Diss. ETH No. 18403

Multivariate Hawkes Processes

THOMAS LINIGER

Diss. ETH No. 18403

Multivariate Hawkes Processes

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of

Doctor of Sciences

presented by

THOMAS JOSEF LINIGER

Dipl. Math. ETH
born 9. November 1978
citizen of Luzern, LU

accepted on the recommendation of

Prof. Dr. Paul Embrechts, examiner
Prof. Dr. Alexander McNeil, co-examiner
Prof. Dr. Pierre Brémaud, co-examiner

2009

to my parents

Abstract

This thesis addresses theoretical and practical questions arising in connection with multivariate, marked, linear Hawkes processes.

On the theoretical side, the following two main topics are discussed: the calculation of moment measures; and the existence and uniqueness of stationary solutions. Two different representations of Hawkes processes are used to answer these problems. First, a Hawkes process is constructed as a recursive Poisson cluster process. This reveals the underlying treelike structure and is used to derive the moment measures. Second, a Hawkes process is defined as the solution to a thinning problem and the associated minimal solution is introduced. Afterwards continuations are defined and their coupling properties are analyzed. These results serve as a basis for answering the questions of existence and uniqueness of the aforementioned thinning problem.

The formulation and proof of these statements requires several results from point process theory, some of which have to be extended first. Notable related topics that are discussed are: translation-invariant measures and their corresponding reduced versions; an extended notion of delta-functions and their usefulness in proving decompositions of moment measures; a concise construction of Poisson cluster fields and the proof of associated moment measure formulas; the formulation of the self-similarity structure of Hawkes processes; a general definition of strong solutions; and finally a general notion of hazard rates and its relation to intrinsic intensity functions.

On the practical side, issues that come up during the implementation of parameter estimation and simulation algorithms are addressed: the parameterization of a reasonably large family of Hawkes processes which is suitable for numerical calculations; the derivation of algorithms which perform the calculations in a numerically efficient way; concrete examples of Hawkes processes; and an illustrative financial case study.

Kurzfassung

Die vorliegende Doktorarbeit beschäftigt sich mit theoretischen und praktischen Fragen, die im Zusammenhang mit multivariaten, markierten Hawkes Prozessen auftauchen.

Auf der theoretischen Seite werden die folgenden zwei Themen behandelt: Die Berechnung der Momentenmasse sowie die Frage nach der Existenz und Eindeutigkeit von stationären Lösungen. Zu diesem Zweck werden zwei verschiedene Repräsentationen eines Hawkesprozesses verwendet: Einerseits wird erläutert, wie ein Hawkesprozess als rekursiver Poissonclusterprozess konstruiert werden kann. Diese Vorgehensweise verdeutlicht die zu Grunde liegende Baumstruktur von Hawkesprozessen und führt schliesslich zur Berechnung der Momentenmasse. Andererseits wird ein Hawkesprozess als Lösung eines bestimmten Verdünnungsproblems charakterisiert. Entsprechend wird eine zugehörige minimale Lösung dieses Problems definiert, sowie sogenannte Fortsetzungslösungen eingeführt. Desweiteren werden die Koppelungseigenschaften zweier Fortsetzungslösungen untersucht. Schliesslich werden diese Resultate verwendet, um die oben erwähnte Frage nach der Existenz und Eindeutigkeit zu beantworten.

Die Formulierung all dieser Aussagen und deren Beweise basiert auf verschiedenen Resultaten aus der Punktprozesstheorie, welche für den gegebenen Zweck angepasst und erweitert werden. Dies trifft insbesondere auf folgende Resultate zu: Translationsinvariante Masse und deren reduzierte Versionen; eine Erweiterung von Kronecker's Delta-Funktion und deren Nutzen in der Herleitung von Zerlegungen von Momentenmassen; eine kompakte Definition von Poissonclusterfeldern und zugehörige Formeln für die Momentenmasse; die Selbstähnlichkeitsstruktur von Hawkesprozessen; eine allgemeingültige Definition von starken Lösungen und schliesslich eine erweiterte Definition einer Hazardrate und der Zusammenhang mit intrinsischen Intensitäsfunktionen.

Auf der angewandten Seite werden folgende Themen betrachtet, welche bei der Implementation von Parameterschätz- und Simulationsalgorithmen von Relevanz sind: Die Auswahl einer angemessen umfassenden Teilfamilie von Hawkesprozessen, welche geeignet ist für numerische Berechnungen; die Herleitung von numerisch effizienten Algorithmen; sowie ein Fallbeispiel aus dem Finanzmarkt zur Illustration.

Acknowledgment

This thesis was carried out under the supervision of Alexander J. McNeil and Paul Embrechts, to both of whom I owe my sincerest gratitude, not only for their professional advice but also their personal support. To Alexander J. McNeil I am grateful for giving me the space of autonomy and creativity needed to enjoy the work on this thesis and the opportunity to visit him in Edinburgh for a semester when he left Zurich. To Paul Embrechts I am grateful for his encouragement and constant willingness to share his time. Both of my supervisors managed to create a pleasant and motivating atmosphere which made my postgraduate studies all the more enjoyable. Additional thanks go to Pierre Brémaud, who kindly acted as a co-examiner.

ETH Zurich has offered me an excellent academic environment and the opportunity to work as a teaching assistant. This has allowed me to deepen my knowledge in different interesting areas of probability theory and financial mathematics. Many thanks go also to my former colleagues from the financial and insurance mathematics research group at ETH Zurich.

Part of this thesis has been supported by the *National Center of Competence in Research* with their program *Financial Valuation and Risk Management* (NCCR FINRISK) and by the *European Science Foundation* (ESF) with their project *Advanced Mathematical Methods for Finance* (AMaMeF). The received financial support is gratefully acknowledged.

Contents

Chapter 1

Introduction

This thesis is concerned with theoretical questions that arise in the theory of multivariate, marked, linear Hawkes processes. It addresses among other issues the following two main topics: the calculation of second order moment measures; the existence and uniqueness of stationary solutions of the aforementioned class of point processes. A secondary, practical aim that lead towards this thesis was the development of flexible software for simulation and estimation of Hawkes processes.

Organization of the Thesis. After an introductory chapter; the five main chapters can be roughly split in two groups. Chapters 1–4 analyze Hawkes processes using their representation as recursive Poisson cluster processes. In contrast, Chapter 5 uses the theory of intensity processes, so that other aspects of Hawkes processes can be described. For the calculation of moment measures, the first point of view turns out to be the appropriate choice. The representation in terms of intensity processes however is more suitable for certain practical activities such as parameter estimation with the maximum likelihood method.

The six chapters contain the following material:

- This first chapter starts with a historical overview and gives many references. The references serve as a pointer to other results which are related to the material of the thesis. The second section introduces Hawkes processes in an informal way and the remainder focuses on application and numerical issues concerning Hawkes processes.

- The second chapter introduces point configurations, analyzes translation-

invariant measures and discusses several decompositions of product measures.

- The third chapter defines point processes, introduces cluster processes and derives moment measure formulas for general cluster processes as well as Poisson cluster processes.

- The fourth chapter defines Hawkes processes and several related processes. As a preparation for the next chapter, decompositions for product measures of Hawkes processes are derived.

- The fifth chapter uses the previous results to derive explicit and implicit formulas for the first and second order moment measures of Hawkes processes.

- The sixth chapter introduces canonical probability spaces, intensity measures and Hawkes intensity functions. This formalism is then used to give an alternative definition of a Hawkes process as a strong solution of a thinning problem. The issues of existence, uniqueness and stability of such solutions are then addressed.

As one can see from the list of contents, the statements are not kept together with their proofs. Instead, the proofs are collected at the end of each chapter, since they are sometimes quite long and would unnecessarily disturb the flow of the presentation. I hope the thesis is more readable using this somewhat unconventional layout.

Contributions. The thesis is built mainly on material from the book [DVJ03] and on a series of papers initiated by Pierre Brèmaud. Results that are new, partially new, or reformulations of previous results are scattered throughout the thesis and cannot be located in a single theorem. Each chapter starts with a short overview, a detailed list of references and a list of results that cannot be found readily in this form in the literature, at least to my knowledge. These introductions at the beginning of each chapter contain more specific details than I want to give here.

Hawkes processes are very versatile processes, not only from a practical point of view but also from a theoretical one. Some of the techniques used to analyze Hawkes processes are applicable in more general settings than is first apparent. To demonstrate the usefulness of some of these techniques, I often present them in a more general form than would be necessary for the

development of later results. This hopefully means that these techniques can be appreciated as results of independent interest. I give a short list of some of the results that I think may be interesting:

Chapter 1. The introductory chapter is not simply a summary but contains some material that cannot be found in the main part of the thesis. It is supposed to be a self-contained exposition giving the information which is relevant for empirical applications. Moreover, I give specific examples of Hawkes processes and list relevant numerical algorithms. Since I spent a considerable time with implementing algorithms for Hawkes processes, I included some of the hard learned lessons in the form of suitable parameterizations and algorithms. This means that the numerical algorithms are well-tested and ready for use.

Chapter 2. Notable results in Chapter 2 are Definition 2.13 and Theorem 2.14, which explain how one can isolate the one-dimensional symmetry of translation-invariant measures. Definition 2.28 gives an extended notion of delta-functions, and this allows one to give an alternative Definition 2.36 of factorial products of point configurations. Ordered partitions in Definition 2.24 and the associated decomposition of Theorem 2.32 allow one to write a ordinary product of point configuration as a sum of factorial products, see Theorem 2.42. It would be tedious to prove this result without an appropriate notation.

Chapter 3. Definition 3.9 of an extended event space allows one to enumerate the points of a point process in a consistent way, independent of the actual number of points; see Definition 3.12. The construction of a cluster family in Definition 3.19 is based on a family of generating clusters, see Definition 3.17, which are shifted appropriately. Theorem 3.22 gives a very convenient formula for expectations of an integral where the integrator is a factorial product measure. Using factorial decomposition from the first chapter, this leads to compact, but non-trivial moment measure formulas, see e.g. Equation (3.7). These formulas are more general than required later, but I left them in this form since they are useful in their own right.

Chapter 4. The third chapter introduces Hawkes processes and related point processes. The main challenge is not of mathematical nature but lies in finding a suitable, consistent notation for the large variety of processes derived from Hawkes processes. Proposition 4.14 and Corollary 4.22 formulate the self-similarity properties of a Hawkes process. Lemma 4.24 contains a series of

decompositions of product measures, which are needed for the next chapter. The reasoning behind these decomposition is explained at the beginning of that chapter.

Chapter 5. This is the last chapter in which Hawkes processes are considered from the point of view of recursive Poisson cluster processes. The main result is Theorem 5.4 for the second order moment measures of a Hawkes process. It gives compact formulas for these moment measures using the notation from Definition 4.9, which is related to multivariate convolutions.

