omega,t,a)]} M(\omega,dt,da,dz) = n-1,$$

where $\lambda(\omega,t,a) = \Psi(\theta_t N(\omega)^{<0})$ and $\lambda'(\omega,t,a) = \Psi(\theta_t N'(\omega)^{<0})$. But since $N(\omega)^{\leq 0} = N'(\omega)^{\leq 0}$ and also $T_i(\omega) = T'_i(\omega)$ and $A_i(\omega) = A'_i(\omega)$ for all $i = 1,\dots,n-1$, we have that $\theta_t N(\omega)^{<0} = \theta_t N'(\omega)^{<0}$ for all $t \leq T_n(\omega)$. Thus, $\lambda(\omega,t,a) = \lambda'(\omega,t,a)$ for all $t \leq T_n(\omega)$, $a \in \mathbb{M}$. This implies that $n = n-1$ which is a contradiction. So, $T_n(\omega) = T'_n(\omega)$.
Similarly, if we assume that $A_n(\omega) \neq A'_n(\omega)$, then this implies that

$$N\left(\omega, \{T_n(\omega)\} \times \{A_n(\omega)\}\right) = \int_{\{T_n(\omega)\}}\int_{\{A_n(\omega)\}}\int_{\langle 0,\lambda_n(\omega,t,a)]} M(\omega,dt,da,dz) = 1,$$

$$N'\left(\omega, \{T_n(\omega)\} \times \{A_n(\omega)\}\right) = \int_{\{T_n(\omega)\}}\int_{\{A_n(\omega)\}}\int_{\langle 0,\lambda'_n(\omega,t,a)]} M(\omega,dt,da,dz) = 0.$$

Now, since $\lambda(\omega,t,a) = \lambda'(\omega,t,a)$ for all $t \leq T_n(\omega)$, $a \in \mathbb{M}$, this leads to contradiction $1 = 0$ and, thus, it follows that $A_n(\omega) = A'_n(\omega)$. The same reasoning allows us to prove the basis of the strong induction (i.e. to show that $T_1(\omega) = T'_1(\omega)$ and $A_1(\omega) = A'_1(\omega)$). □

2.2.5 Theorem (Theorem 2.21 in [35])**.** *Let N_1 and N_2 be two non-explosive marked point processes (possibly on distinct probability spaces) that admit the same intensity functional Ψ on $\mathbb{R}_{>0}$. Assume also that both N_1 and N_2 satisfy the weak initial condition $N_{\leq 0}$. Then we have that $\mathcal{P}^{N_1} = \mathcal{P}^{N_2}$, i.e. the induced probabilities measures on $\mathcal{N}^{\#}_{\mathbb{R}\times\mathbb{M}}$ coincide.*

2.2.3 Hawkes processes – Poisson cluster representation

In the previous section we presented some existing results due to Morariu-Patrichi and Pakkanen [35] and Brémaud [8] which show that under certain assumptions the marked Hawkes process exists and is unique. Hawkes process is the key example in our analysis and is typically introduced through its stochastic intensity (2.5). On the other hand, the model we used throughout this thesis is the marked Poisson cluster process and all our results presented in Chapter 3 and 4 are under this assumption.

So, to be able to apply those results to the marked Hawkes process we need to understand the connection between the general marked Poisson cluster process and the marked Hawkes process (which, at this point, is not clear). It turns out that Hawkes processes have a Poisson cluster representation which is due to [23] (for the non-marked case). In the sequel, we will present the Poisson cluster representation for the marked Hawkes processes.

Underlying Poisson process is homogenous with rate $\nu > 0$ (the so-called ground Process N^0) and clusters are formed by "Poissonian cascades." Let

$$N^0 = \sum_{i \geq 1} \delta_{\Gamma_i, A_i} .$$

be a Poisson point process with intensity $\nu \times Q$ on the space $[0, \infty) \times \mathbb{M}$.

Consider the process G^{A_i} (representing a cluster of points that is superimposed on N^0 after time Γ_i) as a part of the mark attached to N^0 at time Γ_i. Indeed,

$$\sum_{i \geq 1} \delta_{\Gamma_i, A_i, G^{A_i}}$$

can be viewed as a marked Poisson process on $[0, \infty)$ with marks in the space $\mathbb{M} \times \mathcal{N}^{\#g}_{[0,\infty\rangle \times \mathbb{M}}$. We can write

$$G^{A_i} = \sum_{j=1}^{K_i} \delta_{T_{ij}, A_{ij}} ,$$

where $(T_{ij})_{j \geq 1}$ is a sequence of non-negative random variables and for some $\mathbb{N}_0$ valued random variable K_i. If we count the original point arriving at time Γ_i, the actual cluster size is $K_i + 1$.

For this model, the clusters G^A are recursive aggregation of Cox processes, i.e. Poisson processes with random mean measure $\tilde{\mu}_A \times Q$ where $\tilde{\mu}_A$ has the following form

$$\tilde{\mu}_A(B) = \int_B h(s, A) ds , \tag{2.15}$$

for some fertility (or self–exciting) function h, cf. Example 6.4 (c) of [12].

Now, for the ground process $N^0 = \sum_{i \geq 1} \delta_{\Gamma_i, A_i}$ which is a Poisson point process with intensity $\nu \times Q$ on the space $[0, \infty) \times \mathbb{M}$, the cluster process corresponding to a point (Γ, A) satisfies the following recursive relation

$$G^A = \sum_{l=1}^{L_A} \left(\delta_{\tau_l^1, A_l^1} + \theta_{\tau_l^1} G^{A_l^1} \right) , \tag{2.16}$$

where, given A, $N^A = \sum_{l=1}^{L_A} \delta_{\tau_l^1, A_l^1}$ is a Poisson processes with random mean measure $\tilde{\mu}_A \times Q$, the sequence $(G^{A_l^1})_l$ is i.i.d., distributed as G^A, independent of N^A and θ is the time shift operator introduced in Definition 2.2.4. Thus, at any ancestral point (Γ, A) a cluster of points appears as a whole cascade of points to the right in time generated recursively according to (2.16). Note that by definition L_A has Poisson distribution conditionally on A, with mean $\kappa_A = \int_0^\infty h(s, A)ds$. It corresponds to the number of the first generation progeny (A_l^1) in the cascade. Note also that the point processes forming the second generation are again Poisson conditionally on the corresponding first generation mark A_l^1. The cascade G^A corresponds to the process formed by the successive generations, drawn recursively as Poisson processes given the former generation.

The marked Hawkes process is obtained by attaching to the ancestors (Γ_i, A_i) of the marked Poisson process $N^0 = \sum_{i \geq 1} \delta_{\Gamma_i, A_i}$ a cluster of points, denoted by C_i, which contains point $(0, A_i)$ and a whole cascade G^{A_i} of points to the right in time generated recursively according to (2.16) given A_i. Under the assumption

$$\kappa = \mathbb{E} \int h(s, A)ds < 1\,, \tag{2.17}$$

the total number of points in a cluster is generated by a subcritical branching process. Therefore, the clusters are finite almost surely, and we denote their size by K_i+1. It is known and not difficult to show that under the assumption that $\kappa < 1$, the clusters always satisfy

$$\mathbb{E}K_i+1 = \frac{1}{1-\kappa}\,.$$

Observe that the individual clusters are independent by construction and can be represented as

$$C_i := \sum_{j=0}^{K_i} \delta_{\Gamma_i + T_{ij}, A_{ij}}\,, \tag{2.18}$$

with A_{ij} being i.i.d. and $T_{i0} = 0$. We note that in the case when marks do not influence conditional density, i.e. when $h(s, a) = h(s)$, random variable K_i+1 has a so-called Borel distribution with parameter κ, see [21]. Observe also that in general, marks and arrival times of the final Hawkes process N are not independent of each other, rather, in the terminology of [12], the marks in the process N are only unpredictable which means that the distribution of the mark at T_i is independent of locations and marks $\{(T_j, A_j)\}$ for which $T_j < T_i$.

3

On total claim amount for marked Poisson cluster models

3.1 Introduction

The main goal of this chapter is to study asymptotic distribution of the total claim amount in the setting where Cramér–Lundberg risk model is augmented with a Poisson cluster structure. To make this more precise, we model arrival of claims in an insurance portfolio by a marked point process, say

$$N = \sum_{k=1}^{\infty} \delta_{\tau_k, A^k} ,$$

where τ_k's are nonnegative random variables representing arrival times with some degree of clustering and A^k's represent corresponding marks in a rather general metric space $\mathbb{M}$. Observe that we do allow for the possibility that marks influence arrival rate of the future claims. In the language of point processes theory, we assume that the marks are merely unpredictable and not independent of the arrival times [12]. For each marked event, the claim size can be calculated using a measurable mapping of marks to nonnegative real numbers, $f(A^k)$ say. So that the total claim amount in the time interval $[0, t]$ can be calculated as

$$S(t) = \sum_{\tau_k \le t} f(A^k) = \int_{[0,t] \times \mathbb{M}} f(a) N(ds, da) .$$

In the sequel, we aim to determine the effect of the clustering on the quantity $S(t)$, as $t \to \infty$ even in the case when the distribution of the individual claims does not satisfy assumptions of the classical central limit theorem. The section is organized as follows — in the following section we rigorously introduce marked Poisson cluster model and present some specific cluster models which have attracted attention in the related literature, see [17, 28, 42]. As a proposition in Section 3.3 we present the central limit theorem for the total claim amount $S(t)$ in our setting under appropriate second moment conditions. In Section 3.4, we prove a functional limit theorem concerning the sums of regularly varying non-negative random variables when subordinated to an independent renewal process. Based on this, we prove the limit theorem for the total claim amount $S(t)$ in cases when individual claims have infinite variance. Finally in Section 3.5 we apply our results to the models we introduced in Section 3.2. In particular, we give a detailed analysis of the asymptotic behaviour of $S(t)$ for marked Hawkes processes which have been extensively studied in recent years.

3.2 The general marked Poisson cluster model

Consider an independently marked homogeneous Poisson point process with mean measure (νLeb) on the state space $[0, \infty)$ for some constant $\nu > 0$, where Leb denotes Lebesgue measure on $[0, \infty)$, with marks in a completely metrizable (i.e. metrizable with complete metric) separable space. $\mathbb{M}$,

$$N^0 = \sum_{i \geq 1} \delta_{\Gamma_i, A_i} .$$

Marks A_i are assumed to follow a common distribution Q on a measurable space $(\mathbb{M}, \mathcal{B}(\mathbb{M}))$ where $\mathcal{B}(\mathbb{M})$ denotes a corresponding Borel σ–algebra. In other words, N^0 is a Poisson point process with intensity $\nu \times Q$ on the space $[0, \infty) \times \mathbb{M}$. For non–life insurance modelling purposes, the marks can take values in $\mathbb{R}^d$ with coordinates representing the size of claim, type of claim, severity of accident, etc.

Recall, $\mathcal{N}^{\#g}_{\mathbb{R}_{\geq 0} \times \mathbb{M}}$ is the space of boundedly finite point measures such that their ground measure is non-explosive and simple. Assume that at each time Γ_i with mark A_i another point process in $\mathcal{N}^{\#g}_{\mathbb{R}_{\geq 0} \times \mathbb{M}}$ is generated independently, we denote it by G^{A_i}. Intuitively, point process G^{A_i} represents a cluster of points that is superimposed on N^0 after time Γ_i. Formally, there exists a probability kernel K, from $\mathbb{M}$ to $\mathcal{N}^{\#g}_{\mathbb{R}_{\geq 0} \times \mathbb{M}}$, such that, conditionally on N^0, point processes G^{A_i} are independent, a.s. finite and with the distribution equal to $K(A_i, \cdot)$, thus the dependence between the G^{A_i} and A_i is permitted. Based on N^0 and clusters G^{A_i} we define a cluster Poisson process.

In order to keep the track of the cluster structure, we can alternatively consider the process G^{A_i} as a part of the mark attached to N^0 at time Γ_i. Indeed,

$$\sum_{i \geq 1} \delta_{\Gamma_i, A_i, G^{A_i}}$$

can be viewed as a marked Poisson process on $[0, \infty)$ with marks in the space $\mathbb{M} \times \mathcal{N}^{\#g}_{\mathbb{R}_{\geq 0} \times \mathbb{M}}$. We can write

$$G^{A_i} = \sum_{j=1}^{K_i} \delta_{T_{ij},A_{ij}} \,,$$

where $(T_{ij})_{j\geq 1}$ is a sequence of non-negative random variables and for some $\mathbb{N}_0$ valued random variable K_i. If we count the original point arriving at time Γ_i, the actual cluster size is $K_i + 1$. Further, for any original arrival point Γ_i and corresponding random cluster G^{A_i}, we introduce a point process

$$C_i = \delta_{0,A_i} + G^{A_i} \,.$$

Note that K_i may possibly depend on A_i, but we do assume throughout that

$$\mathbb{E}K_i < \infty \,.$$

Finally, to describe the size and other characteristics of the claims together with their arrival times, we use a marked point process N as a random element in $\mathcal{N}^{\#g}_{\mathbb{R}_{\geq 0}\times\mathbb{M}}$ of the form

$$N = \sum_{i=1}^{\infty} \sum_{j=0}^{K_i} \delta_{\Gamma_i+T_{ij},A_{ij}} \,, \tag{3.1}$$

where we set $T_{i0} = 0$ and $A_{i0} = A_i$. In this representation, the claims arriving at time Γ_i and corresponding to the index $j = 0$ are called ancestral or immigrant claims, while the claims arriving at times $\Gamma_i + T_{ij}$, $j \geq 1$, are referred to as progeny or offspring. Moreover, since N is locally finite, one could also write

$$N = \sum_{k=1}^{\infty} \delta_{\tau_k,A^k} \,,$$

with $\tau_k \leq \tau_{k+1}$ for all $k \geq 1$. Note that in this representation we ignore the information regarding the clusters of the point process. Clearly, if the cluster processes G^{A_i} are independently marked with the same mark distribution Q independent of A_i, then all the marks A^k are i.i.d.

The size of claims is produced by an application of a measurable function, say $f : \mathbb{M} \to \mathbb{R}_+$, on the marks. In particular, sum of all the claims due to the arrival of an immigrant claim at time Γ_i equals

$$D_i = \int_{[0,\infty)\times\mathbb{M}} f(a) C_i(dt, da) \,, \tag{3.2}$$

while the total claim size in the period $[0, t]$ can be calculated as

$$S(t) = \sum_{\tau_k \leq t} f(A^k) = \int_{[0,t]\times\mathbb{M}} f(a) N(ds, da) \,.$$

3.2.1 Remark. *In all our considerations, we take into account (without any real loss of generality) the original immigrant claims arriving at times Γ_i as well. In principle, one could ignore these claims and treat Γ_i as times of incidents that trigger, with a possible delay, a cluster of subsequent payments. Such a choice seems particularly useful if one aims to model the so called incurred but not reported (IBNR) claims, when estimating appropriate reserves in an insurance portfolio [34]. In such a case, in the definition of the process N, one would omit the points of the original Poisson process N^0 and consider*

$$N = \sum_{i=1}^{\infty} \sum_{j=1}^{K_i} \delta_{\Gamma_i + T_{ij}, A_{ij}},$$

instead.

3.2.1 Some special models

Several examples of Poisson cluster processes have been studied in the monograph [12], see Example 6.3 therein for instance. Here we study marked adaptation of the first three examples 6.3 (a)-(b) and (c) of [12].

Mixed binomial Poisson cluster process.

Assume that the clusters have the following form

$$G^{A_i} = \sum_{j=1}^{K_i} \delta_{W_{ij}, A_{ij}},$$

with $(K_i, (W_{ij})_{j\geq 1}, (A_{ij})_{j\geq 0})_{i\geq 0}$ being an i.i.d. sequence. Assume moreover that $(A_{ij})_{j\geq 0}$ are i.i.d. for any fixed $i = 1, 2, \ldots$ and that $(A_{ij})_{j\geq 1}$ is independent of $K_i, (W_{ij})_{j\geq 1}$ for all $i \geq 0$. We allow for possible dependence between $K_i, (W_{ij})_{j\geq 1}$ and the ancestral mark A_{i0}, however, we assume that K_i and $(W_{ij})_{j\geq 1}$ are conditionally independent given A_{i0}. As before we assume $\mathbb{E}[K] < \infty$. Observe that we use notation W_{ij} instead of T_{ij} to emphasize relatively simple structure of clusters in this model in contrast with two other models in this section. Such a process N is a version of the so–called Neyman–Scott process, e.g. see Example 6.3 (a) of [12].

Renewal Poisson cluster process.

Assume next that the clusters G^{A_i} have the following distribution

$$G^{A_i} = \sum_{j=1}^{K_i} \delta_{T_{ij}, A_{ij}},$$

where $(T_{ij})_j$ represents a renewal sequence

$$T_{ij} = W_{i1} + \cdots + W_{ij} \,,$$

and we keep all the other assumptions from the model in subsection 3.2.1 (in particular, $(W_{ij})_{j\geq 1}$ are conditionally i.i.d. and independent of K_i given A_{i0}). A general unmarked model of this type is called Bartlett—Lewis model and analysed in [12], see Example 6.3 (b). See also [17] for an application of such a point process to modelling of teletraffic data.

These two simple cluster models were already considered by [34] in the context of insurance applications. In particular, subsection 11.3.2 and example 11.3.5 therein provide expressions for the first two moments of the number of claims in a given time interval $[0, t]$. Both models can be criticized as overly simple, still the assumption that claims (or delayed payouts) are separated by i.i.d. times (as in the renewal Poisson cluster process) often appears in the risk theory (cf. Sparre Andersen model, [2]).

Marked Hawkes process.

Key motivating example in our analysis is the so called (linear) marked Hawkes process. Hawkes processes of this type have a neat Poisson cluster representation due to [23] which we presented in Section 2.2.3.