Chapter 6. The last chapter starts with several definitions concerning measurable sets on point configuration spaces. Definitions 6.2 and 6.4 give non-standard, but equivalent definitions of predictable σ-algebra and predictable processes. Definition 6.12 differentiates between two different versions of unpredictable marks. The notion of an intrinsic intensity function in Definition 6.13 formalizes an idea which can be found implicitly in the literature. The usefulness of this definition can be seen e.g. in Corollary 6.40. Definition 6.33 of a strong solution is formulated without taking reference to Hawkes processes to point out that the same definition could be used also in the context of other point processes. Definition 6.38 introduces an extended notion of a hazard rate, which is then used to calculate the extinction probability in Proposition 6.39. Definition 6.46 describes in detail the construction of a stationary solution and calls the resulting process the minimal solution. Continuations, which are introduced in Definition 6.35, are a useful tool for some of the proofs but are also of interest if one deals with asymptotic simulation algorithms. Propositions 6.50 and 6.52 show that continuations are closely related to strong solutions and minimal solutions. The uniqueness of strong solutions is shown by splitting the probability space as explained in Definition 6.53. The method that is used to prove the uniqueness in Theorem 6.55 could also be used if the process were defined on an arbitrary probability space instead of a canonical probability space.

1.1 A Short History of Hawkes Processes

I give a short overview of the history of Hawkes processes and illustrate with a few examples how Hawkes processes have been applied in quite different areas. Because they are so versatile, Hawkes processes have successfully been used in such diverse fields as seismology, epidemiology, neurophysiology and network modeling. A very rich source of theoretical results as well as concrete examples of Hawkes processes and their applications can be found in the book [DVJ03].

Origin of Hawkes Processes. In the seventies, an interesting new point process model was introduced. It implemented the idea that is variously known as self-excitement, self-excitation or self-similarity. Clearly, the concept of branching and self-exciting behavior was not new. Indeed, processes with such characteristics had already been treated in [Ker64]. But what was missing was a concrete, tractable point process model with a self-exciting behavior. The search for a model for earthquake occurrences inspired Hawkes in his paper [Haw71] to consider self-exciting point process models. He gave a first precise definition of such a model. From then on, this type of process would commonly be known as a Hawkes process.

There are several equivalent ways of defining Hawkes processes. Originally, they were described based on the intensity process. Moreover, the fact that a Hawkes process has an underlying branching structure gives some additional insight. It is immediately clear that the branching coefficient ρ of the underlying branching process is important for the overall behavior of the process. But even without calling ρ the branching coefficient, Hawkes found in his paper [Haw71] the sufficient condition for existence, which is $\rho < 1$. This first paper mentions also some possible applications, such as epidemics or neuronal networks, and this turned out to be quite an accurate prediction. But it took some time until the first serious applications of Hawkes processes appeared, initiated by Ogata in seismology with his paper [Oga88] and related papers.

Sometime later, relations to other types of processes were discovered. The connection to Poisson cluster processes and to branching processes was worked out in the paper [HO74]. This observation allowed the authors to use results from the Galton-Watson theory, which lead to the interpretation of ρ as the branching coefficient. In the same paper, the authors also derive an implicit equation for the probability generating functional. An interesting special case occurs if the transfer function, sometimes also called the density function, is an exponential function. This special case was further analyzed in [Oak75]. It

turns out that under this assumption, a Hawkes process is actually a continuous time Markov process. This allows one to derive explicit expressions for the distributions of counts and intervals that can otherwise only be described in implicit form.

Further Progress. Already in his first paper, Hawkes briefly discussed the multivariate case. Parallel to new developments for univariate Hawkes processes, all relevant results were carried over to the multivariate case. In [Ada75], the probability generating functional, the distribution of the forward recurrence time and the distribution of the cluster length were extended to the multivariate case.

For a short time, there was some hope that Hawkes processes might play the same role as autoregressive time series in discrete time. One hoped that one could approximate the spectrum of arbitrary point process with a suitably chosen Hawkes process. But this hope was soon disappointed, since the spectral measure of a Hawkes process can not be chosen general enough for this purpose.

Nevertheless, because of the flexibility of Hawkes processes, this idea was pursued further in the paper [OA82]. The authors consider the case where the transfer function is a linear combination of Laguerre polynomials. In this case, the associated spectral measure has a particularly simple form. This idea was then empirically applied to earthquake catalogues in the paper [OAK82].

Relation to Branching Processes. There is also a different way of looking at a Hawkes process. If one ignores the location of the event, i.e. if one ignores the time dimension, a Hawkes process is simply a branching process. In other words, the events are related to each other in the way as ancestors and offspring are related to each other. As a mental picture, one can think of one or several trees, where each node corresponds to one of the events. The root nodes corresponds to immigrants and the branches correspond to the descendants of immigrants.

The fact that a Hawkes process has an underlying branching process allows one to use standard results from the theory of branching processes. Considering for example a univariate Hawkes process without marks, then the distribution of the cluster size can be calculated.

Otter-Dwass Theorem. The so-called Otter-Dwass formula is a recurrence relation for the distribution of the total number of events. It was first discovered in [Dwa69]. It can be applied in the case of a univariate Hawkes process without

marks and one can show that the total progeny has a so-called Borel-Tanner distribution, see [Tan53] or Section 7.2.2 of [NLJ05]. More about the general theory of branching-processes can be found in [Jag75].

An alternative proof for the Otter-Dwass formula was given in [Boy71], which uses an argument based on formal power series. Underlying is a certain functional equation for the moment generating function of the total number of events. Originally, this functional equation was found by [Goo49] and [Ott49], see also [Fel68].

Multivariate Extension of Otter-Dwass. In the multivariate case the situation is unfortunately not as pleasant. [Ott48] showed that the problem of finding the distribution of the total progeny in the univariate case can be reformulated as the solution of a certain functional equation. This functional equation can then be solved using Lagrange expansion. Later on, Lagrange expansion was generalized to the multivariate case by [Goo60]. This paper also mentions the connection to multivariate branching processes.

Independent of the advances in the theory of branching processes, [CS72] defined a new family of probability distribution, which they called the *Lagrange distributions*. Since both use the same mathematical tools, it is not surprising that there is an intimate connection between Lagrange distributions and branching processes. This connection was pointed out once more in [Goo75].

More about the family of Lagrangian probability distribution can be found in the book [CF06].

Parameter Estimation. Over the following years, Hawkes processes received more and more attention, especially in the context of seismology. Early development can be found in the papers [VJ75] and [Ada76]. One problem that arose was the lack of an efficient method for parameter estimation. For the first empirical applications, the parameters were estimated using spectral analysis techniques. But this method was considered less and less satisfactory.

What is now considered the classical maximum likelihood method for point processes was first described in the paper [Rub72], and applied to Hawkes processes in the papers [VJ78] and [Oza79]. In the accompanying paper [VJO82], the authors use the maximum likelihood method and analyze a catalogue of earthquake data. The properties of the maximum likelihood estimator was analyzed in [Oga78]. Since the computational power at that time was not sufficient for real-world data sets, the numerical parameter estimation remained difficult.

Advancement Within the Scope of Seismology. Hawkes processes did not get much attention for some time afterwards. But they were considered again in the paper [Oga88] as one candidate model among others for the prediction of earthquakes. In this paper, different models were compared with the help of residual analysis and Hawkes processes turned out to be superior to alternative models at that time. The specific Hawkes process used by Ogata is known today as the ETAS (Epidemic Type Aftershock-Sequences) model.

More and more extensions of the original specification of a Hawkes process were proposed. To this day, extensions of the original ETAS model are one of the standard tools in seismology. Basically, all these models belong to the family of Hawkes processes. A good review of the early evolvement of the Hawkes process in the context of seismology can be found in [UOM95] and [Oga99].

Ogata's Modified Thinning Algorithm. At a first glance, it seems easy to simulate a Hawkes process. Since a Hawkes process can be defined via its stochastic intensity process, the following observation leads automatically to a simulation procedure: Assume the past of a Hawkes process is known up to a given time, say t. Then the Hawkes process behaves as it were a Poisson process, at least in the short time period after time t until the first event occurs.

This leads to a straight forward simulation method which is not specific to Hawkes processes. At least theoretically, this method can be used to simulate all point process models which are defined by their stochastic intensity process.

This simulation algorithm and its theoretical foundation go back to a thinning procedure given [LS79]. The general idea is to construct a Hawkes process as the stochastic thinning of a homogeneous Poisson process according to the intensity process. In the context of Hawkes processes, this simulation method was first used in [Oga81]. This algorithm is sometimes called *Ogata's modified thinning algorithm*. Later on, it was extended to the multivariate case and to space-time Hawkes processes in [MVJ92] and [Oga98].

For practical applications, this is still the standard algorithm. A more comprehensive description can be found in the book [DVJ03].

Simulation of Stationary Hawkes Process. Often one wants to simulate the stationary version of the process in a finite time window. Unfortunately, the standard method for the simulation method described above does not work in this case, as the past of the process is not know and cannot be simulated, at least not completely.

If one ignores the past of the process and simply starts to simulate the process at some given time, one speaks about an *approximate simulation*. In this case, one is actually simulating a transient version and not the stationary version of the process. But if one simulates for a long enough time interval, the so-called *burn-in period*, then the transient version converges to the stationary one. This is a consequence of the stability properties, which are described below. Since the deviation between the transient and stationary process becomes negligible, this simulation method is good enough for practical applications.

Perfect Simulation. A simulation method which directly simulates the stationary version without approximation is a so-called *perfect simulation* method. The idea is to incorporate somehow the effect of past observations without actually simulating the past of the process. In the context of point processes, this type of simulation method has first been described in [BK02].

The idea is to use a so called *coupling from the past* algorithm. The original idea goes has been developed for Markov chains and goes back to [PW96], see also [FT98].

A coupling from the past method for a univariate Hawkes process was first given in [MR05] This paper also explains that the general coupling from the past method is not good enough. An additional problem specific to Hawkes processes arises, but this can be solved using what the authors call a *dominated coupling from the past* algorithm. This extended version of the original coupling from the past algorithm has first been described in [KM00].

Once an exact simulation algorithm had been established, it was possible to compare the performance of the approximate simulation algorithm. In the paper [MR06], various measures for the deviation of the approximate from the exact simulation algorithm are given.

Development of Theoretical Foundation. The theoretical foundation of Hawkes processes was lifted to a higher level by a series of papers initiated by Pierre Brémaud. The first important paper in the series is [BM96], where sufficient conditions for the existence of stationary versions were given. Actually, the more general case of non-linear Hawkes processes was considered. In the precursor work [BM94] by the same authors, similar techniques had been used, but not in the context of Hawkes processes.

This paper also discusses the conditions under which a non-stationary version converges to an associated stationary version. One of the essential ideas in the proof is an iterative procedure in the spirit of the proof of Picard's existence

theorem for ordinary differential equations. Actually, the same technique had already been used by [Ker64] to derive similar results, long before the modern formulation of point process theory. Moreover, some results concerning multivariate Hawkes processes were given.

The existence and stability results were then extended in [Mas98] to the multivariate case and to the case of marked Hawkes processes. Some more papers followed, which extended the theoretical foundation further. An important step was also the calculation of the spectral measures in [BM02].

An interesting situation occurs if the branching coefficient is exactly one, i.e. if the branching process is critical. In [BM01], it was shown that it is possible to construct Hawkes processes in this critical domain. Such a process is in some sense self-sustaining, i.e. it does not need new immigrants arriving over time any more.

For approximate simulation methods it is important to have a rough estimate of how large the burn-in period should be. Therefore, one needs to know more about the rate of convergence of a transient Hawkes process to its stationary version. In the univariate case, bounds for the rate of convergence were derived in [BNT02]. The results were then extended in [Tor02] to more general processes, including multivariate Hawkes processes.

More Abstract Notions of Self-Excitement. The concept of self-excitement was considered from a more abstract point of view in [KS96]. The authors define self-exciting behavior as an abstract property of the stochastic intensity process which defines the point process. This more general class of positively self-exciting point processes also contains Hawkes processes.

In the paper [VJ05], a class of self-similar measures is introduced and analyzed. But self-similarity is a concept defined for random measures and has a-priori nothing to do with self-excitement. But it was shown in this paper that a subset of Hawkes processes are also self-similar.

Applications in Network Modeling. Hawkes processes, especially non-linear Hawkes processes, define quite a large class of point processes models. It is therefore no surprise that several classical models can be reinterpreted as Hawkes processes, sometimes with a few modifications. A list of such models can be found in the paper [Mas98], For example the loss network model studied by [Kel85] and [FY96] is actually a non-linear Hawkes process. Moreover, this paper proposes Hawkes processes for the description of spontaneously excitable

random media. Another area where Hawkes processes might be used are neural networks. A suitable specification of a Hawkes process is given in [BM96].

Applications in Finance. If one wants to model stochastic processes in finance with Hawkes processes, there is an additional difficulty, that people using Hawkes processes in seismology and other areas of science did not come across. In seismology, one has the advantage that the theoretical foundation of geophysics gives some strong directions of how one should choose the parameterizing functions of a Hawkes process. But as this is not the case for financial data, one first has to gain some experience about suitable choices of Hawkes processes.

Empirical comparisons suggest that Hawkes processes have some of the typical characteristics of financial time series. Financial data have been analyzed using Hawkes processes e.g. in the papers [CDDM05] and [Bow07]. In the book [MFE05], self-exciting processes are used for the calculation of conditional risk measures, such as the Value-at-Risk.

Hawkes processes have also been used in connection with risk processes, that is to model the surplus process of an insurance portfolio. Some first results were obtained in [AA06]. The authors derived a Cramér-Lundberg approximation of the ruin probability and other asymptotic estimates. Actually, the paper considers only shot noise Cox processes, and not a Hawkes process. But the results were extended to Hawkes processes in [ST08].

Applications in Credit Default Modeling. Another area of finance where Hawkes processes have been considered is credit default modeling. Hawkes processes have been proposed as models for the arrival of company defaults in a bond portfolio, starting with the papers [GT05] and [GG05].

A important observation is also the fact that Hawkes processes of Markov type, i.e. Hawkes processes with exponential transfer functions, are actually affine jump-diffusion processes. This relation was described first in [EGG07]. With the help of the theory of affine jump-diffusions, one can then analyze price processes related to certain credit derivatives.

Pricing formulas in the context of affine jump-diffusion processes have been derived in [DPS00]. These processes have the advantage that analytical formulas for Laplace and Fourier transforms can be found. This leads to analytical solutions for various pricing problems.

1.2 Hawkes Processes Informally

This section gives a short introduction to Hawkes processes and discusses some of the main notions. The focus will be on a specific class of multivariate, marked Hawkes processes and associated algorithms. It is *not* the purpose of this chapter to prove results nor to introduce the definitions in a rigorous way. Instead the concepts are presented intuitively.

The targeted audience for this introduction is therefore anyone who wants to have a quick overview of what Hawkes processes are all about and how they could be used in practice, without entering into technical details. The following definitions, propositions and theorems are all stated in an informal way, but I give references to the rigorous formulation in the main part of the thesis.

There are basically two ways how one can define and represent Hawkes processes; and each one has its advantages and disadvantages. This thesis covers both of them: Firstly, a Hawkes process can be constructed as a recursive Poisson cluster process, see Chapter 4. Secondly, it can be defined in terms of its intensity function, see Chapter 6. In this first chapter I will only consider the latter representation, i.e. I will give an informal definition of Hawkes processes using intensity processes.

Classification of Hawkes Models. There is no generally accepted definition of a Hawkes process. The original definition given by Hawkes in his paper [Haw71] is quite specific. Today, one usually calls a far larger class of point process models Hawkes processes. This thesis treats only *linear* Hawkes processes, see Definitions 4.20 and 6.19. But one can also define more general, non-linear Hawkes processes.

An additional distinction can be made whether the Hawkes process is univariate or multivariate or whether it is unmarked or marked. A mark is an additional value attached to each of the points and carries some information about this point. In this introduction I consider the case of multivariate, marked Hawkes processes and assume for simplicity that the marks are real numbers.

It is clear that if one wants to go the full way and finally implement algorithms for Hawkes processes, one has to restrict oneself to a reasonable subfamily of all possible Hawkes process models. This is the reason why I consider in this introduction not the most general case of multivariate Hawkes processes. But the subset of Hawkes processes defined below is still large enough for the majority of serious empirical applications.

Parameterization and Parameter Domain. I will consider only Hawkes processes whose transfer functions are separable in some specific way. A similar, but less restrictive condition is also discussed in the main part of the thesis; see Definition 6.25. One can easily find different parameterizations of a given Hawkes intensity function, which all lead to exactly the same Hawkes process. From a theoretical point of view this does not matter, but for numerical procedures it can make a big difference. Let me explain in a few words what sort of problems an inappropriate parameterization can bring.

As it is the case for all non-trivial models, the model parameter vector, say $\boldsymbol{\theta}$, cannot be arbitrarily but has to lie in some set of valid values, say $D \subseteq \mathbb{R}^n$. The set D consists of all parameter values where the model is well-defined. For Hawkes processes, the restrictions on the parameters are fortunately quite tractable. The precise results can be found in Chapter 6, but I will give a summary below.

Even though these conditions are mathematical fairly simple, they are not trivial. It is clear that choosing a parameter vector $\boldsymbol{\theta}$ that lies outside of D leads most likely to meaningless or absurd results. It is also clear that if $\boldsymbol{\theta}$ is inside D but close to the boundary, the same numerical problems may occur, even though the parameter vector is valid from a mathematical point of view.

Assume one wants to estimate the parameters of a Hawkes model given some data using a standard numerical minimization algorithms. It is then important that the restrictions on $\boldsymbol{\theta}$ are as simple as possible, since otherwise it is difficult to prevent the minimization algorithm from moving outside the domain of valid parameter values D. For the same reason, the range of valid values for one parameter should depend on as few other parameters as possible.

Hawkes processes with separable transfer functions are especially suitable for numerical implementation. This is the reason why I restrict myself in this introduction to these functions. What this actually means is explained in Definition 1.9 below.

What Does Multivariate Mean?. The term *multivariate point process* can mean two different things. To avoid any misunderstanding, I will call the two types of multivariate point processes *genuine* and *pseudo* multivariate point processes. An informal definition is:

1.1 Definition (Multivariate point process). In both cases, let $d \geq 1$ be the number of components. Assume the points are indexed over some countable set $\mathscr{I}$.

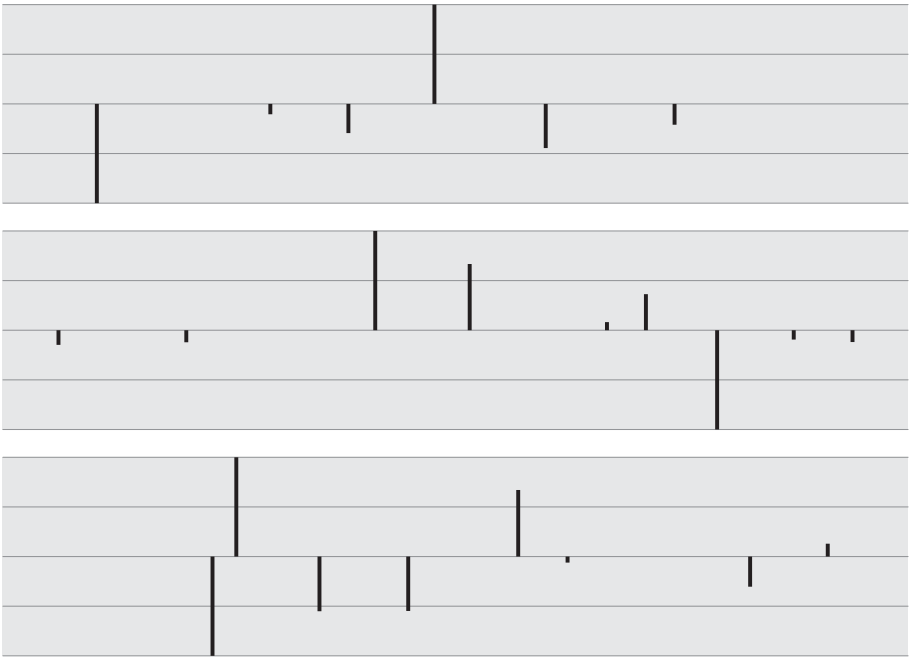

Figure 1.1: Genuine multivariate point process

(1) *Genuine multivariate point process.* The events are triples of the form (t_i, d_i, x_i), where $t_i \in \mathbb{R}$, $d_i \in \{1, \ldots, d\}$ and $x_i \in \mathbb{R}$, for $i \in \mathscr{I}$. The three components have the interpretation:

t_i : time,
d_i : component index,
x_i : mark value.

Consider the i-th event, which occurred at time t_i: The component index d_i looks like it is another mark in addition to x_i. But this would be the wrong interpretation. Instead, the component index d_i assigns the event t_i to the component d_i.

(2) *Pseudo multivariate point process.* The events are given by tuples of the form $(t_i, x_{i,1}, \ldots, x_{i,d})$, where $x_{i,j} \in \mathbb{R}$, for $i \in \mathscr{I}$, $j \in \{1, \ldots, d\}$. The

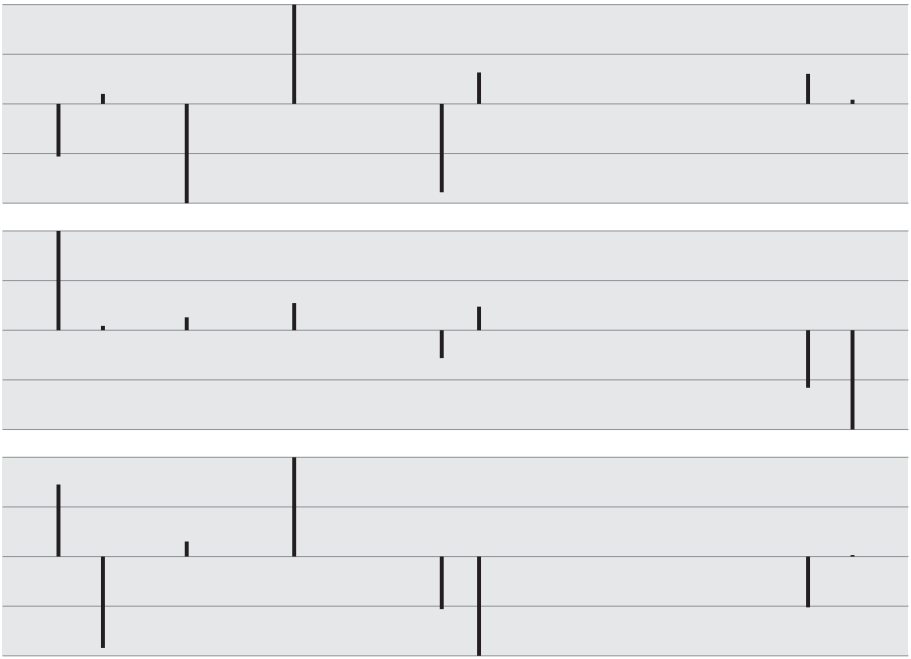

Figure 1.2: Pseudo multivariate point process

components have the interpretation:

t_i : time,
$x_{i,j}$: j-component of the vector-valued mark $\boldsymbol{x}_i$.