Recal, Hawkes processes are typically introduced through their conditional intensity. More precisely, a point process $N = \sum_k \delta_{\tau_k, A^k}$, represents a Hawkes process of this type if the random marks (A^k) are i.i.d. with distribution Q on the space $\mathbb{M}$, while the arrivals (τ_k) have the conditional intensity of the form

$$\lambda_t = \lambda(t) = \nu + \sum_{\tau_i < t} h(t - \tau_i, A^i) \,,$$

where $\nu > 0$ is a constant and $h : [0, \infty) \times \mathbb{M} \to \mathbb{R}_+$ is assumed to be integrable in the sense that $\int_0^\infty \mathbb{E}h(s, A)ds < \infty$. Observe, ν is exactly the constant which determines the intensity of the underlying Poisson process N^0 due to the Poisson cluster representation of the linear Hawkes processes, cf. [23]. Observe, λ is $\mathcal{F}_t$–predictable, where $\mathcal{F}_t$ stands for an internal history of N, $\mathcal{F}_t = \sigma\{N(I \times S) : I \in \mathcal{B}(\mathbb{R}), I \subset (-\infty, t], S \in \mathcal{B}(\mathbb{M})\}$. Moreover, A^n's are assumed to be independent of the past arrival times τ_i, $i < n$, see also [8]. Writing $N_t = N((0, t] \times \mathbb{M})$, one can observe that (N_t) is an integer valued process with non-decreasing paths. The role of intensity can be described heuristically by the relation

$$\mathbb{P}(dN_t = 1 \mid \mathcal{F}_{t-}) \approx \lambda_t dt \,.$$

Stationary version.

In any of the three examples above, the point process N can be clearly made stationary if we start the construction in (3.1) on the state space $\mathbb{R} \times \mathbb{M}$ with a Poisson process $\sum_i \delta_{\Gamma_i}$ on the whole real line. The resulting stationary cluster process is denoted by N^*. Still, from applied perspective, it seems more interesting to study the nonstationary version where both the ground process N^0 and the cluster process itself have arrivals only from some point onwards, e.g. in the interval $[0, \infty)$ as for instance in [28].

Stability of various cluster models, i.e. convergence towards a stationary distribution in appropriate sense has been extensively studied for various point processes. For instance, it is known that the unmarked Hawkes process on $[0, \infty)$ converges to the stationary version on any compact set and on the positive line under the condition that

$$\int_0^\infty sh(s)ds < \infty\,, \tag{3.3}$$

see [12], p. 232. Using the method of Poisson embedding, originally due to [29], [9] (Section 3) obtained general results on stability of Hawkes processes, even in the non–linear case.

3.3 Central limit theorem

As explained in Section 2, the total claim amount for claims, arriving before time t, can be written as

$$S(t) = \sum_{\tau_k \leq t} f(A_k) = \int_0^t \int_{\mathbb{M}} f(u) N(ds, du)\,.$$

The long term behaviour of $S(t)$ for general marked Poisson cluster processes is the main goal of our study. As before, by Q we denote the probability distribution of marks on the space $\mathbb{M}$.

Moreover, unless stated otherwise, we assume that the process starts from 0 at time $t = 0$, that is $N(-\infty, 0] = 0$.

In the case of the Hawkes process, the process $N_t = N([0,t] \times \mathbb{M})$, $t \geq 0$ which only counts the arrival of claims until time t has been studied in the literature before. It was shown recently under appropriate moment conditions, that in the unmarked case multi-type Hawkes processes satisfy (functional) central limit theorem, see [3]. Karabash and Zhu [28] showed that N_t satisfies central limit theorem even in the more general case of non-linear Hawkes process and that linear but marked

Hawkes have the same property. In the present section we describe the asymptotic behaviour of the total claim amount process $(S(t))$ for a wide class of marked Poisson cluster processes, even in the case when the total claim process has heavy tails, and potentially infinite variance or infinite mean.

It is useful in the sequel to introduce random variable

$$\tau(t) = \inf\{n : \Gamma_n > t\}\,,\ t \geq 0\,.$$

Recall from (3.2) the definition of D_i as

$$D_i = \int_{[0,\infty)\times\mathbb{M}} f(u)C_i(ds, du) = \sum_{j=0}^{K_i} f(A_{ij}) = \sum_{j=0}^{K_i} X_{ij}\,,$$

where $K_i + 1 = C_i[0, \infty)$ denotes the size of the ith cluster and where we denote $X_{ij} = f(A_{ij})$. As before, D_i has an interpretation as the total claim amount coming from the ith immigrant and its progeny. Note that D_i's form an i.i.d. sequence because the ancestral mark in every cluster comes from an independently marked homogeneous Poisson point process.

Observe that in the nonstationary case we can write

$$S(t) = \sum_{i=1}^{\tau(t)} D_i - D_{\tau(t)} - \varepsilon_t\,,\ t \geq 0\,, \tag{3.4}$$

where the last error term represents the leftover or the residue at time t, i.e. the sum of all the claims arriving after t which belong to the progeny of immigrants arriving before time t, that is

$$\varepsilon_t = \sum_{0\leq\Gamma_i\leq t,\, t<\Gamma_i+T_{ij}} f(A_{ij}) \quad t \geq 0\,.$$

Clearly, in order to characterize limiting behaviour of $S(t)$, it is useful to determine moments and the tail behaviour of random variables D_i for each individual cluster model. To simplify the notation, for a generic member of an identically distributed sequence or an array, say (D_n), (A_{ij}), we write D, A etc. Under the conditions of existence of second order moments and the behaviour of the residue term ε_t, it is not difficult to derive the following proposition.

3.3.1 Proposition. *Assume the marked Poisson cluster model defined in Section 2. Suppose that $\mathbb{E}D^2 < \infty$ and that $\varepsilon_t = o_P(\sqrt{t})$ then, for $t \to \infty$,*

$$\frac{S(t) - t\nu\mu_D}{\sqrt{t\nu\mathbb{E}D^2}} \xrightarrow{d} N(0, 1)\,, \tag{3.5}$$

where $\mu_D = \mathbb{E}D$ and $\nu > 0$ is the constant which determines the intensity of the underlying Poisson point process N^0.

Proof. Recall, $\varepsilon_t = o_P(\sqrt{t})$ means that $\lim_{t\to\infty} \frac{\varepsilon_t}{\sqrt{t}} = 0$.
Denote the first term on the r.h.s. of (3.4) by

$$S^D(t) = \sum_{i=1}^{\tau(t)} D_i \ t \geq 0 \, .$$

An application of the central limit theorem for two-dimensional random walks, see [19, Section 4.2, Theorem 2.3] yields

$$\frac{S^D(t) - t\nu\mu_D}{\sqrt{t\nu \mathbb{E} D^2}} \xrightarrow{d} N(0,1) \, ,$$

as $t \to \infty$. Since we assumed $\varepsilon_t / \sqrt{t} \xrightarrow{P} 0$, it remains to show that

$$\frac{D_{\tau(t)}}{\sqrt{t}} \xrightarrow{P} 0 \quad t \to \infty.$$

However, this follows at once from [19, Theorem 1.2.3] for instance, or from the fact that in this setting sequences (Γ_n) and (D_n) are independent.

□

Note that Proposition 3.3.1 holds for f taking possibly negative values as well. However, when modelling insurance claims, non-negativity assumption seems completely natural, and in the heavy tail case our proofs actually depend on it, cf. the proof of Theorem 3.4.1. In the special case $f \equiv 1$, one obtains the central limit theorem for the number of arrivals in time interval $[0, t]$. Related results have appeared in the literature before, see for instance [11] or [28] . The short proof above stems from the classical Anscombe's theorem, as presented in Gut [19, Chapter IV] (cf. [11, Theorem 3 ii]) unlike the argument in [28] which relies on martingale central limit theorem and seems not easily extendible, especially for heavy tailed claims we consider next.

3.3.1 Remark. *It is not too difficult to find examples where the residue term is not negligible. Consider renewal cluster model of subsection 3.2.1 with $K = 1$, $X = 1$. Let W_{i1} be i.i.d. and regularly varying with index $\alpha < 1/2$. Then ε_t has Poisson distribution with parameter $\mathbb{E}[W\mathbb{I}_{W<t}] \to \infty$ and thus, by Karamata's theorem, $\varepsilon_t/\sqrt{t}$ tends to infinity in probability. Similarly, one can show that $\mathrm{Var}(\varepsilon_t/\sqrt{t}) = \mathbb{E}[W\mathbb{I}_{W<t}]/t \to 0$ so that $(\varepsilon_t - \mathbb{E}[W\mathbb{I}_{W<t}])/\sqrt{t}$ tends to zero in probability. Thus (3.5) does not hold any more but instead we have*

$$\frac{S(t) - t\nu\mu_D + \mathbb{E}[W\mathbb{I}_{W<t}]}{\sqrt{t\nu \mathbb{E} D^2}} \xrightarrow{d} N(0,1) \, , \qquad t \to \infty \, .$$

3.4 Infinite variance stable limit

It is known that if the claims are sufficiently heavy tailed, properly scaled and centred sums $S(t)$ may converge to an infinite variance stable random variable. In the case of random sums $S_n = X_1 + \cdots + X_n$ of i.i.d. random variables, the corresponding statement is true if and only if the claims are regularly varying with index $\alpha \in (0,2)$. For the Cramér–Lundberg model, i.e. when $N = N_0$, with i.i.d. regularly varying claims of index $\alpha \in (1,2)$, corresponding limit theorem follows from Theorem 4.4.3 in [19]. A crucial step in the investigation of the heavy tailed case is to determine the tail behaviour of the random variables of (3.2).

For regularly varying D_i with index $\alpha \in (1,2)$, limit theory for two-dimensional random walks in Section 4.2 of [19] still applies. Note, if one can show that D_i's have regularly varying distribution, then there exists a sequence (a_n) , $a_n \to \infty$, such that

$$nP(D > a_n) \to 1\,, \qquad n \to \infty,$$

and an α–stable random variable G_α such that $S_n^D = D_1 + \cdots + D_n\,, \qquad n \to \infty$, satisfies

$$\frac{S_n^D - n\mu_D}{a_n} \xrightarrow{d} G_\alpha\,, \tag{3.6}$$

where $\mu_D = \mathbb{E}D_i$. It is also known that the sequence (a_n) is regularly varying itself with index $1/\alpha$, see [38]. In the sequel ,we also set $a_t = a_{\lfloor t \rfloor}$ for any $t \geq 1$.

3.4.1 Case $\alpha \in (1,2)$

In this case, the arguments of the previous section can be adopted to show.

3.4.1 Proposition. *Assume the marked Poisson cluster model introduced in Section 2. Suppose that D_i's are regularly varying with index $\alpha \in (1,2)$ and that $\varepsilon_t = o_P(a_t)$, then there exists an α–stable random variable G_α such that for $\mu_D = \mathbb{E}D_i$*

$$\frac{S(t) - t\nu\mu_D}{a_{\nu t}} \xrightarrow{d} G_\alpha\,, \tag{3.7}$$

as $t \to \infty$.

Proof. The proof again follows from the representation (3.4), by an application of Theorem 4.2.6 from [19] on random walks (Γ_n) and (S_n^D) together with relation (3.6). By assumption we have $\varepsilon_t/a_{\nu t} \sim \nu^{-1/\alpha}\varepsilon_t/a_t \xrightarrow{P} 0$. To finish the proof, we observe that the sequences (Γ_n) and (D_n) are independent, hence

$$\frac{D_{\tau(t)}}{a_{\nu t}} \xrightarrow{P} 0, \qquad t \to \infty.$$

□

3.4.2 Case $\alpha \in (0,1)$

In this case, we were not able to find any result of Anscombe's theorem type for two–dimensional random walks of the type used above. Therefore, as our initial step, we prove a theorem which we believe is new and of independent interest. It concerns partial sums of i.i.d. nonnegative regularly varying random variables, say (Y_n), subordinated to an independent renewal process. More precisely, set $V_n = Y_1 + \cdots + Y_n\,, n \geq 1$. Suppose that the sequence (Y_n) is independent of another i.i.d. sequence of nonnegative and nontrivial random variables (W_n). Denote by

$$\sigma(t) = \sup\{k : W_1 + \cdots + W_k \leq t\}$$

the corresponding renewal process, where we set $\sup \emptyset = 0$. Recall that for regularly varying random variables Y_i's there exists a normalizing sequence (a_n) such that in the central limit theorem $\frac{Y_1+\cdots+Y_n}{a_n} \xrightarrow{d} G_\alpha$, where G_α is an $\alpha-$stable distribution. The limiting behaviour of the process $V_{\sigma(t)}$ was considered by [1] in the case when $W_i's$ are themselves regularly varying with index less then or equal to 1.

Since $\sigma(t)/t \xrightarrow{as} \nu$, if $0 < \mathbb{E}W_i = 1/\nu < \infty$, one may expect that $V_{\sigma(t)}$ has similar asymptotic behaviour as $V_{\nu t}$ for $t \to \infty$. It is not too difficult to make this argument rigorous if for instance $\mathbb{E}W_i^2 < \infty$, because then $(\sigma(t) - t\nu)^2/t$, $t > 0$, is uniformly integrable by Gut (2009), Section 2.5. The following functional limit theorem gives precise description of the asymptotic behaviour of $V_{\sigma(t)}$ whenever W_i have a finite mean.

3.4.1 Theorem. *Suppose that (Y_n) and (W_n) are independent non-negative i.i.d. sequences of random variables such that Y_i's are regularly varying with index $\alpha \in (0,1)$, and such that $0 < 1/\nu = \mathbb{E}W_i < \infty$. Then in the space $D[0,\infty)$ endowed with Skorohod's J_1 topology*

$$\frac{V_{\sigma(t\cdot)}}{a_{\nu t}} \xrightarrow{d} G_\alpha(\cdot)\,, \qquad t \to \infty, \tag{3.8}$$

where $(G_\alpha(s))_{s\geq 0}$ is an α–stable subordinator.

Proof. Since Y_i's are regularly varying, it is known, [38, 39], that the following point process convergence holds as $n \to \infty$

$$Z'_n = \sum_i \delta_{\frac{i}{n},\frac{Y_i}{a_n}} \xrightarrow{d} Z_\alpha \sim \mathrm{PRM}(Leb \times d(-y^{-\alpha}))\,, \tag{3.9}$$

with respect to the vague topology on the space of Radon point measures on $[0,\infty) \times (0,\infty]$. Abbreviation PRM stands for Poisson random measure indicating that the limit is a Poisson process. Starting from (3.9), it was shown in [39, Chapter 7] for instance, that for an α–stable subordinator $G_\alpha(\cdot)$ as in the statement of the theorem

$$V'_n(\cdot) = \frac{V_{\lfloor n\cdot\rfloor}}{a_n} \xrightarrow{d} G_\alpha(\cdot)\,, \qquad n \to \infty,$$

in Skorohod's J_1 topology on the space $D[0,\infty)$. Observe that since $\alpha \in (0,1)$, no centering is needed, and that one can substitute the integer index n by a continuous index $t \to \infty$. Note further that we have the joint convergence

$$(Z'_t, V'_t) \xrightarrow{d} (Z_\alpha, G_\alpha), \qquad t \to \infty, \tag{3.10}$$

in the product topology on the space of point measures and càdlàg functions. Moreover, it is known that the jump times and sizes of the α–stable subordinator G_α correspond to the points of the limiting point process Z_α .

The space of point measures and the space of càdlàg functions $D[0,\infty)$ are both Polish, in respective topologies, therefore, Skorohod's representation theorem applies. Thus, we can assume that convergence in (3.10) holds a.s. on a certain probability space $(\Omega, \mathcal{F}, P)$, and in particular there exists $\Omega' \subseteq \Omega$, such that $P(\Omega') = 1$ and for all $\omega \in \Omega'$, $V'_t \to G_\alpha$ in J_1 and $Z'_t \to Z_\alpha$ in vague topology. By Chapter VI, Theorem 2.15 in [25], for any such ω there exists a dense set $B = B(\omega)$ of points in $[0,\infty)$ such that

$$V'_t(s) \to G_\alpha(s)\,, \qquad t \to \infty,$$

for every $s \in B$, where actually B is simply the set of all nonjump times in the path of the process G_α. On the other hand, it is known that in J_1 topology, on some set Ω'' such that $P(\Omega'') = 1$,

$$\frac{\sigma(t\cdot)}{t\nu} \to id(\cdot)\,, \qquad t \to \infty, \tag{3.11}$$

where id stands for the identity map. This follows directly by an application of Theorem 2.15 in Chapter VI of [25]. Moreover, by Proposition VI.1.17 in [25], the convergence in (3.11) holds locally uniformly on $D[0,\infty)$.

Consider now for fixed $t > 0$ and $\omega \in \Omega' \cap \Omega''$

$$V_t(s) = \frac{V_{\sigma(ts)}}{a_{\nu t}}\,, \quad s \geq 0\,.$$

From (3.11) we may expect that $V_t(s) \approx V'_{t\nu}(s)$. Indeed, for any fixed $0 < \delta < 1$ and all large t, we know that $\lfloor tc\nu(1-\delta)\rfloor \leq \sigma(tc) \leq \lfloor tc\nu(1+\delta)\rfloor$, which by monotonicity of the sums implies

$$\frac{V_{\lfloor tc\nu(1-\delta)\rfloor}}{a_{\nu t}} \leq \frac{V_{\sigma(tc)}}{a_{\nu t}} \leq \frac{V_{\lfloor tc\nu(1+\delta)\rfloor}}{a_{\nu t}}\,.$$

Now, for $c(1-\delta)$ and $c(1+\delta)$ in B, the left hand side and the right hand side above converge to $G_\alpha(c(1-\delta))$ and $G_\alpha(c(1+\delta))$. Thus, if we consider $c \in B$ and let $\delta \to 0$, then

$$\frac{V_{\sigma(tc)}}{a_{\nu t}} \to G_\alpha(c)\,, \qquad t \to \infty, \tag{3.12}$$

for all $\omega \in \Omega' \cap \Omega''$ and thus with probability 1.

By Theorem 2.15 in Chapter VI in [25], to prove (3.8), it remains to show that for all $\omega \in \Omega' \cap \Omega''$ and $c \in B$, as $t \to \infty$

$$\sum_{0<s\leq c} |\Delta V_t(s)|^2 = \sum_{i<\sigma(tc)} \left(\frac{Y_i}{a_{t\nu}}\right)^2 \to \sum_{0<s\leq c} |\Delta G_\alpha(s)|^2 , \tag{3.13}$$

where, for an arbitrary càdlàg process $X(t)$ at time $t \geq 0$, we denote $\Delta X(t) = X_t - X_{t-}$. Observe that

$$G_{\alpha/2}(c) := \sum_{0<s\leq c} |\Delta G_\alpha(s)|^2$$

defines an $\alpha/2$–stable subordinator and that the squared random variables Y_i^2 are again regularly varying with index $\alpha/2$ with the property that $n\mathbb{P}(Y_i^2 > a_n^2) \to 1$. A similar approximation argument as for (3.12) shows that (3.13) indeed holds, which concludes the proof. □

Assume now that $P(D > x) = x^{-\alpha}\ell(x)$ for some slowly varying function ℓ and $\alpha \in (0,1)$. Select a sequence $a_n \to \infty$ such that $nP(D > a_n) \to 1$, as $n \to \infty$. Under suitable conditions on the residue term ε_t we obtain the following.

3.4.2 Proposition. *Assume that D_i's are regularly varying with index $\alpha \in (0,1)$ and that $\varepsilon_t = o_P(a_t)$. Then, there exists an α–stable random variable G_α such that*

$$\frac{S(t)}{a_{\nu t}} \xrightarrow{d} G_\alpha , \tag{3.14}$$

as $t \to \infty$.

Proof. The proof follows roughly the same lines as the proof of Proposition 3.4.1, but here we rely on an application of the previous theorem to the random walks (Γ_n) and (S_n^D). Just, instead of Y_i's and W_i's we have D_i's and an independent sequence of i.i.d. exponential random variables with parameter ν.

□

3.4.1 Remark. *One can consider total claim amount in the period $[0,t]$ for the stationary model of subsection 3.2.1, i.e.*

$$S^*(t) = \int_0^t \int_{\mathbb{M}} f(u) N^*(ds,du) , \qquad t \geq 0 .$$

Here again, $S^*(t)$ has a similar representation as in (3.4) but with an additional term on the right hand side, i.e.

$$S^*(t) = \sum_{i=1}^{\tau(t)} D_i - D_{\tau(t)} - \varepsilon_t + \varepsilon^*_{0,t} , \qquad t \geq 0 ,$$

where

$$\varepsilon^*_{0,t} = \sum_{\Gamma_i \le 0,\ 0<\Gamma_i+T_{ij}<t} X_{ij}.$$

Clearly, by stationarity

$$\varepsilon_t = \sum_{0\le\Gamma_i\le t, t<\Gamma_i+T_{ij}} X_{ij} \stackrel{d}{=} \varepsilon^-_t = \sum_{-t\le\Gamma_i\le 0, 0<\Gamma_i+T_{ij}} X_{ij}. \tag{3.15}$$

Hence, $\varepsilon_t = o_P(a_t)$ *yields* $\varepsilon^-_t = o_P(a_t)$ *for any sequence* (a_t) *and therefore*

$$\varepsilon^*_{0,t} \le \varepsilon^-_t + \sum_{\Gamma_i<-t,\ 0<\Gamma_i+T_{ij}<t} X_{ij} = \sum_{\Gamma_i<-t,\ 0<\Gamma_i+T_{ij}<t} X_{ij} + o_P(a_t)\,.$$

In particular, conclusions of propositions 3.3.1, 3.4.1 and 3.4.2 hold for random variables $S^*(t)$ *too under the additional assumption that*

$$\tilde{\varepsilon}_t := \sum_{\Gamma_i\le -t,\ 0<\Gamma_i+T_{ij}<t} X_{ij} = o_P(a_t). \tag{3.16}$$

3.5 Total claim amount for special models

As we have seen in the previous two sections, it is relatively easy to describe asymptotic behaviour of the total claim amount $S(t)$ as long as we are able to determine the moments and tail properties of the random variables D_i and the residue random variable ε_t in (3.4) (and also $\tilde{\varepsilon}_t$ in (3.16) for the stationary version). However, this is typically a rather technical task, highly dependent on an individual Poisson cluster model. In this section we revisit three models introduced in Subsection 3.2.1, characterizing for each of them the limiting distribution of the total claim amount under appropriate conditions. Note that the cluster sum D for all three models admits the following representation

$$D \stackrel{d}{=} \sum_{j=0}^{K} X_j,$$

for $(X_j)_{j\ge 0}$ i.i.d. copies of $f(A)$ and some integer valued K such that $\mathbb{E}[K] < \infty$. Throughout, we assume that the random variables K and $(X_j)_{j\ge 1}$ are independent. The sum $\sum_{j=1}^{K} X_j$ has a so called compound distribution. Its first two moments exist under the following conditions

- if $\mathbb{E}[X] < \infty$ and $\mathbb{E}[K] < \infty$, then $\mu_D = \mathbb{E}D = (1 + \mathbb{E}[K])\mathbb{E}[X] < \infty$,
- if $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[K^2] < \infty$, then $\mathbb{E}D^2 = (\mathbb{E}[K] + 1)\mathbb{E}[X^2] + (\mathbb{E}[K^2] + \mathbb{E}[K])\mathbb{E}[X]^2 < \infty$.

The tail behaviour of compound sums was often studied under various conditions (see [13, 17, 24, 41]). We list below some of these conditions, which are applicable to our setting.

(RV1) If X is regularly varying with index $\alpha > 0$ and $\mathbb{P}(K > x) = o(\mathbb{P}(X > x))$, then $\mathbb{P}(D > x) \sim (\mathbb{E}[K] + 1)\mathbb{P}(X > x)$ as $x \to \infty$, see [17, Proposition 4.1],

(RV2) If K is regularly varying with index $\alpha \in (1, 2)$ and $\mathbb{P}(X > x) = o(\mathbb{P}(K > x))$, then $\mathbb{P}(D > x) \sim \mathbb{P}(K > x/\mathbb{E}[X])$ as $x \to \infty$, see [41, Theorem 3.2] or [17, Proposition 4.3],

(RV3) If X and K are both regularly varying with index $\alpha \in (1, 2)$ and tail equivalent, see [16, Definition 3.3.3], then $\mathbb{P}(D > x) \sim (\mathbb{E}[K] + 1)\mathbb{P}(X > x) + \mathbb{P}(K > x/\mathbb{E}[X])$ as $x \to \infty$, [13, Theorem 7].

We will refer to the last three conditions as the sufficient conditions **(RV)**.

3.5.1 Mixed binomial cluster model

Recall from subsection 3.2.1 that the clusters in this model have the following form

$$G^{A_i} = \sum_{j=1}^{K_i} \delta_{W_{ij}, A_{ij}} .$$

Assume:

- $(K_i, (W_{ij})_{j \geq 1}, (A_{ij})_{j \geq 0})_{i \geq 0}$ constitutes an i.i.d. sequence ,
- $(A_{ij})_{j \geq 0}$ are i.i.d. for any fixed i ,
- $(A_{ij})_{j \geq 1}$ is independent of $K_i, (W_{ij})_{j \geq 1}$ for all $i \geq 0$,
- $(W_{ij})_{j \geq 1}$ are conditionally i.i.d. and independent of K_i given A_{i0}.

Thus we do not exclude the possibility of dependence between $K_i, (W_{ij})_{j \geq 1}$ and the ancestral mark A_{i0}. For any $\gamma > 0$, we denote by

$$A,\ X_j,\ K, W_j,\ m_A\ , m_A^{(\gamma)}\ ,$$

generic random variables with the same distribution as A_{ij}, $X_{ij} = f(A_{ij})$, K_i, W_{ij}, $\mathbb{E}[K_i \mid A_{i0}]$ and $\mathbb{E}[K_i^{\gamma} \mid A_{i0}]$ respectively. Using the cluster representation, one can derive the asymptotic properties of $S(t)$. Let us first consider the Gaussian CLT under appropriate second moment assumptions. Denote by $\mathbb{P}(W \in \cdot \mid A)$ the distribution of W_{ij}'s given A_{i0}.

3.5.1 Corollary. *Assume that* $\mathbb{E}[X^2] < \infty$ *and* $\mathbb{E}[K^2] < \infty$. *If*

$$\sqrt{t}\mathbb{E}[m_A \mathbb{P}(W > t \mid A)] \to 0, \qquad t \to \infty, \tag{3.17}$$

then the relation (3.5) *holds.*

Observe that (3.17) is slightly weaker than the existence of the moment $\mathbb{E}[K\sqrt{W}] < \infty$.

Proof. It follows from the compound sum representation of D that $\mathbb{E}D^2 < \infty$ as soon as $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[K^2] < \infty$. By Proposition 3.3.1, it remains to show that $\varepsilon_t = o_P(\sqrt{t})$. In order to do so, we use the Markov inequality

$$\mathbb{P}(\varepsilon_t > \sqrt{t}) \leq \frac{\mathbb{E}[\varepsilon_t]}{\sqrt{t}} = \frac{\mathbb{E}\left[\sum_{0 \leq \Gamma_i \leq t} \sum_{j=1}^{K_i} \mathbb{I}_{t < \Gamma_i + W_{ij}} f(A_{ij})\right]}{\sqrt{t}}.$$

We use Lemma 7.2.12 of [34] with $f(s) = \sum_{j=1}^{K_i} \mathbb{I}_{W_{ij} > t-s} f(A_{ij})$ in order to compute the r.h.s. term as

$$\begin{aligned} \frac{\int_0^t \mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{W_{ij} > t-s} f(A_{ij})\right] \nu ds}{\sqrt{t}} &= \frac{\nu \mathbb{E}[X] \int_0^t \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{W_{ij} > t-s} \mid A_{i0}\right]\right] ds}{\sqrt{t}} \\ &= \frac{\nu \mathbb{E}[X] \int_0^t \mathbb{E}[m_A \mathbb{P}(W > x \mid A)] dx}{\sqrt{t}}. \end{aligned}$$

Notice that the last identity is obtained thanks to the independence of K_i and $(W_{ij})_{j \geq 0}$ conditionally on A_{i0}. We conclude by the L'Hôpital's rule that this converges to 0 under (3.17). □

For regularly varying D of order $1 < \alpha < 2$, we obtain the corresponding limit theorem under weaker assumptions on the tail of the waiting time W.

3.5.2 Corollary. *Assume that one of the conditions* **(RV)** *holds for* $1 < \alpha < 2$, *so that* D *is regularly varying. When*

$$t^{1+\delta-1/\alpha}\mathbb{E}[m_A \mathbb{P}(W > t \mid A)] \to 0, \qquad t \to \infty, \tag{3.18}$$

for some $\delta > 0$ *the relation* (3.7) *holds.*

The condition (3.18) is slightly weaker than assuming $\mathbb{E}[m_A W^{1+\delta-1/\alpha}] < \infty$. Notice that when $\alpha \to 1^+$ and K is independent of W, this condition boils down to the existence of a δ'th moment of W for any strictly positive δ.

Proof. By definition, (a_t) satisfies $t\mathbb{P}(D > a_t) \to 1$ as $t \to \infty$ and (a_t) is regularly varying with index $1/\alpha$. Applying the Markov inequality as in the proof of Corollary 3.5.1, we obtain

$$\mathbb{P}(\varepsilon_t > a_t) \leq \frac{\mathbb{E}[\varepsilon_t]}{a_t} = \frac{\nu \mathbb{E}[X] \int_0^t \mathbb{E}[m_A \mathbb{P}(W > s \mid A)]ds}{a_t}.$$

The claim follows now by the L'Hôpital's rule and the relation $t^{1/\alpha-\delta} = o(a_t)$ for any $\delta > 0$. □

3.5.1 Remark. *In the context of the mixed binomial model, consider the total claim amount of the stationary process denoted by $S^*(t)$ which takes into account also the arrivals in the interval $(-\infty, 0)$, see Remark 3.4.1. Assume for simplicity that K_i's and (W_{ij})'s are unconditionally independent. Then $\tilde{\varepsilon}_t$ from (3.16) is $o_P(a_t)$ under the same conditions as in Corollaries 3.5.1 and 3.5.2, where we set $a_t = \sqrt{t}$ in the former case. Indeed, we will show that*

$$\mathbb{E}\tilde{\varepsilon}_t = \mathbb{E}\left(\sum_{\Gamma_i \leq -t,\, 0<\Gamma_i+W_{ij}<t} X_{ij}\right) = \mathbb{E}\left(\sum_{-t\leq\Gamma_i\leq 0, t<\Gamma_i+W_{ij}} X_{ij}\right),$$

so that $\mathbb{E}\tilde{\varepsilon}_t = o_P(a_t)$ as well since the r.h.s. is dominated by $\mathbb{E}\varepsilon_t^- = o(a_t)$, cf. (3.15).

Note first that under assumption of the last two corollaries, individual claims have finite expectation, i.e. $\mathbb{E}X < \infty$. So it suffices to show that

$$I_1 := \mathbb{E}\tilde{\varepsilon}_t / \mathbb{E}X = \mathbb{E} \sum_{\Gamma_i < -t} \sum_{j=1}^{K_i} \mathbb{I}_{0<\Gamma_i+W_{ij}<t} = \mathbb{E} \sum_{-t<\Gamma_i<0} \sum_{j=1}^{K_i} \mathbb{I}_{\Gamma_i+W_{ij}>t} =: I_2 .$$

From I_1, I_2 we subtract respectively l.h.s. and r.h.s. of the equality

$$\mathbb{E} \sum_{-2t<\Gamma_i<-t} \sum_{j=1}^{K_i} \mathbb{I}_{\Gamma_i+W_{ij}\in(0,t)} \mathbb{I}_{W_{ij}\in(t,2t]} = \mathbb{E} \sum_{-t<\Gamma_i<0} \sum_{j=1}^{K_i} \mathbb{I}_{\Gamma_i+W_{ij}\in(t,2t)} \mathbb{I}_{W_{ij}\in(t,2t]},$$

where the equality follows by the stationarity of the underlying Poisson process, to obtain

$$J_1 = \mathbb{E} \sum_{-\infty<\Gamma_i<-t} \sum_{j=1}^{K_i} \mathbb{I}_{0<\Gamma_i+W_{ij}<t} \mathbb{I}_{W_{ij}>2t} = \mathbb{E}K \int_{-\infty}^{-t} \nu ds \int_{-s\vee 2t}^{t-s} dF_W(u)$$

and

$$J_2 = \mathbb{E} \sum_{-t<\Gamma_i<0} \sum_{j=1}^{K_i} \mathbb{I}_{\Gamma_i+W_{ij}>t} \mathbb{I}_{W_{ij}>2t} = \mathbb{E}K \int_{-t}^{0} \nu ds \int_{2t}^{\infty} dF_W(u)$$

where F_W denotes the distribution function of delays (W_{ij}). Finally, note that

$$J_1 = \mathbb{E}K \int_{2t}^{\infty} dF_W(u) \int_{-u}^{t-u} \nu ds = J_2 .$$

Since we assumed that $\mathbb{E}[K] < \infty$, the regular variation property of D with index $\alpha \in (0,1)$ can arise only through the claim size distribution, see Proposition 4.8 in [17]. It turns out that in such a heavy tailed case, no additional assumption on the waiting time W is needed.

3.5.3 Corollary. *Assume that X is regularly varying of order $0 < \alpha < 1$, then the relation (3.14) holds.*

Proof. Observe that one cannot apply Markov inequality any more because $\mathbb{E}D = \infty$. Instead, we use the fact that $\sum_{j=1}^{t} X_j / a_t$ converges because X and D have equivalent regular varying tails. Recall from (3.15) that

$$\varepsilon_t = \sum_{0 \le \Gamma_i \le t, t < \Gamma_i + W_{ij}} X_{ij}.$$

We denote the (increasing) number of summands in the r.h.s. term by $J_t = \#\{i, j : 0 \le \Gamma_i \le t, t < \Gamma_i + W_{ij}\}$. We can apply Proposition 3.4.2 after observing that $\sum_{j=1}^{J_t} X_j / a_{J_t}$ is a tight family of random variables, because J_t is independent of the array (X_{ij}). Writing

$$\frac{\varepsilon_t}{a_t} \overset{d}{=} \frac{\sum_{j=1}^{J_t} X_j}{a_{J_t}} \frac{a_{J_t}}{a_t}, \tag{3.19}$$

and observing that a_t is regularly varying with index $1/\alpha$, we obtain the desired result provided that $J_t = o_P(t)$. It is sufficient to show the convergence to 0 of the ratio

$$\begin{aligned} \frac{\mathbb{E}[J_t]}{t} &= \frac{\mathbb{E}[\#\{i, j : 0 \le \Gamma_i \le t, t < \Gamma_i + W_{ij}\}]}{t} \\ &= \frac{\mathbb{E}\left[\sum_{0 \le \Gamma_i \le t} \sum_{j=1}^{K_i} \mathbb{I}_{t \le \Gamma_i + W_{ij}}\right]}{t}. \end{aligned}$$

Using similar calculation as in the proof of Corollary 3.5.1 (setting $X = 1$), we obtain an explicit formula for the r.h.s. term as

$$\frac{\nu \int_0^t \mathbb{E}[m_A \mathbb{P}(W > x \mid A)] dx}{t} \to 0, \qquad t \to \infty,$$

the convergence to 0 following from a Cesarò argument. □

3.5.2 Renewal cluster model

Recall from subsection 3.2.1 that the clusters of this model have the following form

$$G^{A_i} = \sum_{j=1}^{K_i} \delta_{T_{ij}, A_{ij}} ,$$

where

- $T_{ij} = W_{i1} + \cdots + W_{ij}$,
- while $(K_i, (W_{ij})_{j\geq 1}, (A_{ij})_{j\geq 0})_{i\geq 0}$ constitutes an i.i.d. sequence satisfying the assumptions listed in the previous subsection 3.5.1.