Consider again the i-th event, which occurred at time t_i: This time, there is no component information d_i and conceptually one may think that the event t_i lies in component one. On the other hand, there are d mark values $x_{i,j}$, which define a mark vector $\boldsymbol{x}_i := (x_{i,1}, \ldots, x_{i,d})$. ♢

The difference between a genuine and a pseudo multivariate point process is also illustrated in the Figures 1.1 and 1.2. Both times, a realization of a three-dimensional point process is displayed.

Note that Definition 1.1 is an informal definition. Compare this with the formal Definitions 2.33 and 3.4. Indeed, there are some technical conditions that need to be mentioned: Clearly, the time and mark values need to be random variables, i.e. they need to be measurable functions defined on some

probability space. Moreover, the point process should not be "explosive", i.e. there should be no subsequence, say $(k_i)_{i\geq 1} \subseteq \mathscr{I}$, such that the limit $\lim_{i\to\infty} t_{k_i}$ exists. In order to exclude processes of this type, one usually assumes that the point process is locally-finite, i.e. that in every bounded time interval there are only finitely many points. Moreover, a Hawkes process is always a simple point process. This means that no two points occur at the same time, i.e. there are no multiple points. Hence, one always has $t_k \neq t_l$, for $k \neq l$,

At the moment it does not matter what the set $\mathscr{I}$ is. But when I discuss concrete algorithms later, it is better to have a specific set. Since $\mathscr{I}$ is countable, one can always assume that $\mathscr{I} := \mathbb{Z}$. In this case, assume that the points are ordered in the obvious way, i.e. that $t_k < t_l$, if $k < l$.

Event Spaces. From a mathematical point of view, a pseudo multivariate point process is actually a *univariate* point process with mark space $\mathbb{R}^d$. When I give the rigorous definition of point configurations in the Chapter 2, I will not distinguish between genuine and pseudo multivariate processes. Instead, I will introduce the more general concept of *event spaces* and *mark spaces*.

A pseudo multivariate process is then point process with event space $\mathbb{E} := \mathbb{R}$ and mark space $\mathbb{X} := \mathbb{R}^d$. On the other hand, a genuine multivariate process is a point process with event space $\mathbb{E} := (\mathbb{R}, \ldots, \mathbb{R})$ and mark space $\mathbb{X} := (\mathbb{R}, \ldots, \mathbb{R})$. I mention event and mark spaces only to point out that the distinction between genuine and pseudo multivariate point processes can be avoided if one uses a more abstract notion of point processes.

In this introduction I will not use more abstract concepts and I therefore always distinguish between genuine and pseudo multivariate point processes. This also means that I have to give separate definitions, once for the genuine case and once for the pseudo multivariate case.

Analogy to Time Series. A short remark for the reader who is more familiar with multivariate time series than with multivariate point processes:

1.2 Remark (Alternative interpretation). A univariate, marked point process is in some sense a time series with random, unevenly spaced time intervals. Similarly, a pseudo multivariate point process is in some sense a multivariate time series with unevenly spaced events in time. On the other hand, a genuine multivariate point process is conceptually something else and one cannot easily find an analogy in terms of time series. ♢

Integral Notation. In dealing with point processes it is common to use integral notation to express sums over all events of the point process. The deeper reason for this notation is the fact that a point process is actually a random measure, see Definitions 3.1 and 3.4. The following notation treats only the case of a univariate point process, but the extension to a multivariate one is trivial:

1.3 NOTATION (INTEGRAL NOTATION). Let (t_i, x_i) be an enumeration of all events of N, where i lies in some countable set $\mathscr{I}$. The usual notation for the sum over all events is:

$$\int_{\mathbb{R}\times\mathbb{R}} f(t,x)N(dt \times dx) \equiv \sum_{i\in\mathscr{I}} f(t_i, x_i),$$

where f is some arbitrary function such that the sum is well-defined. $\diamondsuit$

Hawkes Intensity Function. As explained, one can define a Hawkes process either as a recursive Poisson cluster process, or one can define a Hawkes process by specifying its intensity function. In this introduction, I will only look at the second representation. The reason is that in order to calculate the likelihood function one needs the intensity process, and therefore this is the suitable representation if one wants to fit a Hawkes model to a set of empirical data.

In Chapter 6 so-called *transfer functions* are introduced, see Definition 6.19. I will not give the definition of transfer functions here, since in this first chapter I consider only a special case of transfer functions, so-called *separable* transfer functions, see Definition 6.25. To avoid technical notions, I directly define the Hawkes intensity process using decay and impact functions, see Definition 1.9 below.

Immigration Intensity. Before I give the definition of the Hawkes intensity function, I introduce and explain all the relevant pieces in turn. The immigration intensities are just constants, but what is important is their interpretation: Every event of a Hawkes process is either an immigrant or a descendant. The immigration intensities govern the frequency at which new immigrants arrive in each of the components:

1.4 Definition (Immigration intensity). Let $d \geq 1$ be the number of components. Depending on the version of multivariate Hawkes process, assume the following constants are given:

(1) *Genuine multivariate.* Let $\eta_j \geq 0$, for $j \in \{1, \ldots, d\}$.

(2) *Pseudo multivariate.* Let $\eta \geq 0$.

The parameters above are called the *immigration intensities.* ♢

Decay Function. A Hawkes process is characterized by the following property: Whenever an event occurs, the intensity is increased for some time, i.e. events arrive at a higher frequency. This increase causes secondary events, which are called descendants. How fast this effect decays in time is governed by the so-called *decay functions*:

1.5 Definition (Decay function). Assume an event has occurred at time $s \in \mathbb{R}$ and fix some $t > s$. Let $\Delta t := t - s$ denote the time lag. The functions w_j, w given below are called the *decay functions*:

(1) *Genuine multivariate.* Let w_j, for $j \in \{1, \ldots, d\}$, be functions of the form:

$$w_j : \mathbb{R}_+ \to \mathbb{R}_+.$$

Interpretation: If there is an event at time s, the intensity of component j at time t is increased proportional to $w_j(\Delta t)$.

(2) *Pseudo multivariate.* Let w be a function of the form:

$$w : \mathbb{R}_+ \to \mathbb{R}_+.$$

Interpretation: If there is an event at time s, the intensity at time t is increased proportional to $w(\Delta t)$. ♢

Note that the intensity is not increased by the amount $w(\Delta t)$, but only by a value proportional to $w(\Delta t)$. This becomes clear if one looks at Definition 1.9, since $w(\Delta t)$ appears as multiplicative factor in the definition of the intensity function. When numerical algorithms are discussed later, the functions given in the next definition will be quite convenient:

1.6 Definition (Related functions). Let $w_{j=1,\ldots,d}$ and w be decay functions, as in Definition 1.5.

(i) *Cumulative decay function.* Define for $j \in \{1, \ldots, d\}$ and $t \geq 0$ the functions:

$$\bar{w}_j(t) := \int_0^t w_j(s)ds \qquad \text{and} \qquad \bar{w}(t) := \int_0^t w(s)ds.$$

The first part refers to the genuine multivariate case and the second part to the pseudo multivariate case.

(ii) *Quantile function.* Let $\varepsilon > 0$, usually a very small number. In the genuine case, define for all $j \in \{1, \ldots, d\}$ the quantile $q_j > 0$ such that the following equation is satisfied:

$$\int_{q_j}^{\infty} w_j(t)dt = \varepsilon.$$

In the pseudo multivariate case, take w instead of w_j and define $q > 0$ correspondingly. ♢

Impact Function. The amount by which the intensity increases after an event does not only depend on the time lag but also on the mark value of the triggering event. The influence of the mark value is governed by the impact functions. The impact functions describe again only the relative effect of an event, since they appear as a multiplicative factor in the definition of the intensity function.

1.7 Definition (Impact function). The functions g_k given below are called the *impact functions*:

(1) *Genuine multivariate.* Let g_k, for $k \in \{1, \ldots, d\}$, be a family of measurable functions of the form $g_k : \mathbb{R} \to \mathbb{R}_+$. Interpretation: Assume the triggering event is in component d_i and has mark value x_i. The intensity of all other components is then increased proportional to $g_{d_i}(x_i)$.

(2) *Pseudo multivariate.* The family of impact function $g_{k=1,\ldots,d}$ is defined in the same way as above; but the interpretation is different: Assume the triggering event has the vector-valued mark $\boldsymbol{x}_i \in \mathbb{R}^d$. The intensity is then increased proportional to $\prod_{k=1}^{d} g_k(x_{i,k})$. ♢

Note that in the genuine case, there are d impact functions but only one mark value x_i. One always takes the impact function g_{d_i} corresponding to the component d_i of the triggering event. This emphasizes once more the fundamental difference between the genuine and the pseudo multivariate case.

1.8 Definition (Branching coefficients). The parameters $\vartheta_{jk}, \vartheta$ given below are called the *branching coefficients* and the matrix Q is called the *branching matrix*:

(1) *Genuine multivariate.* Let $\vartheta_{jk} \geq 0$ be constants, for $j, k \in \{1, \dots, d\}$, and define the matrix:

$$Q := \Big(\vartheta_{jk}; j, k \in \{1, \dots, d\}\Big).$$

Interpretation: Given that there is an event in component k, the intensity of component j is increased proportional to ϑ_{jk}.

(2) *Pseudo multivariate.* Let $\vartheta \geq 0$ be a constant, and for consistency with the previous case, define the 1×1-matrix $Q := \vartheta$. Interpretation: Every event increases the intensity by an amount proportional to ϑ. ♢

Now all required ingredients have been introduced and I can finally define the Hawkes intensity function. A rigorous definition of an intensity process is given in Definition 6.10 and Remark 6.11. But roughly speaking, the intensity value $\lambda(t)$ at time t measures the probability that an event occurs in the infinitesimal time interval $[t, t + dt)$.

1.9 Definition (Hawkes intensity process). Let $d \geq 1$ be the number of components.

(1) *Genuine multivariate.* Let w_j, g_k, η_j and ϑ_{jk} be as in the definitions above. Define for $j \in \{1, \dots, d\}$ and $t \in \mathbb{R}$ the family of Hawkes intensity processes by:

$$\lambda_j(t) := \eta_j + \sum_{k=1}^{d} \vartheta_{jk} \int_{(-\infty,t)\times\mathbb{R}} w_j(t-s) g_k(x) N_k(ds \times dx). \tag{1.1}$$

(2) *Pseudo multivariate.* Let w, g_k, η and ϑ be as in the definitions above. Define for $t \in \mathbb{R}$ the Hawkes intensity process by:

$$\lambda(t) := \eta + \vartheta \int_{(-\infty,t)\times\mathbb{R}^d} w(t-s) \Big[\prod_{k=1}^{d} g_k(x_k)\Big] N(ds \times d\boldsymbol{x}). \tag{1.2}$$

♢

Note that a genuine multivariate Hawkes process has d intensity processes, whereas a pseudo multivariate Hawkes processes has only one intensity process. Once more, this indicates that the two versions are quite different.

1.10 Remark (Hawkes intensity process). The following comments refer to the genuine case but apply to the pseudo case as well:

(i) The functions w_j, g_k and the parameter ϑ_{jk} are determined only up to

multiplication by a constant. Below additional normalizing conditions are introduced, so that in the end all terms are uniquely determined.

(*ii*) The case where one has a mixture of marked and unmarked components is included in this specification for the Hawkes intensity. Indeed, assume component $k \in \{1, \dots, d\}$ has no marks. Then define the *void* impact function:

$$g_k := 1.$$

With this impact function, the unmarked component k can be treated as if it were a marked component. The void impact function is introduced only for notational convenience in order to treat all components in a unified way. ♢

Mark Distribution. The intensity process given above is the so-called time-intensity process. It describes only the dynamics of the ground process, i.e. the process without the marks. For a full specification of a marked point process, one needs to know the so-called time-space-intensity process. For the theoretical background, see Definitions 6.9, 6.10 and 6.19.

At this point, it is not necessary to introduce another theoretical concept. Indeed, the Hawkes processes I consider here have marks which are independent of the past of the process. In other words, the mark distribution is not influenced by previous events and stays always the same. It is therefore enough to specify the mark distribution, and then the complete dynamics of the Hawkes process are defined.

1.11 Definition (Mark distribution). Let f_j, for $j \in \{1, \dots, d\}$, be probability densities on $\mathbb{R}$, the so-called *mark densities*.

(*1*) *Genuine multivariate.* Given that there is an event in component j at time t, the mark X associated to this event is independent of the past of the process and has distribution:

$$X \sim f_j.$$

(*2*) *Pseudo multivariate.* Given that there is an event t, the vector-valued mark $\boldsymbol{X} \in \mathbb{R}^d$ is obtained as follows: The components X_j are independent of each other and also independent of the past of the process and the

distribution of $\boldsymbol{X}$ is given by:

$$\boldsymbol{X} = \big(X_1, \dots, X_d\big), \text{ where } X_j \sim f_j. \qquad \diamondsuit$$

Note that in both cases there are d mark densities f_j, but the way they appear in the definition is quite different.

As explained in Remark 1.10, one needs additional assumptions if one wants the branching coefficients ϑ_{jk} to be uniquely determined. To this end, I assume that the decay functions w_j and the impact functions g_k always satisfy the following conditions:

1.12 Convention (Normalizing conditions). In the genuine multivariate case, assume that the following two normalizing conditions are satisfied:

$$\int_0^\infty w_j(t)dt = 1 \qquad \text{and} \qquad \int_{-\infty}^\infty g_k(x) f_k(x) dx = 1,$$

for all $j, k \in \{1, \dots, d\}$. In the pseudo multivariate case, replace w_j by w. $\diamondsuit$

Let me explain the reasoning behind these normalizing conditions:

1.13 Remark (Advantages and disadvantages). The normalizing conditions are specifically designed such that the parameters ϑ_{jk} correspond to the branching coefficients of the underlying branching process. This is the reason why ϑ_{jk} are called the *branching coefficients* in the first place. Without these normalizing conditions, the parameters ϑ_{jk} could not be interpreted in this way. There are advantages and disadvantages that come with these normalizing conditions:

(*i*) The big advantage is that the branching coefficients ϑ_{jk} appear explicitly as parameters in the model. As explained below, the branching coefficients determining whether the Hawkes model is well-defined or not. Without these normalizing conditions, the branching coefficients would be non-trivial functions of the model parameters.

(*ii*) There is also a disadvantage: Assume f_k is the mark distribution of component k. Then the associated impact function g_k needs to be specifically adapted such that the normalizing condition is satisfied. Especially for numerical implementation, this means a lot of additional work; but the benefits of this parameterization by far outweigh the disadvantages. $\diamondsuit$

To get a better understanding of what these normalizing conditions imply in a concrete situation, have a look at the list of examples of distributions and impact functions given in Section 1.3 below.

Hawkes Process. I give in Chapter 6 a precise definition of a Hawkes process. Actually, I will consider so-called *strong solutions* of a certain thinning equation. A Hawkes process is then defined as the solution to this thinning equation, see Definition 6.33. It is easier to derive theoretical properties if the Hawkes process is defined on a specific probability space, which I called the *canonical probability space*, see Definitions 6.3, 6.28 and 6.30.

For the purposes of this chapter, the underlying probability space is not important, and I therefore define a Hawkes process directly with the help of the associated Hawkes intensity process. See also Definition 6.10 for a precise definition of an intensity process.

1.14 Definition (Hawkes process). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

(1) *Genuine multivariate.* Let N be a genuine multivariate point process as in the first part of Definition 1.1. Assume each component N_k, $k \in \{1, \dots, d\}$ has intensity λ_k, as given in Equation (1.1).

(2) *Pseudo multivariate.* Let N be a pseudo multivariate point process as in the second part of Definition 1.1. Assume N has intensity λ as given in Equation (1.2).

In this case, N is called a *Hawkes process.* ♢

As already mentioned, the branching coefficients ϑ_{jk} determine whether the Hawkes model is well-defined or not. The regularity conditions I give below rely on the *spectral radius* of the branching matrix Q:

1.15 Definition (Spectral radius). Let Q be a $d \times d$-matrix and denote the spectrum of Q by $\Lambda(Q)$. The *spectrum* is the set of all (in general complex-valued) eigenvalues of Q. The *spectral radius* is then defined as:

$$\operatorname{Spr}(Q) := \max\big\{|\lambda| : \lambda \in \Lambda(Q)\big\}. \qquad \diamondsuit$$

Since the branching coefficients ϑ_{jk} are explicit model parameters, it is easy to determine whether the model is well-defined or not. The next theorem is a reformulation of Theorem 6.55:

1.16 Theorem (Regularity condition, informal). In order for a Hawkes model to be well-defined, the following conditions need to be satisfied:

(*i*) The branching matrix satisfies:

$$\mathrm{Spr}(Q) < 1.$$

(*ii*) The decay functions satisfy, for all $j \in \{1, \dots, d\}$:

$$\int_0^\infty t w_j(t) dt < \infty.$$

In case of a pseudo multivariate model take w instead of w_j. $\diamondsuit$

Likelihood Function. The standard way to estimate the parameters of a Hawkes process model is to maximize the likelihood function. In order to define the likelihood function, one first has to fix an *observation period* D, i.e. the time interval during which the data has been collected. I will always use the observation period $D := [T_*, T^*]$, for fixed times $T_* < T^*$.

The compensator will be useful for the calculation of the likelihood function. Recall Definition 1.6 of the cumulative decay functions $\bar{w}_j$ and $\bar{w}$:

1.17 Definition (Compensator). For all $t \in D$ define in the genuine and pseudo multivariate case:

$$\Lambda_j(t) := \int_{T_*}^t \lambda_j(s) ds \qquad \text{and} \qquad \Lambda(t) := \int_{T_*}^t \lambda(s) ds.$$

This is the general definition. For a Hawkes process one obtains:

(*1*) *Genuine multivariate.* For $j \in \{1, \dots, d\}$ and $t \in [T_*, T^*]$:

$$\Lambda_j(t) = \sum_{k=1}^d \vartheta_{jk} \int_{(-\infty,t)\times\mathbb{R}} \Big[\bar{w}_j(t-u) - \bar{w}_j(T_*-u)\Big] g_k(x) N_k(du \times dx) + \eta_j(t - T_*),$$

where the convention is used that $\bar{w}_j(t) = 0$, if $t < 0$.

(*2*) *Pseudo multivariate.* For $t \in [T_*, T^*]$:

$$\Lambda(t) = \vartheta \int_{(-\infty,t)\times\mathbb{R}} \Big[\bar{w}(t-u) - \bar{w}(T_*-u)\Big] \Big[\prod_{k=1}^d g_k(x_k)\Big] N(du \times d\boldsymbol{x}) + \eta(t - T_*),$$

where the convention is used that $\bar{w}(t) = 0$, if $t < 0$. $\diamondsuit$

The general definition of the likelihood function of a point process is given in Definition 6.16. The likelihood function for a Hawkes process is derived in Proposition 6.27. It is a function of the model parameters, $\boldsymbol{\theta}$ say, and the n observed events. The vector $\boldsymbol{\theta}$ is the concatenation of the parameters of the decay and impact functions, the branching coefficients and the parameters of the mark distributions. Moreover, the likelihood depends on the choice of the observation period D. In summary, the likelihood is a function of the following form, for the two versions of multivariate Hawkes processes:

$$L_{D,n}\Big(\boldsymbol{\theta}; (t_i, d_i, x_i)_{i=1,\dots,n}\Big) \qquad \text{and} \qquad L_{D,n}\Big(\boldsymbol{\theta}; (t_i, x_{i,1}, \dots, x_{i,d})_{i=1,\dots,n}\Big).$$

To avoid cluttering the notation, I write from now on simply L.

1.18 Proposition (Hawkes likelihood). Let N be a Hawkes process and assume one has observed N in the time interval $D = [T_*, T^*]$.

(1) *Genuine multivariate.* The log-likelihood is given by:

$$\begin{aligned}\log L = &\sum_{j=1}^{d} \int_{[T_*,T^*]\times\mathbb{R}} \log \lambda_j(t) N_j(dt \times dx) \\ &+ \sum_{j=1}^{d} \int_{[T_*,T^*]\times\mathbb{R}} \log f_j(x) N_j(dt \times dx) - \sum_{j=1}^{d} \Lambda_j(T^*).\end{aligned}$$

(2) *Pseudo multivariate.* The log-likelihood is given by:

$$\begin{aligned}\log L = &\int_{[T_*,T^*]\times\mathbb{R}^d} \log \lambda(t) N(dt \times d\boldsymbol{x}) \\ &+ \sum_{k=1}^{d} \int_{[T_*,T^*]\times\mathbb{R}^d} \log f_k(x_k) N(dt \times d\boldsymbol{x}) - \Lambda(T^*).\end{aligned}$$ $\diamondsuit$

Consider the integrals above: Note that the integrands in these integrals either depend on t or on the mark values x or x_k, but not on both. This becomes relevant later, when I discuss numerical algorithms for the calculation of the likelihood function.

Residual Process. Once one has fitted a Hawkes model to some given data, one usually wants to assess how good the model fits the data. A standard

method to check the goodness-of-fit is to consider the so-called residual process. One basically obtains the residual process as follows: Assume the estimated Hawkes model is correct; in other words, assume that the data have been generated by this Hawkes model. Moreover, assume one is able to somehow exactly calculate the intensity process. Then transform the point process in time and space according to the estimated model, as indicated in the next definition. The resulting point process is then the residual process.