The total claim amount coming from the ith immigrant and its progeny is again

$$D \stackrel{d}{=} \sum_{j=0}^{K} X_j,$$

for (X_j) i.i.d. copies of $f(A)$. Dealing with the waiting times $T_{ij} = W_{i1} + \cdots + W_{ij}$ is more involved than in the previous model. We obtain first

3.5.4 Corollary. *Suppose* $\mathbb{E}X^2 < \infty$, $\mathbb{E}K^2 < \infty$ *and* $\mathbb{E}[K^2 W^\delta] < \infty$ *for some* $\delta > 1/2$ *then the relation* (3.5) *holds.*

Proof. The proof follows from Proposition 3.3.1. Second moment of D_i's is finite by the moment assumptions on X and K. It remains to show that the residue term satisfies $\varepsilon_t = o_P(\sqrt{t})$. Using Lemma 7.2.12 of [34] with $f(x) = \sum_{j=1}^{K_i} \mathbb{I}_{W_{i1}+\cdots+W_{ij}>x} f(A_{ij})$ similarly as in the proof of Corollary 3.5.1 we obtain

$$\begin{aligned}
\mathbb{E}[\varepsilon_t] &= \int_0^t \mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{W_{i1}+\cdots+W_{ij}>x} f(A_{ij})\right] \nu dx \\
&= \nu \int_0^t \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{W_{i1}+\cdots+W_{ij}>x} f(A_{ij}) \mid K_i, (W_{ij})_{j\geq 1}\right]\right] dx \\
&= \nu \mathbb{E}[X] \int_0^t \mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{W_{i1}+\cdots+W_{ij}>x}\right] dx
\end{aligned}$$

by independence between $K_i, (W_{ij})_{j\geq 1}$ and $(f(A_{ij}))_{j\geq 1}$. The key argument in dealing with the renewal cluster model is the following upper bound

$$\sum_{j=1}^{K_i} \mathbb{I}_{W_{i1}+\cdots+W_{ij}>x} \leq \mathbb{I}_{W_{i1}+\cdots+W_{iK_i}>x} K_i \, . \tag{3.20}$$

Assume with no loss of generality that $\delta \le 1$. By the Markov inequality and the conditional independence of K_i and $(W_{ij})_{j\ge 0}$ conditionally on A_{i0}, we obtain

$$\mathbb{E}\left[\mathbb{I}_{W_{i1}+\cdots+W_{iK_i}>x}K_i \mid A_{i0}\right] \le \frac{\mathbb{E}[K_i(W_{i1}+\cdots+W_{iK_i})^\delta \mid A_{i0}]}{x^\delta} \le m_{A_{i0}}^{(2)}\frac{\mathbb{E}[W^\delta \mid A_{i0}]}{x^\delta}, \tag{3.21}$$

using the notation $m_{A_{i0}}^{(\gamma)} = \mathbb{E}[K_i^\gamma \mid A_{i0}]$ for any $\gamma > 0$. The last inequality follows from the sub-linearity of the mapping $x \mapsto x^\delta$ for $\delta \le 1$. Thus, we obtain for some constant $C > 0$

$$\mathbb{E}[\varepsilon_t] \le \nu\mathbb{E}[X]\mathbb{E}[m_A^{(2)}W^\delta]\int_1^t x^{-\delta}dx + C = O(\mathbb{E}[m_A^{(2)}W^\delta]t^{1-\delta}) = o(\sqrt{t})$$

as $\delta > 1/2$ by assumption. □

Regularly varying claims can be handled with additional care as K may not be square integrable.

3.5.5 Corollary. *Assume that one of the conditions* **(RV)** *holds so that D is regularly varying of order $1 < \alpha < 2$. Suppose further that $\mathbb{E}[K^{1+\gamma}] < \infty$ and $\mathbb{E}[K^{1+\gamma}W^\delta] < \infty$, $\delta > 0$, $\gamma > 0$ and $\delta > (\alpha-\gamma)/\alpha$. Then the relation (3.7) holds.*

Observe that we obtain somewhat stronger conditions than in the mixed binomial case, see Corollary 3.5.2 and remark following it.

Proof. With no loss of generality we assume that $\gamma \le 1$. We use the Markov inequality of order γ

$$\mathbb{P}(\varepsilon_t > a_t) \le \frac{\mathbb{E}[\varepsilon_t^\gamma]}{a_t^\gamma}.$$

Thanks to the the sub-additivity of the function $x \mapsto x^\gamma$ we have

$$\mathbb{E}[\varepsilon_t^\gamma] = \mathbb{E}\left[\left(\sum_{0\le\Gamma_i\le t}\sum_{j=1}^{K_i}\mathbb{I}_{\Gamma_i+W_{i1}+\cdots+W_{ij}>t}f(A_{ij})\right)^\gamma\right] \le \mathbb{E}\left[\sum_{0\le\Gamma_i\le t}\left(\sum_{j=1}^{K_i}\mathbb{I}_{\Gamma_i+W_{i1}+\cdots+W_{ij}>t}f(A_{ij})\right)^\gamma\right]$$

Using Lemma 7.2.12 of [34] with $f(x) = \sum_{j=1}^{K_i}\mathbb{I}_{W_{i1}+\cdots+W_{ij}>x}f(A_{ij})$ similarly as in the proof of Corollary 3.5.1 we obtain

$$\mathbb{E}[\varepsilon_t^\gamma] \le \nu\int_0^t \mathbb{E}\left[\left(\sum_{j=1}^{K_i}\mathbb{I}_{W_{i1}+\cdots+W_{ij}>x}f(A_{ij})\right)^\gamma\right]dx. \tag{3.22}$$

We use Jensen's inequality as follows

$$\mathbb{E}\left[\mathbb{E}\left[\left(\sum_{j=1}^{K_i}\mathbb{I}_{W_{i1}+\cdots+W_{ij}>x}f(A_{ij})\right)^{\gamma}\mid K_i,(W_{ij})_{j\geq 1}\right]\right]$$
$$\leq \mathbb{E}\left[\left(\mathbb{E}\left[\sum_{j=1}^{K_i}\mathbb{I}_{W_{i1}+\cdots+W_{ij}>x}f(A_{ij})\mid K_i,(W_{ij})_{j\geq 1}\right]\right)^{\gamma}\right]$$

so that, using the independence between K_i, $(W_{ij})_{j\geq 1}$ and $(f(A_{ij}))_{j\geq 1}$, one gets

$$\mathbb{E}[\varepsilon_t^{\gamma}] \leq \nu \mathbb{E}[X]^{\gamma}\int_0^t \mathbb{E}\left[\left(\sum_{j=1}^{K_i}\mathbb{I}_{W_{i1}+\cdots+W_{ij}>x}\right)^{\gamma}\right]dx.$$

Using the stochastic domination (3.20), we obtain

$$\mathbb{E}[\varepsilon_t^{\gamma}] \leq \nu \mathbb{E}[X]^{\gamma}\int_0^t \mathbb{E}\left[\mathbb{I}_{W_{i1}+\cdots+W_{iK_i}>x}K_i^{\gamma}\right]dx.$$

With no loss of generality we assume $0<\delta<1$. Applying the Markov inequality of order δ conditionally on A_{i0} as in (3.21), we have

$$\mathbb{E}[\mathbb{I}_{W_{i1}+\cdots+W_{iK_i}>x}K_i^{\gamma}\mid A_{i0}] \leq m_{A_{i0}}^{(1+\gamma)}\frac{\mathbb{E}[W_i^{\delta}\mid A_{i0}]}{x^{\delta}}.$$

Plugging in this bound in the previous inequality, we obtain for some $C>0$,

$$\mathbb{E}[\varepsilon_t^{\gamma}] \leq \nu \mathbb{E}[X]^{\gamma}\mathbb{E}\left[m_A^{(1+\gamma)}W^{\delta}\right]t^{1-\delta}+C=o(a_t^{\gamma})$$

as $1-\delta<\gamma/\alpha$ by assumption. □

3.5.6 Corollary. *If X is regularly varying of order $\alpha\in(0,1)$ then the relation (3.14) holds.*

Proof. We use the same arguments as in the proof of Corollary 3.5.3 in order to obtain (3.19). The desired result follows if one can show that $J_t=\#\{i,j:0\leq\Gamma_i\leq t, t<\Gamma_i+W_{i1}+\cdots+W_{ij}\}=o_P(t)$. Using the Markov's inequality, it is enough to check that $\mathbb{E}[J_t]/t=o(1)$. Following the same reasoning than in the proof of Corollary 3.5.5, we estimate the moment of J_t similarly as the one of ε_t in (3.22):

$$\mathbb{E}[J_t] \leq \int_0^t \mathbb{E}\left[\sum_{j=1}^{K_i}\mathbb{I}_{W_{i1}+\cdots+W_{ij}>x}\right]\nu dx \leq \int_0^t \mathbb{E}\left[K_i\mathbb{I}_{W_{i1}+\cdots+W_{iK_i}>x}\right]\nu dx.$$

We used again the stochastic domination (3.20) to obtain the last upper bound. From a Cesaró argument, the result will follow if

$$\mathbb{E}\left[K_i\mathbb{I}_{W_{i1}+\cdots+W_{iK_i}>x}\right]\to 0, \qquad x\to\infty.$$

One can actually check this negligibility property because the random sequence $K_i\mathbb{I}_{W_{i1}+\cdots+W_{iK_i}>x}\to 0$ a.s. by finiteness of $W_{i1}+\cdots+W_{iK_i}$ and because the sequence is dominated by K_i that is integrable. □

3.5.3 Marked Hawkes process

Recall from Section 2.2.3 that the clusters of the Hawkes model satisfy recursive relation (2.16). In other words, the clusters G^{A_i} represent a recursive aggregation of Poisson processes with random mean measure $\tilde{\mu}_A \times Q$ which satisfies $\kappa = \mathbb{E}\int h(s,A)ds < 1$.

In general, it is not entirely straightforward to see when the moments of D are finite. However, note that D_i's are i.i.d. and satisfy distributional equation

$$D \stackrel{d}{=} f(A) + \sum_{j=1}^{L_A} D_j\,, \tag{3.23}$$

where L_A has the Poisson distribution conditionally on A, with mean κ_A where $\kappa_A = \int_0^\infty h(s,A)ds$. Recall from (2.17) that $\kappa = \mathbb{E}\kappa_A < 1$. The D_j's on the right hand side are independent of κ_A and i.i.d. with the same distribution as D. Conditionally on A, the waiting times are i.i.d. with common density $h(t,A)/\kappa_A$, $t \geq 0$. Thus, one can relate the clusters of the Hawkes process with those of a mixed binomial process from Section 3.5.1 with $K = L_A$. In order to obtain the asymptotic properties of $S(t)$ one still needs to characterize the moment and tail properties of D.

Consider the Laplace transform of D, i.e. $\varphi(s) = \mathbb{E}e^{-sD}$, for $s \geq 0$. Also, recall the Laplace transform of a Poisson compound sum is of the form

$$\mathbb{E}\left[e^{-s\sum_{j=1}^{M} Z_j}\right] = \mathbb{E}\left[e^{m_A\left(\mathbb{E}e^{-sZ}-1\right)}\right],$$

where M is Poiss(m_A) distributed, independent of the i.i.d. sequence (Z_i) of non-negative random variables with common distribution (see, for instance, Section 7.2.2 in [34]). Note, φ is an infinitely differentiable function for $s > 0$. To simplify the notation, denote by

$$X = f(A)\,,$$

a generic claim size and observe that by (3.23), φ satisfies the following

$$\begin{aligned}\varphi(s) &= \mathbb{E}\left[\mathbb{E}\left(e^{-s(X+\sum_{j=1}^{L_A} D_j)}\middle| A\right)\right] = \mathbb{E}\left[e^{-sX}\mathbb{E}\left(e^{-s\sum_{j=1}^{L_A} D_j}\middle| A\right)\right] \\ &= \mathbb{E}\left[e^{-sX}e^{\kappa_A(\mathbb{E}e^{-sD}-1)}\right] = \mathbb{E}\left[e^{-sX}e^{\kappa_A(\varphi(s)-1)}\right].\end{aligned} \tag{3.24}$$

When $\mathbb{E}[\kappa_A] = \kappa < 1$, it is known that this functional equation has a unique solution φ which further uniquely determines the distribution of D. By studying the behaviour of the derivatives of $\varphi(s)$ for $s \to 0+$, we get the following result.

3.5.1 Lemma. *If $\mathbb{E}X^2 < \infty$ and $\mathbb{E}\kappa_A^2 < \infty$ then*

$$\mathbb{E}D^2 = \frac{\mathbb{E}X^2}{1-\kappa} + \frac{(\mathbb{E}X)^2}{(1-\kappa)^3}\mathbb{E}\kappa_A^2 + 2\frac{\mathbb{E}X}{(1-\kappa)^2}\mathbb{E}(X\kappa_A) < \infty\,.$$

Notice that this expression coincides with the expression in [28], when $X = f(A) \equiv 1$, i.e. in the case when one simply counts the number of claims.

Proof. Differentiating the equation (3.24) with respect to $s > 0$ produces

$$\varphi'(s) = \mathbb{E}\left[e^{-sX} e^{\kappa_A(\varphi(s)-1)} \left(-X + \kappa_A \varphi'(s)\right)\right].$$

As $\mathbb{E}(\kappa_A) = \kappa < 1$ we obtain

$$\varphi'(s) = \frac{-\mathbb{E}\left[e^{-sX} e^{\mu(\varphi(s)-1)} X\right]}{1 - \mathbb{E}\left[\kappa_A^{-sX} e^{\kappa_A(\varphi(s)-1)} \kappa_A\right]} \tag{3.25}$$

As $\varphi(s) \le 1$, $s \ge 0$, the integrand in the numerator is dominated by X and the one in the denominator by κ_A. By the dominated convergence argument, $\lim_{s\to 0+} \varphi'(s)$ exists and is equal to

$$\varphi'(0) = \frac{-\mathbb{E}X}{1-\kappa}.$$

In particular $\mathbb{E}D = \mathbb{E}X/(1-\kappa)$. Differentiating (3.24) again produces second moment of D. Indeed, we have

$$\varphi''(s) = \mathbb{E}\left[e^{-sX} e^{\kappa_A(\varphi(s)-1)} \left(\left(-X + \kappa_A \varphi'(s)\right)^2 + \kappa_A \varphi''(s)\right)\right],$$

so that

$$\varphi''(s) = \frac{\mathbb{E}\left[e^{-sX} e^{\kappa_A(\varphi(s)-1)} \left(-X + \kappa_A \varphi'(s)\right)^2\right]}{1 - \mathbb{E}[e^{-sX} e^{\kappa_A(\varphi(s)-1)} \kappa_A]}. \tag{3.26}$$

Here again, applying the dominated convergence theorem twice, one can let $s \to 0+$ and obtain

$$\varphi''(0) = \frac{\mathbb{E}\left(-X + \kappa_A \varphi'(0)\right)^2}{1-\kappa} = \frac{\mathbb{E}\left(X + \kappa_A \mathbb{E}D\right)^2}{1-\kappa}.$$

Which concludes the proof since $X = f(A)$. □

The following theorem describes the behaviour of the total claim amount $(S(t))$ for the marked Hawkes process under appropriate 2nd moment assumptions. Recall from (2.15) that $\tilde{\mu}_A(B) = \int_B h(s, A) ds$.

3.5.1 Theorem. *If $\kappa < 1$, $\mathbb{E}X^2 < \infty$ and $\mathbb{E}[\kappa_A^2] < \infty$ then, in either stationary or nonstationary case, if*

$$\sqrt{t}\mathbb{E}[\tilde{\mu}_A(t, \infty)] \to 0, \qquad t \to \infty, \tag{3.27}$$

then the relation (3.5) holds.

Proof. In order to apply Proposition 3.3.1 one has to check that $\varepsilon_t = o_P(\sqrt{t})$. The proof is based on the following domination argument on ε_t. Recall that one can write

$$N = \sum_i \sum_j \delta_{\Gamma_i + T_{ij}, A_{ij}} = \sum_{k=1}^{\infty} \delta_{\tau_k, A^k}\,,$$

w.l.o.g. assuming that $0 \le \tau_1 \le \tau_2 \le \ldots$. At each time τ_j, a claim arrives generated by one of the previous claims or an entirely new (immigrant) claim appears. In the former case, if τ_j is a direct offspring of a claim at time τ_i, we will write $\tau_i \to \tau_j$. Progeny τ_j then creates potentially further claims. We denote by D_{τ_j} the total amount of claims generated by the arrival at τ_j (counting the claim at τ_j itself as well). Clearly, D_{τ_j}'s are identically distributed as D and even independent if we consider claims which are not offspring of one another. They are also independent of everything happening in the past.

The process N is naturally dominated by the stationary marked Hawkes process N^* which is well defined on the whole real line as we assumed $\kappa = \mathbb{E}\kappa_A < 1$, see discussion at the end of subsection 3.2.1. For the original and stationary Hawkes processes, N and N^*, by λ and λ^*, we denote corresponding predictable intensities. By the construction of these two point processes, $\lambda \le \lambda^*$. Recall that $\tau_i \to \tau_j$ is equivalent to $\tau_j = \tau_i + W_{ik}$, $k \le L^i = L_{A^i}$, where, by assumption, W_{ik} are i.i.d. with common density $h(t, A^i)/\kappa_{A^i}$, $t \ge 0$, and independent of L^i conditionally on the mark A^i of the claim at τ_i. Moreover, conditionally on A^i, the number of direct progeny of the claim at τ_i, denoted by L^i, has Poisson distribution with parameter $\tilde{\mu}_{A^i}$. Therefore, using conditional independence and equal distribution of D's we get

$$\begin{aligned}
\mathbb{E}[\varepsilon_t] &= \mathbb{E}\Big[\sum_{\Gamma_i \le t} \sum_j \mathbb{I}_{\Gamma_i + T_{ij} > t} X_{ij}\Big] \\
&= \mathbb{E}\Big[\sum_{\tau_i \le t} \sum_{\tau_j > t} D_{\tau_j} \mathbb{I}_{\tau_i \to \tau_j}\Big] \\
&= \mathbb{E}\Big[\sum_{\tau_i \le t} \mathbb{E}\Big[\sum_{k=1}^{L^i} D_{\tau_i + W_{ik}} \mathbb{I}_{\tau_i + W_{ik} > t} \mid (\tau_i, A^i)_{i \ge 0}; \tau_i \le t\Big]\Big] \\
&= \mu_D \mathbb{E}\left[\int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty)) N(ds, da)\right]\,,
\end{aligned}$$

where $\tilde{\mu}_a((u,\infty)) = \int_u^\infty h(s,a)ds$. Observe that from projection theorem, see [8], Chapter 8, Theorem 3, the last expression equals to

$$\mu_D \mathbb{E}\left[\int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty)) Q(da) \lambda(s) ds\right]\,.$$

One can further bound this estimate by

$$\mathbb{E}\left[\int_0^t \int_{\mathbb{M}} \bar{\mu}_a((t-s,\infty))Q(da)\lambda^*(s)ds\right] = \int_0^t \int_{\mathbb{M}} \bar{\mu}_a((t-s,\infty))Q(da)\mathbb{E}[\lambda^*(s)]ds$$
$$= \frac{\nu}{1-\kappa}\int_0^t \int_{\mathbb{M}} \bar{\mu}_a((t-s,\infty))Q(da)ds.$$

Here we used Fubini's theorem, and the expression $\mathbb{E}\left[\lambda^*(s)\right] \equiv \nu/(1-\kappa)$. Observe that this expectation is constant since N^* is a stationary point process, to show that it equals $\nu/(1-\kappa)$, note that

$$\begin{aligned}
\mu^* = \mathbb{E}\lambda^*(s) = \mathbb{E}&\left[\nu + \int_{-\infty}^s \int_{\mathbb{M}} h(s-u,a)N^*(du,da)\right] \\
&= \nu + \int_{-\infty}^s \int_{\mathbb{M}} h(s-u,a)\mathbb{E}(\lambda^*(u))duQ(da) \\
&= \nu + \mu^* \int_{-\infty}^s \mathbb{E}h(s-u,A)du \\
&= \nu + \mu^* \int_0^\infty \mathbb{E}h(v,A)dv\,,
\end{aligned}$$

see also [12], Example 6.3(c). Hence, $\mu^* = \nu + \mu^* \cdot \kappa$ and $\mu^* = \nu/(1-\kappa)$ as we claimed above. Now, we have

$$\mathbb{E}\varepsilon_t \leq \frac{\nu}{1-\kappa}\int_0^t \int_{\mathbb{M}} \bar{\mu}_a((t-s,\infty))Q(da)ds = \frac{\nu}{1-\kappa}\int_0^t \mu_D \int_s^\infty \mathbb{E}[h(u,A)]duds\,. \tag{3.28}$$

Hence the residual term is bounded in expectation by the expression we obtained in the mixed binomial case in Section 3.5.1. Thus, the result will follow from the proof of Corollary 3.5.1 under the condition (3.17) which is further equivalent to (3.27) thanks to the expression of the density of the waiting times.

Dividing the last expression by $\sqrt{t}$ and applying L'Hôpital's rule, proves the theorem for the nonstationary or pure Hawkes process, see [28] where the same idea appears in the proof of Theorem 1.3.2.

To show that the the central limit theorem holds in the stationary case, note that $S(t)$ now has a similar representation as in (3.4) but with an additional term on the right hand side, i.e.

$$S(t) = \sum_{i=1}^{\tau(t)} D_i - D_{\tau(t)} - \varepsilon_t + \varepsilon_{0,t}\,,\ t \geq 0\,, \tag{3.29}$$

where

$$\varepsilon_{0,t} = \sum_{\Gamma_i \leq 0,\ 0<\Gamma_i+T_{ij}<t} X_{ij}\,.$$

Similar computation provides

$$
\begin{aligned}
\mathbb{E}\varepsilon_{0,t} = \mathbb{E}\sum_{\Gamma_i\leq 0}\sum_{j}\mathbb{I}_{0<\Gamma_i+T_{ij}<t}X_{ij} &= \mathbb{E}\sum_{\tau_i\leq 0}\sum_{0<\tau_j<t} D_{\tau_j}\mathbb{I}_{\tau_i\to\tau_j} \\
&= \mathbb{E}\left[\sum_{\tau_i\leq 0}\mu_D\mathbb{E}\left(\sum_{0<\tau_j<t}\mathbb{I}_{\tau_i\to\tau_j}\middle|\mathcal{F}_0\right)\right] \\
&= \mu_D\mathbb{E}\left[\sum_{\tau_i\leq 0}\tilde{\mu}_{A^i}((0-\tau_i,t-\tau_i))\right] \\
&= \mu_D\mathbb{E}\left[\int_{-\infty}^{0}\int_{\mathbb{M}}\tilde{\mu}_a((-s,t-s))N^*(ds,da)\right].
\end{aligned}
$$

where we denote $\tilde{\mu}_a(B) = \int_B h(s,a)ds$ and $\mathcal{F}_0$ stands for the internal history of the process up to time 0, i.e. $\mathcal{F}_0 = \sigma\{N(I\times S) : I\in\mathcal{B}(\mathbb{R}), I\subset(-\infty,0], S\in\mathcal{B}(\mathbb{M})\}$. Again, by the projection theorem, see [8], Chapter 8, Theorem 3, the last expression equals to

$$
\mu_D\mathbb{E}\left[\int_{-\infty}^{0}\int_{\mathbb{M}}\tilde{\mu}_a((-s,t-s))\lambda^*(s)dsQ(da)\right].
$$

Which is further equal to

$$
\begin{aligned}
&\mu_D\int_{-\infty}^{0}\int_{\mathbb{M}}\tilde{\mu}_a((-s,t-s))\mathbb{E}\left[\lambda^*(s)\right]dsQ(da) \\
&= \mu_D\frac{\nu}{1-\kappa}\int_{-\infty}^{0}\int_{\mathbb{M}}\tilde{\mu}_a((-s,t-s))dsQ(da) \\
&= \mu_D\frac{\nu}{1-\kappa}\int_{-\infty}^{0}\mathbb{E}\tilde{\mu}_A((-s,t-s))ds \\
&= \mu_D\frac{\nu}{1-\kappa}\int_{0}^{\infty}\mathbb{E}\tilde{\mu}_A((s,s+t))ds \\
&= \mu_D\frac{\nu}{1-\kappa}\int_{0}^{\infty}\mathbb{E}\int_{s}^{s+t}h(u,A)duds \\
&= \mu_D\frac{\nu}{1-\kappa}\int_{0}^{\infty}\mathbb{E}(t\wedge u)h(u,A)du \\
&= \mu_D\frac{\nu}{1-\kappa}\left(\int_{0}^{t}\mathbb{E}[uh(u,A)]du + t\int_{t}^{\infty}\mathbb{E}[h(u,A)]du\right).
\end{aligned}
$$

Notice that the second term in the last expression divided by $\sqrt{t}$ tends to 0 by (3.27). Using integration by parts for the first term, we have

$$
\int_{0}^{t}\mathbb{E}[uh(u,A)]du = t\int_{t}^{\infty}\mathbb{E}[h(s,A)]ds + \int_{0}^{t}\int_{u}^{\infty}\mathbb{E}[h(s,A)]dsdu.
$$

The first integral on the r.h.s. divided by $\sqrt{t}$ tends to 0 under (3.27). The last term divided by $\sqrt{t}$ also tends to 0 by an application of the L'Hôpital rule as in the non-stationary case.

Finally, we observe that $\varepsilon_{0,t}/\sqrt{t} \xrightarrow{P} 0$ and the result in the stationary case is proved. □

Observe that (3.27) is substantially weaker than (3.3) in the unmarked case. Namely the former condition only requires that the total residue due to the claims on the compact interval $[0,t]$ is of the order $o(\sqrt{t})$ in probability. In particular, in the unmarked case, the central limit theorem holds for the stationary and the non-stationary case even if (3.3) is not satisfied, i.e. even when non-stationary process is not convergent.

As we mentioned above, there are related limit theorems in the literature concerning only the counting process N_t, see [28], but in the contrast to their result, our proof does not rely on the martingale central limit theorem, it stems from rather simple relations (3.4) and (3.29).

In the following example, we consider some special cases of Hawkes processes for which a closed form expression for the 2nd moment $\mathbb{E}D^2$ can be found.

3.5.1 Example. *(Marked Hawkes processes with claims independent of the cluster size) Assume that the random measure* (2.15)

$$\tilde{\mu}_A(B) = \int_B h(s,A)ds\,,$$

on $\mathbb{R}_+$ and the corresponding claim size $X = f(A)$ are independent. In particular, this holds if $\tilde{\mu}_A(B) = \int_B h(s)ds$, for some integrable function h, i.e. when $\tilde{\mu}_A$ is a deterministic measure and we actually have standard Hawkes process with independent marks. In this special case $K+1$ is known to have the so–called Borel distribution, see [21].

Using the arguments from the proof of Lemma 3.5.1, one obtains $\mu_D = \mathbb{E}D_i = \mathbb{E}X/(1-\kappa)$. Similarly the variance of D_i's is finite as the variance of a compound sum, and equals

$$\sigma_D^2 = \frac{\sigma_X^2}{1-\kappa} + \frac{\kappa(\mathbb{E}X)^2}{(1-\kappa)^3}\,,$$

cf. Lemma 2.3.4 in [34]. Hence $\mathbb{E}D^2$ in Theorem 3.5.1 has the form

$$\mathbb{E}D^2 = \sigma_D^2 + \mu_D^2 = \frac{\sigma_X^2}{1-\kappa} + \frac{(\mathbb{E}X)^2}{(1-\kappa)^3}.$$

In the special case, when the claims are all constant, say $X = f(A) \equiv c > 0$, direct calculation yields $\mathbb{E}D = c/(1-\kappa)$, with $\kappa = \mathbb{E}[\kappa_A]$, and

$$\varphi''(0) = \frac{\mathbb{E}\left(-c + \kappa_A \varphi'(0)\right)^2}{1-\kappa} = c^2 \frac{\operatorname{Var}\kappa_A + 1}{(1-\kappa)^3}\,,$$

obtaining

$$\mathbb{E}D^2 = \varphi''(0) = c^2 \frac{\operatorname{Var}\kappa_A + 1}{(1-\kappa)^3},$$

in particular, for $c = 1$ we recover expression in [28].

In the rest of this subsection, we study marked Hawkes process in the case when D_i's are regularly varying with index $\alpha < 2$. Using the result of [24], one can show that when the individual claims $X = f(A)$ are regularly varying, this property is frequently passed on to the random variable D under appropriate moment assumptions on κ_A. However, using the specific form of the Laplace transform for D given in (3.24), one can show regular variation of D under weaker conditions. This is the content of the following lemma.

3.5.2 Lemma. *Assume that $\kappa < 1$ and that $X = f(A)$ is regularly varying with index $\alpha \in (0,1) \cup (1,2)$. When $\alpha \in (1,2)$, assume additionally that $Y = X + \kappa_A \mu_D$ is regularly varying of order α. Then the random variable D is regularly varying with the same index α.*

Proof. We will use Karamata's Tauberian Theorem, as formulated and proved in Theorem 8.1.6 of [7]. In particular, the equivalence between (8.1.12) and (8.1.11b) in [7] yields the following.

3.5.2 Theorem. *The nonnegative random variable X is regularly varying with a noninteger tail index $\alpha > 0$, i.e. $\bar{F}(x) \sim x^{-\alpha}\ell(x)$ as $x \to \infty$ if and only if*

$$\varphi^{(\lceil\alpha\rceil)}(s) \sim c s^{\alpha-\lceil\alpha\rceil}\ell(1/s), \qquad s \to 0+,$$

for some slowly varying function ℓ and a constant depending only on α: $c = -\Gamma(\alpha+1)\Gamma(1-\alpha)/\Gamma(\alpha-\lfloor\alpha\rfloor)$.

Consider first the case $0 < \alpha < 1$. By differentiating once the expression for the Laplace transform, we obtain the identity (3.25)

$$\varphi'(s) = \frac{-\mathbb{E}\left[e^{-sX}e^{\kappa_A(\varphi(s)-1)}X\right]}{1-\mathbb{E}[e^{-sX}e^{\kappa_A(\varphi(s)-1)}\kappa_A]}, \qquad s > 0.$$

We are interested in the behaviour of this derivative as $s \to 0+$. Using the inequality $|1-e^{-x}| = 1-e^{-x} \le x$, we have

$$\left|\varphi'(s) - \frac{-\mathbb{E}\left[e^{-sX}X\right]}{1-\mathbb{E}[e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A]}\right| \le \frac{\mathbb{E}\left[\kappa_A(1-\varphi(s))e^{-sX}X\right]}{1-\mathbb{E}[e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A]}$$
$$\le \frac{1-\varphi(s)}{s}\frac{\mathbb{E}\left[\kappa_A e^{-sX}sX\right]}{1-\mathbb{E}[e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A]}.$$

As $e^{-sX}sX \le e^{-1}$, we prove that $\mathbb{E}\left[\kappa_A e^{-sX}sX\right] = o(1)$ as $s \to 0+$ by dominated convergence. As in the proof of Lemma 3.5.1 the denominator $1 - \mathbb{E}[e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A]$ is controlled thanks to dominated convergence as well. Moreover, using again $1 - e^{-x} \le x$ and denoting $\varphi_X(s) = \mathbb{E}[e^{-sX}]$ the Laplace transform of X, we have

$$0 \le \varphi_X(s) - \varphi(s) \le \mathbb{E}[e^{-sX}\kappa_A(1-\varphi(s))] \le \kappa(1-\varphi(s))$$

so that

$$1 - \varphi(s) \le \frac{1}{1-\kappa}(1-\varphi_X(s)).$$

Collecting all those bounds and using the identity $\varphi_X'(s) = -\mathbb{E}[e^{-sX}X]$, we obtain

$$\left|\varphi'(s) - \frac{\varphi_X'(s)}{1 - \mathbb{E}[e^{-sX}e^{\kappa_A(\varphi(s)-1)}\kappa_A]}\right| = o\left(\frac{1-\varphi_X(s)}{s}\right), \qquad s \to 0^+. \tag{3.30}$$

The regular variation of the random variable D follows now from the regular variation of the random variable X by two consecutive applications of Theorem 3.5.2. First, as X is regularly varying of order $0 < \alpha < 1$, applying the direct part of the equivalence in Theorem 3.5.2 we obtain

$$\varphi_X'(s) \sim cs^{\alpha-1}\ell(1/s), \qquad s \to 0^+.$$

Applying Karamata's theorem, i.e. the equivalence between (8.1.9) and (8.1.11b) in [7, Theorem 8.1.6], we obtain $(1-\varphi_X(s))/s = O(\varphi_X'(s))$ as $s \to 0+$. Using (3.30) and the limiting relation

$$\mathbb{E}[e^{-sX}e^{\kappa_A(\varphi(s)-1)}\kappa_A] \to \kappa, \qquad s \to 0^+,$$

we obtain

$$\varphi'(s) \sim \frac{\varphi_X'(s)}{1-\kappa} \sim \frac{cs^{\alpha-1}\ell(1/s)}{1-\kappa}, \qquad s \to 0^+.$$

Finally, applying the reverse part of Theorem 3.5.2, we obtain

$$\bar{F}_D(x) \sim \frac{\ell(x)x^{-\alpha}}{1-\kappa} = \frac{\bar{F}_X(x)}{1-\kappa}, \qquad x \to \infty.$$

The case $1 < \alpha < 2$ can be treated similarly, under the additional assumption that $Y = X + \kappa_A\mu_D$ is regularly varying. We will again show that $P(D > x) \sim (1-\kappa)^{-1}P(Y > x)$ as $x \to \infty$. To prove this equivalence, recall the identity (3.26)

$$\varphi''(s) = \frac{\mathbb{E}\left[e^{-sX}e^{\kappa_A(\varphi(s)-1)}\left(-X+\kappa_A\varphi'(s)\right)^2\right]}{1 - \mathbb{E}[e^{-sX}e^{\kappa_A(\varphi(s)-1)}\kappa_A]}.$$

As $\alpha > 1$, we have that $\mathbb{E}[Y] < \infty$ and thus $\mathbb{E}[X] < \infty$ and $E[D] = \mu_D = (1-\kappa)^{-1}\mathbb{E}[X]$. Observe that, for any $s > 0$,

$$\left|\varphi''(s) - \frac{\mathbb{E}\left[e^{-sY}Y^2\right]}{1-\mathbb{E}[e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A]}\right|$$
$$= \left|\frac{\mathbb{E}\left[e^{-sX-\kappa_A(1-\varphi(s))}\left(-X+\kappa_A\varphi'(s)\right)^2\right] - \mathbb{E}\left[e^{-sY}Y^2\right]}{1-\mathbb{E}[e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A]}\right|.$$

Let us decompose the numerator into two terms

$$\underbrace{\mathbb{E}\left|e^{-sX-\kappa_A(1-\varphi(s))}\left(\left(-X+\kappa_A\varphi'(s)\right)^2 - Y^2\right)\right|}_{I_1}$$
$$+\underbrace{\mathbb{E}\left|\left(e^{-sY}-e^{-sX-\kappa_A(1-\varphi(s))}\right)Y^2\right|}_{I_2}.$$

Using the identity $a^2 - b^2 = (a-b)(a+b)$, I_1 is bounded by

$$I_1 \le (\mu_D + \varphi'(s))\mathbb{E}\left[e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A\left(2X + \kappa_A(\mu_D - \varphi'(s))\right)\right]$$
$$\le \frac{\mu_D+\varphi'(s)}{s}\left(\mathbb{E}\left[2\kappa_A e^{-sX}sX\right] + \mathbb{E}\left[\kappa_A e^{-\kappa_A(1-\varphi(s))}s(\mu_D - \varphi'(s))\right]\right).$$

As $e^{-sX}sX \le e^{-1}$ then $\mathbb{E}\left[2\kappa_A e^{-sX}sX\right] = o(1)$ as $s \to 0^+$ by dominated convergence. By convexity of $\varphi(s)$ we have $1-\varphi(s) \ge -\varphi'(s)s$ for any $s > 0$. Thus

$$e^{-\kappa_A(1-\varphi(s))}(-\varphi'(s)s) \le e^{-\kappa_A(-\varphi'(s)s)}(-\varphi'(s)s) \le e^{-1}$$

and the dominated convergence argument also applies to the second integrand as $-\varphi'(s)s \le 1-\varphi(s) = o(1)$. We obtain $I_1 = o((\mu_D + \varphi'(s))/s)$ as $s \to 0^+$. In order to control the rate of $(\mu_D + \varphi'(s))/s$, we notice that $\varphi(s)$ is μ_D Lipschitz on $s \ge 0$ so that $|1-\varphi(s)| = 1-\varphi(s) \le \mu_D s$. Then

$$sX + \kappa_A(1-\varphi(s)) \le sX + s\kappa_A\mu_D = sY$$

and we bound

$$\varphi'(s) \le \frac{-\mathbb{E}\left[e^{-sY}X\right]}{1-\mathbb{E}[e^{-sY}\kappa_A]}$$
$$\le \frac{\varphi_Y'(s)}{1-\mathbb{E}[e^{-sY}\kappa_A]} + \frac{\kappa\mu_D}{1-\mathbb{E}[e^{-sY}\kappa_A]}$$

where $\varphi_Y(s) = \mathbb{E}[e^{-sY}]$ denotes the Laplace transform of Y. It yields to the estimates $\mu_D + \varphi'(s) = O(\mu_D + \varphi_Y'(s)) + O(\kappa - \mathbb{E}[e^{-sY}\kappa_A])$. Using again that $1 - e^{-x} \le x$ on the second term we obtain that $I_1 = o((\mu_D + \varphi_Y'(s))/s) + o(1)$ as $s \to 0^+$.