Note that it is not possible to calculate the intensity process exactly in practice, since one can observe the Hawkes process only during a finite time interval but the intensity process is a function of the infinite past of the process. For the following construction, use the notation given in Definition 1.1:

1.19 Definition (Residual process). Assume one has observed n events of a Hawkes process in the time interval D. Depending on the type of the process, apply the following transformation:

(*1*) *Genuine multivariate.* The residual process is given by the following time and space transformed point process:

$$(\tau_1, d_1, \chi_1), \quad \dots \quad , (\tau_n, d_n, \chi_n).$$

Note that this is again a d-dimensional, genuine point process. The pair (τ_i, χ_i), for $1 \leq i \leq n$, is obtained as follows:

$$\tau_i := \Lambda_{d_i}(t_i) \qquad \text{and} \qquad \chi_i := F_{d_i}(t_i),$$

where F_j is the cumulative distribution function associated to the mark density f_j, see Definition 1.11.

(*2*) *Pseudo multivariate.* This time, the residual process is of the form:

$$(\tau_1, \chi_{1,1}, \dots, \chi_{1,d}), \quad \dots \quad , (\tau_n, \chi_{n,1}, \dots, \chi_{n,d}).$$

The residual process is again of the same class as the original process, i.e. it is a d-dimensional, pseudo multivariate point process. The tuples $(\tau_i, \chi_{i,1}, \dots, \chi_{i,d})$ are calculated according to:

$$\tau_i := \Lambda(t_i), \ \chi_{i,1} := F_1(x_{i,1}), \quad \dots \quad , \chi_{i,d} := F_d(x_{i,d}). \qquad \diamondsuit$$

The next theorem states that given the model is correct, the residual process is a Poisson process with iid standard uniformly distributed marks. This statement

goes back to [Wat64], [Mey71] and [Pap72]. Modern formulations are given in Theorem T16 of Section II.6 of [Bré81] and in Theorem 7.4.I of [DVJ03].

One could formulate the following theorem for the same observation period $[T_*, T^*]$ that has been used for the likelihood function. But the theorem is easier to formulate if one assumes, at least conceptually, that the observation period is the unbounded interval $[T_*, \infty)$. Loosely speaking, if one takes the interval $[T_*, T^*]$, one observes only a finite number of points. The residual process has then also a finite number of points and can therefore not be a Poisson process on $\mathbb{R}^+$. Indeed, the residual process would be a truncated Poisson process. To avoid this complication, I take instead the infinite time interval $[T_*, \infty)$:

1.20 Theorem (Random time change, informal). Take a Hawkes process with strictly positive immigration intensities, i.e. assume that $\eta_j, \eta > 0$; see Definition 1.4. Moreover, assume that the mark distributions F_j are absolutely continuous.

(1) *Genuine multivariate.* Assume the compensator processes Λ_j are known. Define the residual process in the same way as in the first part of Definition 1.19, but this time assume the observation period is $[T_*, \infty)$. Then define the residual process:

$$(\tau_1, d_1, \chi_1),\ (\tau_2, d_2, \chi_2), \quad \dots \quad .$$

Let R denote be the representation of this sequence as a random measure. Then the projections $R_j(dt \times \mathbb{R})$, for $j \in \{1, \dots, d\}$, are independent standard Poisson processes on $\mathbb{R}_+ \times \mathbb{R}$. Moreover, the marks are independent of $R_j(dt \times \mathbb{R})$ and are iid uniformly distributed on $[0, 1]$.

(2) *Pseudo multivariate.* Assume the compensator process Λ is known and the observation period is $[T_*, \infty)$. Define the time transformed point process R, consisting of the sequence of events:

$$(\tau_1, \chi_{1,1}, \dots, \chi_{1,d}),\ (\tau_2, \chi_{2,1}, \dots, \chi_{2,d}), \quad \dots \quad .$$

Then the projection $R(dt \times \mathbb{R}^d)$ is a standard Poisson process on $\mathbb{R}_+ \times \mathbb{R}^d$ and the marks $\boldsymbol{\chi}_i := (\chi_{i,1}, \dots, \chi_{i,d})$ are independent of $R(dt \times \mathbb{R}^d)$ and iid uniformly distributed on $[0, 1]^d \subseteq \mathbb{R}^d$. ♢

The projections $R_j(dt \times \mathbb{R})$ and $R(dt \times \mathbb{R}^d)$ are sometimes also called the *ground processes*, see Section 6.4 in [DVJ03].

The above theorem can be used to assess the goodness-of-fit of a Hawkes

model. If the Hawkes model is inappropriate the residual process is not a Poisson process and one should be able to detect this. At least, the residual process can be analyzed graphically to detect deviations from the Poisson assumption. Moreover, statistical tests can be used to quantify this deviation, which then leads to a classical goodness-of-fit test statistic.

Simulation. Let me finish this short exposition of Hawkes processes with the discussion of a simulation algorithm. I will only look at the genuine multivariate case. The corresponding algorithm for the pseudo multivariate case is easier, since one basically can use the simulation algorithm for a univariate Hawkes process. The presented algorithm is a multivariate extension of Algorithm 7.5.IV in [DVJ03], where it is called *Ogata's modified thinning algorithm.*

The simulation algorithm consist of two loops: The inner loop determines for every component a candidate event and the outer loop selects among the candidate events the next true event. I explain below the workings of the two loops.

Inner Loop. The inner loops starts off directly after a new event has been generated, say at time t in component j. It determines a sequence $\tau_1, \ldots, \tau_d$ of potentially new events, each τ_k belonging to component k, and returns this sequence to the outer loop. The outer loop then selects exactly one of these d potential events and nominates it as the new event, say at time τ in component r. Indeed, it defines $\tau := \min\{\tau_1, \ldots, \tau_d\}$ and sets r such that $\tau = \tau_r$. The outer loop then discards all other events $\tau_k \neq \tau_r$ and starts the inner loop afresh.

To understand why there are two stages, one has to look more closely at how the inner loop determines a candidate event. Conceptually, the inner loop samples *simultaneously* an event for each component, since it is not yet clear in which component the next event will lie. The true next event is then determined by the rule "the first one wins".

Two-Stage Rejection Sampling. Let me explain this idea from another perspective: Assume the inner loop is sampling a candidate event for component $k \in \{1, \ldots, d\}$. Clearly, whenever an event occurs in one of the other components $l \neq k$, this also affects the intensity process of component k. But unfortunately, at this stage in the algorithm it is not yet known in which component the next event will occur.

Since the inner loop somehow has to make an assumption about the other components, it simply pretends that there are no events in the other components. Notice that this assumption is indeed valid in the time interval (t, τ_r), i.e. until the first event τ_r occurs, but becomes invalid after time τ_r.

Now assume the inner loop generates a new event in component k at time τ_k. If another event has occurred in the mean time, i.e. if $\tau_l < \tau_k$ for $l \neq k$, then the assumption under which τ_k has been generated has become wrong, and in consequence the event τ_k is invalid. But if on the other hand τ_k is the first event, i.e. if $\tau_k < \tau_l$ for all $l \neq k$, the assumption that there are no events in (t, τ_k) is justified and τ_k is a valid event.

It is now the task of the outer loop to sort out the invalid events generated by the inner loop. The outer loop simply compares all the potential events generated by the inner loop and rejects all of them except the first one. Hence, the true next event is $\tau := \min\{\tau_1, \ldots, \tau_d\}$.

Thinning procedure. Let me now explain the inner loop in more detail: Since the inner loop simply assumes that the other components contain no events after time t, one can calculate the intensity process λ_k after time t. Actually one can calculate the so-called *hazard rate*, which I denote by λ_k^+, Simply speaking, the hazard rate is the intensity process under the additional, artificial assumption that there are no events after time t. I write "artificial" since the point process usually has events afterwards, and it is therefore not an assumption in a strict sense. See also Definition 6.38, for a more precise definition of hazard rates.

Now assume one has calculated the hazard rate λ_k^+. It is then quite simple to generate the next event. Basically one needs to generate an inhomogeneous Poisson process with hazard rate λ_k^+, and then take the first event. The inner loop does this by constructing a thinning of a homogeneous Poisson process. But there is one problem one has to address: A straight-forward thinning procedure could become very inefficient, since the intensity could decrease to zero very fast, and hence a large number of trial events would be rejected.

Adaptive Thinning. The algorithm solves this efficiency problem as follows: It assumes the hazard rate is monotonically decreasing after time t. It starts with a homogeneous Poisson process with constant intensity $\lambda := \lambda_k^+(t)$. Then the first event of this homogeneous Poisson is generated, say at time s, and the algorithm checks whether it belongs to the thinned process or not. In case it does, this event becomes the candidate event τ_k. Otherwise, the algorithm

takes a new homogeneous Poisson process, this time with constant intensity $\lambda := \lambda_k^+(s)$. This procedure is repeated until a valid point is found.

Burn-In Period. For simplicity, the algorithm I present below starts at time $t_0 := 0$ and assumes an *empty initial state*, see also Definition 6.34. This means the algorithm starts with no points defined initially; but one could easily change that. If the regularity conditions from Theorem 1.16 are satisfied, then the simulated Hawkes process converges in a strong sense to an associated stationary Hawkes process. The details of this convergence are explained in Proposition 6.52, and especially Theorem 6.55 and its proof.

This implies that the first few events generated by this simulation algorithm should not be used. In other words, one has to let the algorithm run for some time, the so-called *burn-in period*, before one uses the generated events.

1.21 Algorithm (Simulation). Consider a d-dimensional, genuine multivariate Hawkes process. The notation $\lambda_j(t)$ for the intensity process used so far is not suitable for the description of this algorithm, since it might not always be clear which points are used to calculate the intensity. To specify exactly which points are included in the calculation, I use the following notation:

$$\lambda_j(t|n) \equiv \lambda_j\Big(t\Big|(t_1,d_1,x_1),\dots,(t_n,d_n,x_n)\Big).$$

This means that the intensity at time t is calculated assuming the n events given above define the past of the process.

(*I*) *Outer loop initialization.* Define $t_0 := 0$, but this is not yet counted as an event.

(*II*) *Outer loop iteration.* The iteration variable is n and iterates through $n \geq 1$. First start the inner loop:

(*i*) *Inner loop initialization.* Define for $j \in \{1,\dots,d\}$ the values:

$$\tau_{n,j,0} := t_{n-1}.$$

Note that the intensity process $t \mapsto \lambda_j(t)$ is left-continuous with right-hand side limits. Hence, the following right-hand limit is well-defined:

$$\lambda_{n,j,0} := \lambda_j(t_{n-1}-|n-1) \equiv \lim_{t \searrow t_{n-1}} \lambda_j(t|n-1).$$

Note that the event at time t_{n-1} increases the intensity in the open interval (t_{n-1}, ∞). In other words, the intensity is not yet influenced by this event exactly at time t_{n-1}. This explains why one has to take the right-hand limit, since otherwise event number $n-1$ would not be taken into account.

(*ii*) *Inner loop iteration.* The iteration variable is r and iterates through $r \geq 1$. Sample a standard exponential random variable $E_{n,j,l}$ and define:

$$\tau_{n,j,r} := \tau_{n,j,r-1} + E_{n,j,r}/\lambda_{n,j,r-1}, \quad \lambda_{n,j,r} := \lambda_j(\tau_{n,j,r}|n-1).$$

Note that $\lambda_j(\cdot|n-1)$ by definition is the intensity based on the first $n-1$ points sampled so far. Since $\tau_{n,j,r}$ does not correspond to any of these points, one does not need to take the right-hand limit as before. Actually, the intensity is continuous at $\tau_{n,j,r}$, so it does not matter. Now sample a standard uniform random variable $U_{n,j,l}$ and define:

$$u_{n,j,r} := U_{n,j,r}\lambda_{n,j,r-1}.$$

Check the condition $u_{n,j,r} \leq \lambda_{n,j,r}$, and if so stop the iteration.