We now turn to the term I_2 that we identify as

$$\begin{aligned} I_2 &= \mathbb{E}\left|\left(e^{sX+\kappa_A(1-\varphi(s))-sY}-1\right)e^{-sX-\kappa_A(1-\varphi(s))}Y^2\right| \\ &= \mathbb{E}\left|\left(e^{-\kappa_A(\mu_D s-(1-\varphi(s)))}-1\right)e^{-sX-\kappa_A(1-\varphi(s))}Y^2\right|. \end{aligned}$$

As $1-\varphi(s) \le \mu_D s$ the term in the absolute value is negative for $s>0$ and

$$I_2 = \mathbb{E}\left[\left(1-e^{-\kappa_A(\mu_D s-(1-\varphi(s)))}\right)e^{-sX-\kappa_A(1-\varphi(s))}Y^2\right].$$

Using again the basic inequality $1-e^{-x} \le x$ for $x \ge 0$ we obtain the new estimate

$$\begin{aligned} I_2 &\le \mathbb{E}\left[\kappa_A(s\mu_D-(1-\varphi(s))e^{-sX-\kappa_A(1-\varphi(s))}Y^2\right] \\ &\le \frac{s\mu_D-(1-\varphi(s))}{s^2}\mathbb{E}\left[\kappa_A e^{-sX-\kappa_A(1-\varphi(s))}(sY)^2\right]. \end{aligned}$$

We have

$$(sY)^2 \le (sX+\kappa_A(1-\varphi(s)))^2+\kappa_A^2(s\mu_D-(1-\varphi(s)))^2.$$

As $e^{-x}x^2 \le 4e^{-2}$ for any $x>0$, we prove that

$$\mathbb{E}\left[\kappa_A e^{-(sX+\kappa_A(1-\varphi(s))}(sX+\kappa_A(1-\varphi(s)))^2\right] = o(1)$$

as $s \to 0^+$ by dominated convergence. It remains to bound the term

$$e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A^2(s\mu_D-(1-\varphi(s)))^2$$

uniformly for $s>0$ sufficiently small. As $1-\varphi(s) \sim s\mu_D$ as $s \to 0^+$, we have $0 \le s\mu_D-(1-\varphi(s)) \le 1-\varphi(s)$ for s sufficiently small. Then we obtain

$$e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A^2(s\mu_D-(1-\varphi(s)))^2 \le e^{-\kappa_A(1-\varphi(s))}\kappa_A^2(1-\varphi(s))^2 = o(1)$$

where the negligibility follows from dominated convergence and the basic inequality $e^{-x}x^2 \le 4e^{-2}$ for any $x>0$. We obtain

$$I_2 = o\left(\frac{s\mu_D-(1-\varphi(s))}{s^2}\right), \qquad s \to 0^+.$$

Similar computation than above yields

$$\begin{aligned} 0 \le \varphi(s)-\varphi_Y(s) \le \mathbb{E}\left[\kappa_A(s\mu_D-(1-\varphi(s))e^{-sX+\kappa_A(\varphi(s)-1))}\right] \\ \le \kappa(s\mu_D-(1-\varphi(s)). \end{aligned}$$

Thus as $(s\mu_D - (1-\varphi(s)) \leq (s\mu_D - (1-\varphi_Y(s))/(1-\kappa)$ and $\mathbb{E}[Y] = \mathbb{E}[X] + \kappa\mu_D = \mu_D$ we conclude that

$$\left|\varphi''(s) - \frac{\varphi_Y''(s)}{1-\mathbb{E}[e^{-sX-\kappa_A(1-\varphi(s))}\kappa_A]}\right| = o\left(1 + \frac{\mathbb{E}[Y]+\varphi_Y'(s)}{s} + \frac{s\mathbb{E}[Y]-(1-\varphi_Y(s))}{s^2}\right),$$

as $s \to 0^+$. Let us first apply Theorem 3.5.2 on Y so that $\varphi_Y''(s)$ is $\alpha-2$ regularly varying around 0. Applying Karamata's theorem again, i.e the equivalences between (8.1.11b) and (8.1.9), (8.1.11b) and (8.1.10) in [7, Theorem 8.1.6] assert respectively that $(s\mathbb{E}[Y] - (1-\varphi_Y(s)))/s^2 = O(\varphi_Y''(s))$ and $(\mathbb{E}[Y] + \varphi_Y'(s))/s = O(\varphi_Y''(s))$ as $s \to 0^+$. We then obtain

$$\varphi''(s) \sim \frac{\varphi_Y''(s)}{1-\kappa} \sim \frac{cs^{\alpha-2}\ell(1/s)}{1-\kappa}, \qquad s \to 0+,$$

and finally $\bar{F}(x) \sim \bar{F}_Y(x)/(1-\kappa)$, $x \to \infty$, by applying the reverse part of Theorem 3.5.2. □

We are now ready to characterize the asymptotic behavior of $S(t)$ in the regularly varying case.

3.5.3 Theorem. *Assume that the assumptions of Lemma 3.5.2 hold.*
i) If $\alpha \in (0,1)$ and there exists $\delta > 0$ such that

$$t^{\delta}\mathbb{E}[\tilde{\mu}_A(t,\infty)] \to 0. \tag{3.31}$$

as $t \to \infty$, then there exists a sequence (a_n), $a_n \to \infty$, and an α–stable random variable G_α such that

$$\frac{S(t)}{a_{\lfloor \nu t\rfloor}} \xrightarrow{d} G_\alpha\,.$$

ii) If $\alpha \in (1,2)$ and

$$t^{1+\delta-1/\alpha}\mathbb{E}[\tilde{\mu}_A(t,\infty)] \to 0\,, \tag{3.32}$$

as $t \to \infty$ holds for some $\delta > 0$, then there exists a sequence (a_n), $a_n \to \infty$, and an α–stable random variable G_α such that

$$\frac{S(t) - t\nu\mu_D}{a_{\lfloor \nu t\rfloor}} \xrightarrow{d} G_\alpha\,.$$

Proof. The proof is based on the representation (3.4), and application of Propositions 3.4.1 and 3.4.2. In either case, it remains to show that

$$\varepsilon_t = o_P(a_t)\,.$$

Consider first the case $\alpha \in (1,2)$. Since then $\mu_D = \mathbb{E}D < \infty$, the argument in the proof of Theorem 3.5.1 still yields the bound (3.28) on $\mathbb{E}\varepsilon_t$. Using L'Hôpital's rule again together with condition (3.32), shows that $\mathbb{E}\varepsilon_t = o(t^{1/\alpha-\delta})$, where we assume without loss of generality that $\delta < 1/\alpha$. Since, $a_t = t^{1/\alpha}\ell(t)$ for some slowly varying function ℓ, it follows that $\varepsilon_t/a_t \xrightarrow{P} 0$ as $t \to \infty$.

For $\alpha \in (0,1)$, random variable D has no finite mean. In order to prove that $\varepsilon_t = o_P(a_t)$ we use the Markov inequality of order $0 < \gamma < \alpha$ as $\mathbb{E}[D^\gamma] < \infty$. We will show that under assumption (3.31)

$$\mathbb{E}[\varepsilon_t^\gamma] = o(a_t^\gamma)\,, \qquad t \to \infty.$$

By sub-linearity of $x \to x^\gamma$, $\gamma \leq 1$, we have

$$\begin{aligned}
\mathbb{E}[\varepsilon_t^\gamma] &= \mathbb{E}\Big[\Big(\sum_{\Gamma_i \leq t}\sum_j \mathbb{I}_{\Gamma_i + T_{ij} > t} X_{ij}\Big)^\gamma\Big] \\
&= \mathbb{E}\Big[\Big(\sum_{\tau_i \leq t}\sum_{\tau_j > t} D_{\tau_j}\mathbb{I}_{\tau_i \to \tau_j}\Big)^\gamma\Big] \\
&\leq \mathbb{E}\Big[\sum_{\tau_i \leq t}\sum_{\tau_j > t} D_{\tau_j}^\gamma \mathbb{I}_{\tau_i \to \tau_j}\Big] \\
&= \mathbb{E}\Big[\sum_{\tau_i \leq t}\mathbb{E}\Big[\sum_{k=1}^{L^i} D_{\tau_i + W_{ik}}^\gamma \mathbb{I}_{\tau_i + W_{ik} > t} \mid (\tau_i, A^i)_{i \geq 0}; \tau_i \leq t\Big]\Big] \\
&= \mathbb{E}[D^\gamma]\mathbb{E}\left[\int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty))N(ds,da)\right].
\end{aligned}$$

We can again compare the marked Hawkes process N with a stationary version of it, N^* say. By the same arguments as in the proof of Theorem 3.5.1, we obtain

$$\mathbb{E}\left[\sum_{0 \leq \tau_i \leq t}\sum_{t < \tau_j} \mathbb{I}_{\tau_i \to \tau_j}\right] \leq \frac{\nu}{1-\kappa}\left(\int_0^t \int_u^\infty \mathbb{E}[h(s,A)]dsdu\right).$$

By regular variation of order $1/\alpha$ of (a_t) we have $t^{\gamma/\alpha - \delta'} = o(a_t^\gamma)$ for any $\delta' > 0$. Once again, we use a Cesaró argument to prove that $\mathbb{E}[\varepsilon_t^\gamma] = o(a_t^\gamma)$ under the condition

$$t^{1+\delta'-\gamma/\alpha}\mathbb{E}[\tilde{\mu}_A(t,\infty)] \to 0\,, \qquad t \to \infty\,.$$

As γ can be taken as close as possible to α, the result holds under assumption (3.31). □

3.5.2 Remark. *Theorem 3.5.3 i) and ii) also hold on the stationary version following the same arguments as in the proof of Theorem 3.5.1.*

4

On maximal claim size for marked Poisson cluster models

4.1 Introduction

We want to observe and understand the behaviour of the maximal claim size up until time t, i.e. we are interested in

$$M(t) = \sup \{f(X_k) : \tau_k \leq t\} ,$$

where τ_k represents the arrival time of k'th event and events arrive in clusters. It turns out that with suitable conditions on claim sizes and the leftover effect at time t one can handle maximal claim size, $M(t)$. Namely, if the claim size, X belongs to the maximum domain of attraction of of one of the three extreme value distribution and the number of claims in the leftover effect at time t grows much slower than the normalizing sequence it can be proved that $M(t)$, when properly normalized converges weakly to an extreme value distribution. The case when a random variable X belongs to the maximum domain of attraction (MDA for short) of the Fréchet distribution has been studied the most (so the literature is quite rich, see e.g. [16], [39], [38], [20], [7]) and is often used in practice. On the other hand, the maximum domain of attraction of the Gumbel distribution contains wide range of distributions. It includes sub–exponential distributions (e.g. lognormal distribution), but also distributions with light tails and infinite right-end point (e.g.normal, exponential) and it also contains distributions with finite right-end point.

4.1.1 Remark. *We will discuss only the cases of Fréchet and Gumbel maximum domain of attraction. The case when distributions are in the maximum domain of attraction of Weibull distribution follows easy from the results in the Fréchet case. Nevertheless, we will omit them because distributions in the maximum domain of attraction of Weibull distributions have support bounded to the right (Theorem 3.3.12., p 135 [16]). So, because $x_F < \infty$ they are mostly omitted in modelling extremal events in insurance and finance. As in [16] is stated,* *often distributions with $x_F = \infty$ should be preferred since they allow for arbitrarily large values in a sample. Such distributions typically belong to the maximum domain of attraction of Fréchet or Gumbel distribution. See Section 3.3.2 in [16] for detail.*

In the next section we introduce our general model. In section three we state and prove our result. In the last, fourth section we apply this result on to the three special marked Poisson cluster models where, as our key example, the linear marked Hawkes process plays the most interesting role.

4.2 The general marked Poisson cluster model

The model we used is essentially the same as one introduced in section 3.2.

Recall, to describe the size and other characteristics of the claims together with their arrival times, we use a marked point process N as a random element in $\mathcal{N}^{\#g}_{\mathbb{R}_{\geq 0}\times\mathbb{M}}$ of the form 3.1

$$N = \sum_{i=1}^{\infty}\sum_{j=0}^{K_i} \delta_{\Gamma_i+T_{ij},A_{ij}}\,,$$

where we set $T_{i0} = 0$ and $A_{i0} = A_i$. In this representation, the claims arriving at time Γ_i and corresponding to the index $j = 0$ are called ancestral or immigrant claims, while the claims arriving at times $\Gamma_i + T_{ij}$, $j \geq 1$, are referred to as progeny or offspring. Also, we assume that

$$\mathbb{E}K_i < \infty\,. \tag{4.1}$$

Since N is locally finite, one could also write

$$N = \sum_{k=1}^{\infty} \delta_{\tau_k,A^k}\,,$$

with $\tau_k \leq \tau_{k+1}$ for all $k \geq 1$ and A^k are iid marks (we ignore the information regarding the clusters of the point process). Clearly, if the cluster processes G^{A_i} are independently marked with the same mark distribution Q independent of A_i, then all the marks A^k are i.i.d.

We again assume that the size of claims is produced by an application of a measurable function, say $f : \mathbb{M} \to \mathbb{R}_+$, on the marks. Maximum of all the claims due to the arrival of an immigrant claim at time Γ_i equals

$$H_i = \max\left\{f(A_{ij}) : 0 \leq j \leq K_i\right\}\,, \tag{4.2}$$

while the maximal claim size in the period $[0, t]$ can be calculated as

$$M(t) = \sup\left\{f(A^k) : \tau_k \leq t\right\}.$$

H_i has an interpretation as the maximal claim size coming from the ith immigrant and its progeny. It is useful in the sequel to introduce random variables

$$\tau(t) = \inf\{n : \Gamma_n > t\},\ t \geq 0,$$

Let (W_n) be an iid sequence of Exponential random variables with parameter $\nu > 0$, that is $W_1 \sim Exp(\nu)$, so we have $\mathbb{E}W_1 = \frac{1}{\nu}$.
Then, $\tau(t)$ is a renewal process generated by the sequence (W_n). According to Theorem 5.1. in [19], for every $c \geq 0$,

$$\frac{\tau(tc)}{\nu t} \xrightarrow{as} c, \quad t \to \infty. \tag{4.3}$$

If we denote

$$M^\tau(t) = \sup\{H_i : 1 \leq i \leq \tau(t)\},$$

then we can write

$$M^\tau(t) = M(t) \bigvee H_{\tau(t)} \bigvee \varepsilon_t^m,\ t \geq 0, \tag{4.4}$$

where the last error term represents the leftover effect at time t, i.e. the maximum of all the claims arriving after t which correspond to the progeny of immigrants arriving before time t, that is

$$\varepsilon_t^m = \max\{f(A_{ij}) : 0 \leq \Gamma_i \leq t, t < \Gamma_i + T_{ij}\} \quad t \geq 0. \tag{4.5}$$

Denote the (increasing) number of members in the ε_t^m by

$$J(t) = \#\{(i,j) : 0 \leq \Gamma_i \leq t, t < \Gamma_i + W_{ij}\}.$$

To simplify the notation, for a generic member of an identically distributed sequence or an array, say (H_n), (A_{ij}), we write H, A etc.

4.3 Limiting behaviour of the maximal claim size

4.3.1 Definition. *[20], [16], [7] A random variable X belongs to the maximum domain of attraction of the extreme value distribution G if there exists constants $a_n > 0$, $b_n \in \mathbb{R}$ such that*

$$\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \xrightarrow{d} G, \quad n \to \infty,$$

for an i.i.d. sequence $X, X_1, X_2, \ldots$ of non-degenerate random variables.

4.3.1 Proposition. *Assume that H_i's are in the maximum domain of attraction of Fréchet or Gumbel extreme value distribution G and that*

$$J(t) = o_P(a_t). \tag{4.6}$$

Then

$$\frac{M(t) - b_{\lfloor \nu t \rfloor}}{a_{\lfloor \nu t \rfloor}} \xrightarrow{d} G, \tag{4.7}$$

as $t \to \infty$.

Proof. Similarly as in Propositions 3.3.1, 3.4.1 and 3.4.2 in chapter 3, we relay on (4.4) and on two results in Basrak and Špoljarić [6]. We state them below.

4.3.1 Lemma. *(Lemma 1 in [6]) Assume that $N, (N_t), t \geq 0$ are point processes with values in $\mathcal{N}^{\#}_{\mathbb{R}_{\geq 0} \times \mathbb{E}}$, for a measurable subset $\mathbb{E}$ of $\mathbb{R}^d$. Assume further that $Z, (Z_t), t \geq 0$ are $\mathbb{R}_{\geq 0}$ valued random variables. If $P(N(\{Z\} \times \mathbb{E}) > 0) = 0$ and $(N_t, Z_t) \xrightarrow{d} (N, Z)$, as $t \to \infty$, then*

$$N_t|_{[0,Z_t]\times\mathbb{E}} \xrightarrow{d} N|_{[0,Z]\times\mathbb{E}},$$

as $t \to \infty$.

By $m|_A$ above we denote the restriction of a point measure m on a set A, i.e. $m|_{A(B)} = m(A \cap B)$.

4.3.1 Theorem. *(Theorem 2 in [6]) Suppose (X_n) is an iid sequence such that $X_1 \in MDA(G)$. If the mean step size of the renewal process $(\tau(t))$ is finite ($EW = \frac{1}{\nu} < \infty$), then, for every $c \geq 0$,*

$$N_t\Big|_{\left[0,\frac{\tau(tc)}{\nu t}\right]\times\mathbb{M}} \xrightarrow{d} N\Big|_{[0,c]\times\mathbb{M}}, \quad t \to \infty, \tag{4.8}$$

where N is PRM$(\lambda \times \mu_G)$ with λ represents Lebesgue measure, μ_G and state space $\mathbb{M}$ correspond to G and

$$N_t = \sum_{i=1}^{\infty} \delta_{\left(\frac{i}{\nu t}, \frac{X_i - b_{\lfloor \nu t \rfloor}}{a_{\lfloor \nu t \rfloor}}\right)}$$

Recall from subsection 2.2.1, abbreviation PRM$(\lambda \times \mu_G)$ stands for Poisson random measure with mean measure $\lambda \times \mu_G$ where λ denotes the Lebesgue measure and μ_G represents the measure induced by the nondecreasing function $\log G$ (see for example [39]), indicating that the limit is a Poisson process. If the conditions of the previous theorem are satisfied, then

$$\frac{M^{\tau}(t) - b_{\lfloor \nu t \rfloor}}{a_{\lfloor \nu t \rfloor}} \xrightarrow{d} G, \quad t \to \infty. \tag{4.9}$$

Namely,

$$\mathbb{P}\left(\frac{M^{\tau}(t)-b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}}\leq x\right)=\mathbb{P}\left(N_t|_{\left[0,\frac{\tau(t)}{\nu t}\right]\times\langle x,\infty]}=0\right),$$

it follows

$$\mathbb{P}\left(N_t|_{\left[0,\frac{\tau(t)}{\nu t}\right]\times\langle x,\infty]}=0\right)\to\mathbb{P}\left(N|_{[0,1]\times\langle x,\infty]}=0\right)=G(x),\quad t\to\infty,$$

Equation (4.4) and the previous result yield

$$\frac{M^{\tau}(t)-b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}}=\frac{M(t)-b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}}\bigvee\frac{H_{\tau(t)}-b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}}\bigvee\frac{\varepsilon_t^m-b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}},\ t\geq 0.$$

Next, we have

$$0\leq\frac{H_{\tau(t)}-b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}}\leq\frac{H_{\tau(t)}}{a_{\lfloor \nu t\rfloor}}\leq\frac{\sum_{j=0}^{K_{\tau(t)}}X_{\tau(t)j}}{a_{\lfloor \nu t\rfloor}}\xrightarrow{P}0,\quad t\to\infty.$$

because (Γ_n) and $(\sum_{j=0}^{K_n}X_{nj})$ are independent (see Lemma 2.9 and Lemma 3.5 in [18]). Notice, the centering sequence (b_n) is nonnegative because the size of claims is produced by an application of a measurable function $f:\mathbb{M}\to\mathbb{R}_+$, on the marks and by taking into account that $b_n=F^{\leftarrow}(1-\frac{1}{n})$ [16] . To finish the proof, we will use the assumption (4.6) and the same technique used in Theorem 3.1 [6]. According to (4.9), $M^{\tau}(t)$ behaves as $M_{\lfloor \nu t\rfloor}$, so we define

$$N_t=\sum_{i\geq 1}\delta_{\left(\frac{i}{\lfloor \nu t\rfloor},\frac{X_i-b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}}\right)}.$$

The assumption $X_1\in\mathrm{MDA}(G)$ is equivalent to $N_t\xrightarrow{d}N$, $n\to\infty$ where N is $\mathrm{PRM}(\lambda\times\mu_G)$. Next, by (4.6)

$$\frac{\nu t-J(t)}{\nu t}=1-\frac{1}{\nu}\cdot\frac{J(t)}{t}\xrightarrow{P}1-0=1.$$

Thus, by Proposition 3.1 in [39] we have joint convergence

$$\left(N_t,\frac{\lfloor \nu t\rfloor-J(t)}{\nu t}\right)\xrightarrow{d}(N,1),$$

as $t\to\infty$. An application of Lemma 1 in [6] yields

$$N_t|_{\left[0,\frac{\nu t-J(t)}{\nu t}\right]\times\mathbb{M}}\xrightarrow{d}N|_{[0,1]\times\mathbb{M}},$$

as $t \to \infty$. Since

$$\mathbb{P}\left(\frac{\bigvee_{i=1}^{\lfloor \nu t\rfloor - J(t)} X_i - b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}} \leq x\right) = \mathbb{P}\left(N_t|_{\left[0,\frac{\lfloor \nu t\rfloor - J(t)}{\lfloor \nu t\rfloor}\right]\times\langle x,\infty]} = 0\right),$$

it follows

$$\mathbb{P}\left(N_t|_{\left[0,\frac{\lfloor \nu t\rfloor - J(t)}{\lfloor \nu t\rfloor}\right]\times\langle x,\infty]} = 0\right) \to \mathbb{P}\left(N|_{[0,1]\times\langle x,\infty]} = 0\right) = G(x), \quad t\to\infty.$$

Now we have,

$$\frac{M^\tau(t) - b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}} \sim \frac{\bigvee_{i=1}^{\lfloor \nu t\rfloor - J(t)} - b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}} \bigvee \frac{\bigvee_{i=1}^{J(t)} - b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}},$$

as $t \to \infty$. From two previous results, and by taking into account that $\vee_{i=1}^{J(t)} X_i$ is tight (it is stochastically dominated by $M^\tau(t)$) it follows that

$$\frac{\varepsilon_t^m - b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}} \stackrel{d}{=} \frac{\bigvee_{i=1}^{J(t)} - b_{\lfloor \nu t\rfloor}}{a_{\lfloor \nu t\rfloor}} \xrightarrow{P} 0,$$

which finishes the proof.

□

4.4 Maximal claim size for three special models

As we have seen in the previous section, it is relatively easy to determine asymptotic behaviour of the maximal claim size $M(t)$ as long as we are able to determine properties of the random variables H_i and the leftover effect at time t, ε_t in (4.4). However, this is typically a rather technical task, highly dependent on an individual Poisson cluster model. Similarly as for $S(t)$ in previous chapter 3, in order to characterize limiting behaviour of $M(t)$, it is useful to determine behaviour of random variables H_i for each individual cluster model. In this section we introduce three special models.

But first, we want to find some useful conditions regarding the behaviour of

$$H_i = \max\left\{f(A_{ij}) : 0 \leq j \leq K_i\right\}$$

which represents the maximal claim size in the $i-$th cluster.

Note that the cluster H_i admits the following representation

$$H = \max\left\{X_j : 0 \leq j \leq K\right\},$$

for $(X_j)_{j\geq 0} = f(A_j)_{j\geq 0}$ i.i.d. copies of $X = f(A)$ and some integer valued K such that $\mathbb{E}[K] < \infty$. The tail behaviour of random maxima has been studied before (see [26], [13], [43]). We list below some of these results (conditions) in the case when X is in the maximum domain of attraction of Fréchet distribution, which are applicable to our setting. Recall, the maximum domain of attraction of Fréchet distribution consists of distributions whose right tail is regularly varying with index α, see for example Embrechts et a.l [16].

(MRV1) If X is regularly varying with index $\alpha > 0$ and independent of K, then $\mathbb{P}(H > x) \sim (\mathbb{E}[K]+1)\mathbb{P}(X > x)$ as $x \to \infty$, see [26], Lemma 5.1.;

(MRV2) If X and K are independent and both regularly varying with index $\alpha \in (1,2)$ and tail equivalent, then $\mathbb{P}(H > x) \sim (\mathbb{E}[K]+1)\mathbb{P}(X > x) + \mathbb{P}(K > x/\mathbb{E}[X])$ as $x \to \infty$, [13], Theorem 7.

The next Lemma slightly generalize previous two results because it includes all three domains of attraction (Fréchet, Gumbel and Weibull). It will be proved using one more time results in [6].

4.4.1 Lemma. *If X is in the maximum domain of attraction of Gumbel, Fréchet or Weibull distribution and K a nonnegative random variable with $\mathbb{E}K < \infty$ and independent of X, then H also belongs to the same domain of attraction.*

Proof. Recall that $X \in \mathrm{MDA}(G)$ means that for some sequences (a_n) and (b_n)

$$\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \xrightarrow{d} G, \quad n \to \infty,$$

for an i.i.d. sequence $X, X_1, X_2, \ldots$ of non–degenerate random variables.

This is equivalent to

$$n \cdot \mathbb{P}(X_1 > a_n x + b_n) \to -\log G(x), \quad n \to \infty.$$

Our aim is to show that

$$\frac{\bigvee_{i=1}^n H_i - b_{\lfloor(\mathbb{E}K+1)\cdot n\rfloor}}{a_{\lfloor(\mathbb{E}K+1)\cdot n\rfloor}} \xrightarrow{d} G, \quad n \to \infty,$$

for an i.i.d. sequence $H, H_1, H_2, \ldots$ of non–degenerate random variables. By Definition 4.3.1 the above expression is what we intend to prove.

First we will see that due to the independence of K and X

$$\bigvee_{i=1}^n H_i = \bigvee_{i=1}^n \bigvee_{j=0}^{K_i} X_{ij} \stackrel{d}{=} \bigvee_{i=1}^{K_1'+\cdots+K_n'+n} X_i, \tag{4.10}$$

where K_n and K_n', for every $n \in \mathbb{N}$ are i.i.d random variables, equally distributed as K and represent the number of points in a cluster. Without loss of generality, let $n = 2$. We then have the following:

$$\begin{aligned}
&\mathbb{P}\left(\bigvee_{i=0}^{K_1} X_i \vee \bigvee_{i=0}^{K_2} X_i \leq x\right) = \mathbb{P}\left(X_0 \leq x, \ldots, X_{K_1} \leq x, X_0' \leq x, \ldots, X_{K_2}' \leq x\right) \\
&= \sum_{k_1 \in \mathbb{N}} \sum_{k_2 \in \mathbb{N}} \mathbb{P}\left(X_0 \leq x, \ldots X_{k_2}' \leq x \middle| K_1 = k_1, K_2 = k_2\right) \mathbb{P}(K_1 = k_1)\mathbb{P}(K_2 = k_2) \\
&= \sum_{k_1 \in \mathbb{N}} \sum_{k_2 \in \mathbb{N}} \mathbb{P}\left(X_0 \leq x, \ldots X_{k_1} \leq x, \ldots X_{k_2}' \leq x\right) \mathbb{P}(K_1 = k_1)\mathbb{P}(K_2 = k_2) \\
&= \sum_{k_1 \in \mathbb{N}} \sum_{k_2 \in \mathbb{N}} \left[\mathbb{P}\left(X \leq x\right)\right]^{k_1 + k_2 + 2} \mathbb{P}(K_1 = k_1)\mathbb{P}(K_2 = k_2) \\
&= \sum_{k_1 \in \mathbb{N}} \mathbb{E}\left[\mathbb{P}\left(X \leq x\right)\right]^{k_1 + K_2 + 2} \mathbb{P}(K_1 = k_1) \\
&= \mathbb{E}\mathbb{E}\left[\mathbb{P}\left(X \leq x\right)\right]^{K_1 + K_2 + 2}.
\end{aligned}$$

The third equality follows from the independence between K_1, K_2 and $X_0, \ldots, X_{k_2}'$ and the fourth from the fact that $X_0, \ldots, X_{k_2}'$ are i.i.d. random variables. Similar, on the other side,

$$\begin{aligned}
&\mathbb{P}\left(\bigvee_{i=1}^{K_1' + K_2' + 2} X_i \leq x\right) = \mathbb{P}\left(X_1 \leq x, \ldots, X_{K_1' + K_2' + 2} \leq x\right) \\
&= \sum_{k \in \mathbb{N}} \mathbb{P}\left(X_1 \leq x, \ldots X_{k+2} \leq x \middle| K_1' + K_2' = k\right) \mathbb{P}(K_1 + K_2 = k) \\
&= \sum_{k \in \mathbb{N}} \left[\mathbb{P}\left(X \leq x\right)\right]^{k+2} \mathbb{P}(K_1' + K_2' = k) \\
&= \sum_{k \in \mathbb{N}} \left[\mathbb{P}\left(X \leq x\right)\right]^{k+2} \sum_{k_1 = 0}^{k} \mathbb{P}(K_1' = k_1)\mathbb{P}(K_2' = k - k_1) \\
&= \sum_{k \in \mathbb{N}} \sum_{k_1 = 0}^{k} \left[\mathbb{P}\left(X \leq x\right)\right]^{k+2} \mathbb{P}(K_1 = k_1)\mathbb{P}(K_2 = k - k_1) \\
&= \sum_{k_1 \in \mathbb{N}} \sum_{k_2 \in \mathbb{N}} \left[\mathbb{P}\left(X \leq x\right)\right]^{k_1 + k_2 + 2} \mathbb{P}(K_1 = k_1)\mathbb{P}(K_2 = k_2) \\
&= \mathbb{E}\mathbb{E}\left[\mathbb{P}\left(X \leq x\right)\right]^{K_1 + K_2 + 2}.
\end{aligned}$$

So we see that the (4.10) is true for $n = 2$. For general n the proof follows the same steps.

Denote now by $(T(n))$ the random walk generated by the iid sequence of nonnegative, integer valued random variables (K_n) equally distributed as K. Recall that

$\mathbb{E}K_1 < \infty$ by assumption. By the strong law of large numbers (SLLN) it follows that $\frac{T(n)}{n} \xrightarrow{as} \mathbb{E}K$, as $n \to \infty$ or, equivalently,

$$\frac{T(n)+n}{n \cdot (\mathbb{E}K+1)} \xrightarrow{as} 1, \quad n \to \infty \tag{4.11}$$

(see, for example Theorem 2.4.1, p. 63 in [14]). Next, we consider point process of the form

$$N_n = \sum_{i \geq 1} \delta_{\left(\frac{i}{g(n)}, X_{n,i}\right)}$$

for a nondecreasing function $g : (0, \infty) \to (0, \infty)$ tending to $+\infty$ as $n \to \infty$, with

$$X_{n,i} = \frac{X_i - b_{g(n)}}{a_{g(n)}},$$

where scaling and centering sequences (a_n) and (b_n) are given before. The assumption $X_1 \in \text{MDA}(G)$ is equivalent to $N_n \xrightarrow{d} N, \quad n \to \infty$ where N is $\text{PRM}(\lambda \times \mu_G)$ where λ denotes the Lebesgue measure and μ_G represents the measure induced by the nondecreasing function $\log G$, see for example Corollary 6.1, p. 183 in [39]. Due to the assumption $\mathbb{E}K < \infty$, by Proposition 3.1. (p. 57) in [39] and using (4.11) we have joint convergence of

$$\left(N_n, \frac{T(n)}{n \cdot (\mathbb{E}K+1)}\right) \xrightarrow{d} (N, 1),$$

as $n \to \infty$. The convergence takes place in $\mathcal{N}^{\#g}_{\mathbb{R}_{\geq 0} \times \mathbb{M}} \times \mathbb{R}_{\geq 0}$ endowed with the product topology i.e. topology of vague convergence of point measures and standard topology on $\mathbb{R}_{\geq 0}$ generated by the open intervals. Now, by Lemma 1 in [6] with $g(n) = n \cdot \mathbb{E}K$ and $Z_n = \frac{T(n)}{n \cdot (\mathbb{E}K+1)}$ we have

$$N_n|_{\left[0, \frac{T(n)}{n \cdot (\mathbb{E}K+1)}\right] \times \mathbb{M}} \xrightarrow{d} N|_{[0,1] \times \mathbb{M}},$$

as $n \to \infty$. Since

$$\mathbb{P}\left(\frac{\bigvee_{i=1}^{n} H_i - b_{\lfloor (\mathbb{E}K+1) \cdot n \rfloor}}{a_{\lfloor (\mathbb{E}K+1) \cdot n \rfloor}} \leq x\right) = \mathbb{P}\left(N_n|_{\left[0, \frac{T(n)}{n \cdot (\mathbb{E}K+1)}\right] \times \langle x, \infty]} = 0\right),$$

it follows

$$\mathbb{P}\left(N_n|_{\left[0, \frac{T(n)}{n \cdot (\mathbb{E}K+1)}\right] \times \langle x, \infty]} = 0\right) \to \mathbb{P}\left(N|_{[0,1] \times \langle x, \infty]} = 0\right) = G(x), \quad n \to \infty,$$

which proves the desired result.

□

4.4.1 Mixed binomial Poisson cluster model

Assume again that the clusters have the following form

$$G^{A_i} = \sum_{j=1}^{K_i} \delta_{W_{ij},A_{ij}}\,,$$

with $(K_i,(W_{ij})_{j\geq 1},(A_{ij})_{j\geq 0})_{i\geq 0}$ being an i.i.d. sequence. Assume moreover that $(A_{ij})_{j\geq 0}$ are i.i.d. for any fixed $i = 1,2,\ldots$ and that $(A_{ij})_{j\geq 0}$ is independent of $K_i,(W_{ij})_{j\geq 1}$ for all $i\geq 0$. As before we assume $\mathbb{E}[K]<\infty$.