(*iii*) *Inner loop finalization.* When the iteration has terminated, define:

$$\tau_{n,j} := \tau_{n,j,r}.$$

Then continue in the outer loop.

Back in the outer loop, define the pair (t_n, d_n) as follows:

$$t_n := \min_{j\in\{1,\ldots,d\}} \tau_{n,j}, \qquad d_n := \arg\min_{j\in\{1,\ldots,d\}} \tau_{n,j}.$$

Now sample a random variable X_{n,d_n} with distribution F_{d_n} and define:

$$x_n := X_{n,d_n}.$$

The following triple defines a new event:

Add the triple (t_n, d_n, x_n) to the output of the algorithm.

Check some terminal condition, e.g. check whether the number n of events

generated so far is sufficient. Otherwise continue with the iteration in the outer iteration, now for $n+1$ instead of n. ♢

Note that an actual implementation has to consider some more details: E.g. it is necessary to make sure that none of the two loops repeats infinitely, which could happen if the Hawkes model is not well-defined.

1.3 Examples of Hawkes Processes

So far I gave only the definition and described the properties and interpretation of the decay functions, impact functions and mark distributions. It is now time to give some concrete examples. Clearly, the following list cannot be complete. Instead it mentions some of the more obvious functions one could take. One can obtain already quite many Hawkes models if one considers all possible combinations of decay functions, mark distributions and associated impact functions.

Decay Functions. For simplicity, I consider only the genuine multivariate case. In the pseudo multivariate case, omit the index j in the following definitions. Let $j \in \{1, \dots, d\}$ be the component index. Note that both decay functions given below satisfy the normalizing condition from Convention 1.12.

Exponential Decay Function. Let $\alpha_j > 0$ be parameters and define for $t \geq 0$ the exponential decay function:

$$w_j(t) = \alpha_j \exp\{-\alpha_j t\}.$$

Associated to the decay functions are the cumulative decay functions $\bar{w}_j$ and the quantile functions q_j, see Definition 1.6. For the exponential decay function they are given by:

$$\bar{w}_j(t) = 1 - \exp\{-\alpha_j t\} \qquad \text{and} \qquad q_j = -\frac{\ln(\varepsilon)}{\alpha_j},$$

where $\varepsilon > 0$ is some fixed value.

1.22 Remark. The exponential decay function has an important property: A Hawkes process with exponential decay functions is actually a Markov process if one includes the intensity as part of the state of the process. As a consequence,

there exists a more efficient algorithm for the calculation of the intensity process. This is especially important for the numerical calculation of the intensity process, which is discussed in the next section. Hence, if one has no specific preferences, one should preferably take an exponential decay function. ♢

Power Decay Function. Let $\alpha_j > 2$ and $\beta_j > 0$ be parameters and define for $t \geq 0$ the power decay function:

$$w_j(t) = \frac{(\alpha_j - 1)\beta_j}{(1 + \beta_j t)^{\alpha_j}}.$$

Note that the restriction $\alpha > 2$ is required such that the second condition of Theorem 1.16 is satisfied, i.e. such that $\int_0^\infty t w_j(t) dt < \infty$ holds. The associated cumulative decay functions $\bar{w}_j$ and the quantile functions q_j, for a constant $\varepsilon > 0$, are given by:

$$\bar{w}_j(t) = 1 - (1 + \beta_j t)^{1-\alpha_j} \quad \text{and} \quad q_j = \frac{1}{\beta_j}\Big[\exp\big\{-(\alpha_j - 1)^{-1} \ln \varepsilon\big\} - 1\Big].$$

As a historical side remark, the ETAS model for earthquake modeling uses a power decay function. See the original definition of the ETAS model given in Section 2.1 of the paper [Oga88]. In the context of earthquake modeling, the decay function is called *Omori's Law* and is defined as:

$$\tilde{w}(t) := \frac{K}{(t + c)^p}.$$

where K, c and p are some parameters. Note that Omori's Law does not satisfy the normalizing condition from Convention 1.12. It is therefore not a decay function in the sense used so far; and this explains the notation $\tilde{w}$.

One-Sided Mark Distributions. Below follows a list of common distributions which may be useful in a practical application. Since the impact functions differ for one-sided and two-sided distributions, I split the distributions in two groups and start with the one-sided ones.

Recall from Convention 1.12 that the impact functions need to be adapted to the corresponding distribution. Since the integral in the normalizing condition has to be finite, one cannot take any pair of distribution and impact function. I will focus only on a few standard types of impact functions, although one could think also of more complicated ones:

1.23 Definition (One-sided impact functions). Let $\alpha, \beta, \gamma \geq 0$.

(*1*) *Non-normalized.* Consider the following three functions:

$$\tilde{g}(x) = \alpha + \beta x + \gamma x^2, \qquad \tilde{g}(x) = x^\alpha, \qquad \tilde{g}(x) = \exp\{\alpha x\}.$$

In the first case assume that at least one of α, β, γ is strictly positive; and in the other two cases assume $\alpha > 0$. Note that these functions are not impact functions, since they do not satisfy the normalizing condition given in Convention 1.12. Hence, one could call them *non-normalized* impact functions and this is the reason why the notation $\tilde{g}$ has been used.

(*2*) *Normalized.* Now assume X has some distribution f, see also Definition 1.11. Then define the impact functions:

$$g(x) = \frac{\alpha + \beta x + \gamma x^2}{\alpha + \beta\,\mathbb{E}[X] + \gamma\,\mathbb{E}[X^2]}, \quad g(x) = \frac{x^\alpha}{\mathbb{E}[X^\alpha]}, \quad g(x) = \frac{\exp\{\alpha x\}}{\mathbb{E}\big[\exp\{\alpha X\}\big]}.$$

These impact functions are now the *normalized* versions. Clearly, whether a certain impact function is compatible with a given distribution f depends on whether the associated moments are finite or not.

(*3*) *Degenerate impact function.* As a degenerate case one can always take the void impact function $g(x) = 1$, which simply ignores the marks, see also Remark 1.10. ♢

Note that the normalized impact functions depend not only on the three parameters α, β, γ, but implicitly also on the parameters of the distribution of X. This becomes clear if one looks at the examples given below.

Special Functions. Some of the distributions and impact functions that follow below contain special mathematical functions. For completeness, I list them with their definitions:

1.24 Definition (Special functions). The definitions for the following special functions can be found e.g. in the book [AS64].

(*i*) The *error function* is defined for $x \geq 0$ by:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

It is related as follows to the cumulative distribution function Φ of a standard normal distribution:

$$\Phi(x) = \frac{1}{2}\Big(1 + \mathrm{erf}\Big\{\frac{x}{\sqrt{2}}\Big\}\Big) \qquad \text{and} \qquad \mathrm{erf}(x) = 2\Phi\big(x\sqrt{2}\big) - 1.$$

See Equations 7.1.1 and 7.1.22 of [AS64].

(*ii*) The *gamma function* and *lower incomplete gamma function* are defined for $\kappa > 0$ and $x \geq 0$ by:

$$\Gamma(\kappa) = \int_0^\infty t^{\kappa-1}e^{-t}dt, \qquad \gamma(\kappa, x) = \int_0^x t^{\kappa-1}e^{-t}dt.$$

See Equations 6.1.1 and 6.5.2 in [AS64].

(*iii*) *Gauss' hypergeometric function* is defined as:

$${}_2\mathrm{F}_1(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a}dt.$$

See Equations 15.1.1 and 15.3.1 in [AS64]. For the exact domain where this function is defined, see the same reference. ♢

For all impact functions that follow below, assume the three parameters α, β, γ lie in the appropriate range, as specified in Definition 1.23.

Exponential Distribution. The exponential distribution is defined on the half line $(0, \infty)$ and has one parameter $\lambda > 0$. The density and cumulative distribution function are:

$$f_\lambda(x) = \lambda \exp\{-\lambda x\} \qquad \text{and} \qquad F_\lambda(x) = 1 - \exp\{-\lambda x\}.$$

The first two and higher order moments are:

$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \mathbb{E}[X^2] = \frac{2}{\lambda^2}, \qquad \mathbb{E}[X^\alpha] = \frac{\Gamma(\alpha+1)}{\lambda^\alpha}.$$

Two suitable impact functions are:

$$g(x) = \frac{\lambda^2}{\alpha\lambda^2 + \beta\lambda + 2\gamma}\big(\alpha + \beta x + \gamma x^2\big), \qquad g(x) = \frac{\lambda^\alpha}{\Gamma(\alpha+1)}x^\alpha.$$

Gamma Distribution. The gamma distribution is defined on the half line $(0, \infty)$ and has two parameters $\eta > 0$ and $\sigma > 0$. The density and cumulative distribution function are:

$$f_{\eta,\sigma}(x) = x^{\eta-1} \frac{\exp\{-x/\sigma\}}{\Gamma(\eta)\sigma^\eta} \qquad \text{and} \qquad F_{\eta,\sigma}(x) = \frac{\gamma(\eta, x/\sigma)}{\Gamma(\eta)}.$$

The first two and higher order moments are:

$$\mathbb{E}[X] = \eta\sigma, \qquad \mathbb{E}[X^2] = \eta(\eta+1)\sigma^2, \qquad \mathbb{E}[X^\alpha] = \frac{\sigma^\alpha \Gamma(\eta+\alpha)}{\Gamma(\eta)}.$$

Two suitable impact functions are:

$$g(x) = \frac{\alpha + \beta x + \gamma x^2}{\alpha + \beta\eta\sigma + \gamma\eta(\eta+1)\sigma^2}, \qquad g(x) = \frac{\Gamma(\eta)}{\Gamma(\eta+\alpha)\sigma^\alpha} x^\alpha.$$

Inverse Gamma Distribution. The inverse gamma distribution is defined on the half line $(0, \infty)$ and has two parameters $\eta > 0$ and $\sigma > 0$. The density and cumulative distribution function are:

$$f_{\eta,\sigma}(x) = \frac{\sigma^\eta}{\Gamma(\eta)} x^{-\eta-1} \exp\Big\{\frac{-\sigma}{x}\Big\} \qquad \text{and} \qquad F_{\eta,\sigma}(x) = \frac{\gamma(\eta, \sigma/x)}{\Gamma(\eta)},$$

where γ is the lower incomplete gamma function, see Definition 1.24. A finite first moment requires $\eta > 1$ and a finite second moment requires $\eta > 2$. In this case, the first two moments are:

$$\mathbb{E}[X] = \frac{\sigma}{\eta - 1} \qquad \text{and} \qquad \mathbb{E}[X^2] = \frac{\sigma^2}{(\eta-1)(\eta-2)}.$$

Under the restriction that $\eta > 2$, a suitable impact function is:

$$g(x) = \frac{(\eta-1)(\eta-2)}{\alpha(2 - 3\eta + \eta^2) + \beta(-2\sigma + \eta\sigma) + \gamma\sigma^2}(\alpha + \beta x + \gamma x^2).$$