4.4.1 Corollary. *Assume that X is in the maximum domain of attraction of Fréchet or Gumbel distribution, then the relation (4.7) holds.*

Proof. Recall, we denote the (increasing) number of summands in the r.h.s. term by $J(t) = \#\{(i,j) : 0\leq \Gamma_i\leq t, t<\Gamma_i+W_{ij}\}$. We can apply Proposition 4.3.1 after observing that $J(t) = o_P(t)$. It is sufficient to show the convergence to 0 of the ratio

$$\begin{aligned}\frac{\mathbb{E}[J(t)]}{t} &= \frac{\mathbb{E}[\#\{i,j : 0\leq\Gamma_i\leq t, t<\Gamma_i+W_{ij}\}]}{t}\\ &= \frac{\mathbb{E}\left[\sum_{0\leq\Gamma_i\leq t}\sum_{j=1}^{K_i}\mathbb{I}_{t\leq\Gamma_i+W_{ij}}\right]}{t}.\end{aligned}$$

Using calculation as in the proofs of corollaries 3.5.1 and 3.5.3 in chapter 3, we obtain an explicit formula for the r.h.s. term as

$$\begin{aligned}\frac{\mathbb{E}\left[\sum_{0\leq\Gamma_i\leq t}\sum_{j=1}^{K_i}\mathbb{I}_{t<\Gamma_i+W_{ij})}\right]}{t} &= \frac{\int_0^t\mathbb{E}\left[\sum_{j=1}^{K_i}\mathbb{I}_{W_{ij}>t-s})\right]\nu ds}{t}\\ &= \frac{\nu\int_0^t\mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{K_i}\mathbb{I}_{W_{ij}>t-s}\mid A_{i0}\right]\right]ds}{t}\\ &= \frac{\nu\int_0^t\mathbb{E}[m_A\mathbb{P}(W>x\mid A)]dx}{t}\to 0, \qquad t\to\infty,\end{aligned}$$

From a Cesaró argument, the result will follow if

$$\mathbb{E}[m_A\mathbb{P}(W>x\mid A)]\to 0, \qquad x\to\infty.$$

We can verify this because the random sequence $m_A\mathbb{P}(W>x\mid A)\to 0$ a.s. because $\mathbb{P}(W>x\mid A)\leq 1$ and because the sequence is dominated by m_A that is integrable. □

4.4.2 Renewal cluster model

Assume the same as in previous chapter that the clusters G^{A_i} have the following distribution

$$G^{A_i} = \sum_{j=1}^{K_i} \delta_{T_{ij}, A_{ij}} ,$$

where $(T_{ij})_j$ represents a renewal sequence

$$T_{ij} = W_{i1} + \cdots + W_{ij} ,$$

and we keep all the other assumptions from the model in previous subsection.

4.4.2 Corollary. *Assume that X is in the maximum domain of attraction of Fréchet or Gumbel distribution, then the relation (4.7) holds.*

Proof. The desired result follows if one can show that $J(t) = \#\{(i,j) : 0 \le \Gamma_i \le t, t < \Gamma_i + W_{i1} + \cdots + W_{ij}\} = o_P(t)$. Using the Markov's inequality, it is enough to check that $\mathbb{E}[J(t)]/t = o(1)$. We estimate the moment of $J(t)$ similarly as the one of ε_t in Corollary 3.5.5.

$$\begin{aligned}
\mathbb{E}[J_t] &= \mathbb{E}\left[\sum_{0 \le \Gamma_i \le t} \sum_{j=1}^{K_i} \mathbb{I}_{\Gamma_i + W_{i1} + \cdots + W_{ij} > t} f(A_{ij})\right] \\
&= \int_0^t \mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{W_{i1} + \cdots + W_{ij} > x}\right] \nu dx \\
&\le \int_0^t \mathbb{E}\left[K_i \mathbb{I}_{W_{i1} + \cdots + W_{iK_i} > x}\right] \nu dx.
\end{aligned}$$

We used the stochastic domination (3.20)

$$\sum_{j=1}^{K_i} \mathbb{I}_{W_{i1} + \cdots + W_{ij} > x} \le \mathbb{I}_{W_{i1} + \cdots + W_{iK_i} > x} K_i .$$

to obtain the last upper bound. From a Cesaró argument, the result will follow if

$$\mathbb{E}\left[K_i \mathbb{I}_{W_{i1} + \cdots + W_{iK_i} > x}\right] \to 0, \qquad x \to \infty.$$

One can actually check this negligibility property because the random sequence $K_i \mathbb{I}_{W_{i1} + \cdots + W_{iK_i} > x} \to 0$ a.s. by finiteness of $W_{i1} + \cdots + W_{iK_i}$ and because the sequence is dominated by K_i that is integrable. □

4.4.3 Marked Hawkes processes

Here is important to emphasize that in order to apply Lemma 4.4.1 we need independence between X and K. This is not the case for general marked Hawkes processes introduced in subsection 3.2.1 so we will observe the case when marks do not influence stochastic intensity, i.e. when $h(s,a) = h(s)$. Stochastic intensity is of the form

$$\lambda(s) = \nu + \sum_{\tau_i \leq t} h(t - \tau_i).$$

For this model, the clusters G^A are recursive aggregation of Cox processes, i.e. Poisson processes with random mean measure $\tilde{\mu} \times Q$ where $\tilde{\mu}$ has the form

$$\tilde{\mu}(B) = \int_B h(s)ds\,,$$

for some fertility (or self–exciting) function h. More precisely, if $N^A = \sum_{l=1}^{L} \delta_{\tau_l^1, A_l^1}$ is a Poisson process with random mean measure $\tilde{\mu} \times Q$, the cluster process corresponding to a point (τ, A) satisfies the following recursive relation

$$G^A = \sum_{l=1}^{L} \left(\delta_{\tau_l^1, A_l^1} + \theta_{\tau_l^1} G^{A_l^1} \right)\,,$$

where the sequence on the $(G^{A_l^1})_l$ on the right side is i.i.d., distributed as G^A and independent of N^A.

Under the assumption (2.17),

$$\kappa = \int h(s)ds < 1\,,$$

the total number of points in a cluster $K+1$ has Borel distribution with parameter κ. So, $\mathbb{E}(K+1) = \frac{1}{1-\kappa}$.

In this case, maximal claim size in one cluster is of the form

$$H \stackrel{d}{=} \bigvee_{j=0}^{K} X_j \qquad \text{or} \qquad H \stackrel{d}{=} X \vee \bigvee_{j=1}^{L} H_j.$$

Notice that in this case K and (X_j) are independent.

4.4.3 Corollary. *Assume that X is in the maximum domain of attraction of Fréchet or Gumbel distribution and*

$$\tilde{\mu}(t, \infty) \to 0, \qquad t \to \infty. \tag{4.12}$$

Then the relation (4.7) holds.

Proof. We will use Markov's inequality one more time to show that it is enough to check $\mathbb{E}[J(t)]/t = o(1)$.

$$\begin{aligned}
\mathbb{E}J(t) &= \mathbb{E}\Big[\sum_{\Gamma_i \le t}\sum_{j} \mathbb{I}_{\Gamma_i + T_{ij} > t}\Big] \\
&= \mathbb{E}\Big[\sum_{\tau_i \le t}\sum_{\tau_j > t} L^i \mathbb{I}_{\tau_i \to \tau_j}\Big] \\
&= \mathbb{E}\Big[\sum_{\tau_i \le t} \mathbb{E}\Big[\sum_{k=1}^{L^i} \mathbb{I}_{\tau_i + W_{ik} > t} \mid (\tau_i, A^i)_{i \ge 0}; \tau_i \le t\Big]\Big] \\
&= \mathbb{E}\left[\int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty)) N(ds, da)\right],
\end{aligned}$$

where $\tilde{\mu}_a((u,\infty)) = \int_u^\infty h(s,a)ds$. Observe that from projection theorem, see [8], Chapter 8, Theorem 3, the last expression equals to

$$\mathbb{E}\left[\int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty)) Q(da)\lambda(s)ds\right],$$

One can further bound this estimate by

$$\begin{aligned}
\mathbb{E}\left[\int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty)) Q(da)\lambda^*(s)ds\right] &= \int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty)) Q(da)\mathbb{E}[\lambda^*(s)]ds \\
&= \frac{\nu}{1-\kappa}\int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty)) Q(da)ds
\end{aligned}$$

Here we used Fubini's theorem, and the expression $\mathbb{E}\left[\lambda^*(s)\right] \equiv \nu/(1-\kappa)$. Observe that this expectation is constant since N^* is a stationary point process, to show that it equals $\nu/(1-\kappa)$. Now, we have (3.28),

$$\mathbb{E}J(t) \le \frac{\nu}{1-\kappa}\int_0^t \int_{\mathbb{M}} \tilde{\mu}_a((t-s,\infty)) Q(da)ds = \frac{\nu}{1-\kappa}\int_0^t \int_s^\infty h(u)duds\,.$$

Dividing the last expression by t and applying L'Hôpital's rule, proves the theorem for the nonstationary or pure Hawkes process. □

Bibliography

[1] K. K. Anderson. “A note on cumulative shock models”. *J. Appl. Probab.* 25.1 (1988), pp. 220–223.

[2] S. Asmussen and H. Albrecher. *Ruin probabilities*. Second. Vol. 14. Advanced Series on Statistical Science & Applied Probability. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2010.

[3] E. Bacry et al. “Some limit theorems for Hawkes processes and application to financial statistics”. *Stochastic Process. Appl.* 123.7 (2013), pp. 2475–2499.

[4] B. Basrak and H. Planinić. “A note on vague convergence of measures”. (2018). Available on Arxiv, arXiv:1803.07024v2.

[5] B. Basrak, O. Wintenberger, and P. Žugec. “On total claim amount for marked Poisson cluster models”. *Advances in Applied Probability* 51(2) (2019), pp. 541–569.

[6] B. Basrak and D. Špoljarić. “Extremes of random variables observed in renewal times”. *Statist. Probab. Lett.* 97 (2015), pp. 216–221.

[7] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*. Vol. 27. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1987.

[8] P. Brémaud. *Point processes and queues*. Springer-Verlag, New York-Berlin, 1981.

[9] P. Brémaud and L. Massoulié. “Stability of nonlinear Hawkes processes”. *Ann. Probab.* 24.3 (1996), pp. 1563–1588.

[10] V. Chavez - Demoulin and J. A. McGill. “High-frequency financial data modelling using Hawkes processes”. *Journal of Banking and Finance* 36 (2012), pp. 3415–3426.

[11] D. J. Daley. “Asymptotic properties of stationary point processes with generalized clusters”. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 21 (1972), pp. 65–76.

[12] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I, Vol. II*. Second. Probability and its Applications (New York). Springer-Verlag, New York, 2003.

[13] D. Denisov, S. Foss, and D. Korshunov. “Asymptotics of randomly stopped sums in the presence of heavy tails”. *Bernoulli* 16.4 (2010), pp. 971–994.

[14] R. Durrett. *Probability: theory and examples*. Fourth. Vol. 31. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[15] M. Egesdal et al. “Statistical and stochastic modelling of gang rivalries in Los Angeles”. (2010). SIAM Undergraduate Research, Online.

[16] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*. Vol. 33. Applications of Mathematics (New York). Springer-Verlag, Berlin, 1997.

[17] G. Faý et al. “Modeling teletraffic arrivals by a Poisson cluster process”. *Queueing Syst.* 54.2 (2006), pp. 121–140.

[18] A. Gut. “On the moments and limit distributions of some first passage times”. *The Annals of Probability* 2 (1974), pp. 277–308.

[19] A. Gut. *Stopped random walks*. Second. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2009.

[20] L. de Haan and A. Ferreira. *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[21] F. A. Haight and M. A. Breuer. “The Borel-Tanner distribution”. *Biometrika* 47 (1960), pp. 143–150.

[22] A. G. Hawkes. “Spectra of some self-exciting and mutually exciting point processes”. *Biometrika* 58 (1971), pp. 83–90.

[23] A. G. Hawkes and D. Oakes. “A cluster process representation of a self-exciting process”. *J. Appl. Probability* 11 (1974), pp. 493–503.

[24] H. Hult and G. Samorodnitsky. “Tail probabilities for infinite series of regularly varying random vectors”. *Bernoulli* 14.3 (2008), pp. 838–864.

[25] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*. Second. Vol. 288. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2003.

[26] A. H. Jessen and T. Mikosch. “Regularly varying functions”. *Publ. Inst. Math. (Beograd) (N.S.)* 80(94) (2006), pp. 171–192.

[27] O. Kallenberg. *Random measures*. Akademie-Verlag, Berlin; Academic Press, London-New York, 1976.

[28] D. Karabash and L. Zhu. “Limit theorems for marked Hawkes processes with application to a risk model”. *Stoch. Models* 31.3 (2015), pp. 433–451.

[29] J. Kerstan. “Teilprozesse Poissonscher Prozesse”. *Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes (Liblice, 1962)*. Publ. House Czech. Acad. Sci., Prague, 1964, pp. 377–403.

[30] M. Kirchner. “Perspectives on Hawkes processes”. PhD Thesis. ETH Zürich, 2017.

[31] P. Laub, T. Taimre, and P. Pollett. “Hawkes Processes”. (2015). Available on Arxiv, arXiv:1507.02822.

[32] P. A. W. Lewis and G. S. Shedler. “Simulation of nonhomogeneous Poisson processes with log linear rate function”. *Biometrika* 63.3 (1976), pp. 501–505.

[33] L. Massoulié. “Stability results for a general class of interacting point processes dynamics, and applications”. *Stochastic Process. Appl.* 75.1 (1998), pp. 1–30.

[34] T. Mikosch. *Non-life insurance mathematics*. Second. Universitext. Springer-Verlag, Berlin, 2009.

[35] M. Morariu-Patrichi and M. S. Pakkanen. “Hybrid marked point processes: characterisation, existence and uniqueness”. (2018). Available on Arxiv, arXiv:1707.06970.

[36] Y. Ogata. “Statistical models for earthquake occurrences and residual analysis for point processes”. *J. Amer. Statist. Assoc.* 83 (1988), pp. 9–27.

[37] V. Pernice et al. “How structure determines correlations in neuronal networks”. *PLoS Computational Biology* 85 (2012). :031916.

[38] S. I. Resnick. *Extreme values, regular variation, and point processes*. Vol. 4. Applied Probability. A Series of the Applied Probability Trust. Springer-Verlag, New York, 1987.

[39] S. I. Resnick. *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007.

[40] P. Reynaud-Bouret and S. Schbath. “Adaptive estimation for Hawkes processes; application to genome analysis”. *Ann. Statist.* 38 (2010), pp. 2781–2822.

[41] C. Y. Robert and J. Segers. “Tails of random sums of a heavy-tailed number of light-tailed terms”. *Insurance Math. Econom.* 43.1 (2008), pp. 85–92.

[42] G. Stabile and G. L. Torrisi. “Risk processes with non-stationary Hawkes claims arrivals”. *Methodol. Comput. Appl. Probab.* 12.3 (2010), pp. 415–429.

[43] C. Tillier and O. Wintenberger. “Regular variation of a random length sequence of random variables and application to risk assessment”. *Extremes* 21(1) (2018), pp. 27–56.

[44] L. Zhu. “Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims”. *Insurance Math. Econom.* 53.3 (2013), pp. 544–550.

Curriculum Vitae

Petra Žugec was born in Varaždin on August 17, 1984. She completed her secondary education in Varaždin and in 2003 enrolled in the Department of Mathematics, Faculty of Science, Zagreb. In October 2008, she graduated (BSc in Mathematics) with the thesis "Stable distributions", under the supervision of Izv. prof. dr. sc. Miljenko Huzak.

From August 2008 until September 2009 she worked as an actuary in an insurance company called Allianz Zagreb d.d. Since October 2009, she has been working as a teaching and a research assistant at the Faculty of Organization and Informatics, University of Zagreb.

In 2009, she enrolled in the PhD programme at the Department of Mathematics in Zagreb. During that period, she attended few conferences and colloquiums. She gave talks at: *12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics*, in July 2018 in Vilnius, Lithuania; *5th Croatian Mathematical Congress*, in June 2012 in Rijeka, Croatia; *15th Scientific-Professional Colloquium on Geometry and Graphics*, in September 2011 in Tuheljske toplice, Croatia.

Her interests are in applied probability; in particular, the theory of point processes and the extreme value theory. With Bojan Basrak and Olivier Wintenberger, she published a paper "On total claim amount for marked Poisson cluster model", *Advances in Applied Probability*, 2019.

Životopis

Petra Žugec rođena je 17. kolovoza 1984. godine u Varaždinu gdje je i završila srednju školu. Studij matematike upisala je 2003. godine na PMF–Matematičkom odjelu Svečilišta u Zagrebu. U listopadu 2008. godine diplomirala je na profilu diplomirani inženjer matematike, smjer Financijska i poslovna matematika s diplomskim radom pod naslovom "Stabilne distribucije" pod vodstvom Izv.prof.dr.sc. Miljenka Huzaka.

Od kolovoza 2008. godine do rujna 2009. godine radila je kao aktuar u Allianzu Zagreb d.d.. Od listopada 2008. godine radi u Varaždinu kao znanstveni novak i asistent na Fakultetu Organizacije i Informatike Sveučilišta u Zagrebu gdje radi i danas.

Godine 2009. upisuje zajednički sveučilišni poslijediplomski doktorski studij matematike na PMF-Matematičkom odsjeku. Od tada je sudjelovala na nekoliko skupova. Održala je kratke prezentacije na: *12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics* u srpnju 2018. u Vilniusu, Litva; *Petom hrvatskom matematičkom kongresu* u lipnju 2012. u Rijeci, Hrvatska; *15.om znanstveno-stručnom kolokviju za geometriju i grafiku* u rujnu 2011. u Tuheljskim toplicama, Hrvatska.

Njezino područje interesa je primjenjena vjerojatnost, posebno teorija točkovnih procesa i teorija extremnih vrijednosti. S Bojanom Basrakom i Olivierom Wintenbergerom objavila je članak "On total claim amount for marked Poisson cluster model", *Advances in Applied Probability*, 2019.