Log-normal Distribution. The log-normal distribution is defined on the half line $(0, \infty)$ and has two parameters $\mu \in \mathbb{R}$ and $\sigma > 0$. The density and cumulative distribution function are:

$$f_{\mu,\sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\Big\{\frac{-(\ln x - \mu)^2}{2\sigma^2}\Big\}, \quad F_{\mu,\sigma}(x) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\Big\{\frac{\ln x - \mu}{\sigma\sqrt{2}}\Big\},$$

where erf is the error function, see Definition 1.24. The first two and higher order moments are:

$$\mathbb{E}[X] = \exp\Big\{\mu + \frac{\sigma^2}{2}\Big\}, \qquad \mathbb{E}[X^2] = \exp\big\{2\mu + 2\sigma^2\big\},$$
$$\mathbb{E}[X^\alpha] = \exp\Big\{\alpha\mu + \frac{\alpha^2\sigma^2}{2}\Big\}.$$

Two suitable impact functions are:

$$g(x) = \frac{\alpha + \beta x + \gamma x^2}{\alpha + \beta\exp\{\mu + \sigma^2/2\} + \gamma\exp\{2\mu + 2\sigma^2\}},$$
$$g(x) = \exp\Big\{-\frac{2\alpha\mu + \alpha^2\sigma^2}{2}\Big\}x^\alpha.$$

Pareto Distribution. The Pareto distribution is defined on the half line $[0, \infty)$ and has two parameters $\mu > 0$ and $\rho > 0$. The density and cumulative distribution function are:

$$f_{\mu,\rho}(x) = \frac{\rho\mu^\rho}{(x+\mu)^{\rho+1}} \qquad \text{and} \qquad F_{\mu,\rho}(x) = 1 - \Big(\frac{\mu}{x+\mu}\Big)^\rho.$$

A finite first moment requires $\rho > 1$ and finite second moment requires $\rho > 2$. In this case, the first two moments are:

$$\mathbb{E}[X] = \frac{\mu}{(\rho-1)} \qquad \text{and} \qquad \mathbb{E}[X^2] = \frac{2\mu^2}{(\rho-1)(\rho-2)}.$$

Under the restriction that $\rho > 2$, a suitable impact function is:

$$g(x) = \frac{(\rho-1)(\rho-2)}{\alpha(\rho-1)(\rho-2) + \beta\mu(\rho-2) + 2\gamma\mu^2}(\alpha + \beta x + \gamma x^2).$$

Rayleigh Distribution. The Rayleigh distribution is defined on the half line $(0, \infty)$ and has one parameter $\sigma > 0$. The density and cumulative distribution function are:

$$f_\sigma(x) = \frac{x}{\sigma^2}\exp\Big\{-\frac{x^2}{2\sigma^2}\Big\} \qquad \text{and} \qquad F_\sigma(x) = 1 - \exp\Big\{\frac{-x^2}{2\sigma^2}\Big\}.$$

The first two and higher order moments are:

$$\mathbb{E}[X] = \sigma\sqrt{\pi/2}, \qquad \mathbb{E}[X^2] = 2\sigma^2, \qquad \mathbb{E}[X^\alpha] = \sigma^\alpha 2^{\alpha/2}\Gamma(1+\alpha/2).$$

The moment generating function is:

$$\mathbb{E}\big[\exp\{\alpha X\}\big] = 1 + \sigma\alpha\sqrt{\pi/2}\exp\Big\{\frac{\sigma^2\alpha^2}{2}\Big\}\Big[\mathrm{erf}\Big(\frac{\sigma\alpha}{\sqrt{2}}\Big) + 1\Big].$$

Three suitable impact functions are:

$$g(x) = \frac{\alpha + \beta x + \gamma x^2}{\alpha + \beta\sigma\sqrt{\pi/2} + 2\gamma\sigma^2}, \qquad g(x) = \frac{x^\alpha}{\sigma^\alpha 2^{\alpha/2}\Gamma(1+\alpha/2)},$$
$$g(x) = \frac{\exp\{\alpha x\}}{1 + \sigma\alpha\sqrt{\pi/2}\exp\{\sigma^2\alpha^2/2\}\big[\mathrm{erf}\big(\sigma\alpha/\sqrt{2}\big) + 1\big]}.$$

Weibull Distribution. The Weibull distribution is defined on the half line $(0,\infty)$ and has two parameters $\kappa > 0$ and $\sigma > 0$. The density and cumulative distribution function are:

$$f_{\kappa,\sigma}(x) = \frac{\kappa}{\sigma}\Big(\frac{x}{\sigma}\Big)^{\kappa-1}\exp\Big\{-\Big(\frac{x}{\sigma}\Big)^\kappa\Big\}, \qquad F_{\kappa,\sigma}(x) = 1 - e^{-(x/\sigma)^\kappa}.$$

The first two and higher order moments are:

$$\mathbb{E}[X] = \sigma\Gamma\Big(1+\frac{1}{\kappa}\Big), \quad \mathbb{E}[X^2] = \sigma^2\Gamma\Big(1+\frac{2}{\kappa}\Big), \quad \mathbb{E}[X^\alpha] = \sigma^\alpha\Gamma\Big(1+\frac{\alpha}{\kappa}\Big).$$

Two suitable impact functions are:

$$g(x) = \frac{\alpha + \beta x + \gamma x^2}{\alpha + \beta\sigma\Gamma\big(1+\frac{1}{\kappa}\big) + \gamma\sigma^2\Gamma\big(1+\frac{2}{\kappa}\big)}, \qquad g(x) = \frac{x^\alpha}{\sigma^\alpha\Gamma\big(1+\frac{\alpha}{\kappa}\big)}.$$

Two-sided Mark Distributions. Two-sided distributions are defined on the real line $\mathbb{R}$ instead on the half-line $(0,\infty)$. This time, I consider only the polynomial and exponential impact function, since for the distributions given below a power impact function would lead to intractable formulas. Moreover, the polynomial and exponential impact function are slightly altered to account for negative mark values:

1.25 Definition (Two-sided impact functions). Let $\alpha, \beta, \gamma \geq 0$.

(1) *Non-normalized.* Consider the following three functions:

$$\tilde{g}(x) = \alpha + \beta|x| + \gamma x^2, \qquad \tilde{g}(x) = \exp\{\alpha|x|\}.$$

Since these functions do not necessarily satisfy the normalizing condi-

tion given in Convention 1.12, one could call them again non-normalized impact functions.

(2) *Normalized.* Now assume X has distribution f and define the normalized impact functions by:

$$g(x) = \frac{\alpha + \beta|x| + \gamma x^2}{\alpha + \beta\,\mathbb{E}[|X|] + \gamma\,\mathbb{E}[X^2]}, \qquad g(x) = \frac{\exp\{\alpha|x|\}}{\mathbb{E}\big[\exp\{\alpha|X|\}\big]}. \qquad \diamondsuit$$

The same remarks apply as in the one-sided case: Note that the impact functions do not only depend on the explicit parameters α, β, γ, but also implicitly on the parameters of the associated distribution. Also recall that one can always take the void impact function $g(x) = 1$, which simply ignores the marks.

Centered Normal Distribution. The normal distribution is defined on the whole real line $\mathbb{R}$ and has one parameter $\sigma > 0$. The density and cumulative distribution function are:

$$f_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big\{-\frac{x^2}{2\sigma^2}\Big\}, \qquad F_\sigma(x) = \frac{1}{2}\Big(1 + \operatorname{erf}\Big\{\frac{x}{\sigma\sqrt{2}}\Big\}\Big),$$

where erf is the error function. The first two moments of $|X|$ are:

$$\mathbb{E}[|X|] = \sigma\sqrt{2\pi^{-1}} \qquad \text{and} \qquad \mathbb{E}[X^2] = \sigma^2.$$

The moment generating function of $|X|$ is:

$$\mathbb{E}\big[\exp\{\alpha|X|\}\big] = \exp\big\{\tfrac{1}{2}\alpha^2\sigma^2\big\}\Big(1 + \operatorname{erf}\Big\{\frac{\alpha\sigma}{\sqrt{2}}\Big\}\Big).$$

Two suitable impact functions are:

$$g(x) = \frac{\alpha + \beta|x| + \gamma x^2}{\alpha + \beta\sigma\sqrt{2\pi^{-1}} + \gamma\sigma^2}, \qquad g(x) = \frac{\exp\big\{-\alpha^2\sigma^2/2 + \alpha|x|\big\}}{1 + \operatorname{erf}\big(\alpha\sigma/\sqrt{2}\big)}.$$

Student t Distribution. The Student t distribution is defined on the real line $\mathbb{R}$ has one parameter $\nu > 0$. The density and cumulative distribution are:

$$f_\nu(x) = \frac{\Gamma\big(\frac{1}{2}(\nu+1)\big)}{\Gamma\big(\frac{1}{2}\nu\big)\sqrt{\nu\pi}}\big(1 + x^2/\nu\big)^{-(\nu+1)/2},$$
$$F_\nu(x) = \frac{1}{2} + x\Gamma\Big(\frac{\nu+1}{2}\Big)\frac{{}_2\mathrm{F}_1\big(\frac{1}{2}, \frac{\nu+1}{2}, \frac{3}{2}, -\frac{x^2}{\nu}\big)}{\sqrt{\pi\nu}\Gamma\big(\frac{\nu}{2}\big)},$$

where ${}_2\mathrm{F}_1$ is Gauss' hypergeometric function, see Definition 1.24. A finite first moment requires $\nu > 1$ and a finite second moment requires $\nu > 2$. In this case, the first two moments of $|X|$ are:

$$\mathbb{E}[|X|] = \sqrt{\nu/\pi}\frac{\Gamma\big(\frac{1}{2}(\nu-1)\big)}{\Gamma\big(\frac{1}{2}\nu\big)} \qquad \text{and} \qquad \mathbb{E}[X^2] = \frac{\nu}{\nu-2}.$$

Under the restriction that $\nu > 2$, a suitable impact function is:

$$g(x) = \frac{\sqrt{\pi}\nu\Gamma\big(\frac{1}{2}\nu\big) + \sqrt{\nu}(\nu-2)\Gamma\big(\frac{1}{2}(\nu-1)\big)}{\sqrt{\pi}(\nu-2)\Gamma\big(\frac{1}{2}\nu\big)}\big(\alpha + \beta|x| + \gamma x^2\big).$$

1.4 Numerical Algorithms

This section discusses numerical algorithms for the maximum likelihood parameter estimation and the simulation of Hawkes processes. Since these algorithms are meant to be used in real world applications, care has been taken to choose efficient algorithms.

Often it is numerically more efficient to calculate only an approximation for some expression. The algorithms below calculate only approximations to the intensity and compensator processes. But the deviation from the exact value can be chosen to be as small as possible, so that one still obtains accurate results. Moreover, recall that it is not even theoretically possible to calculate all processes exactly. E.g. the intensity process is defined as a function of the infinite past, but one is obviously only able to observe a finite time period. See also Remark 6.18, which explains the difference between the realized and the observed point process.

In consequence, there are usually two formulas for some expression: An exact one and an approximation. To remain mathematically correct, I make a clear notational distinction between these two:

1.26 NOTATION (THEORETICAL VS. EMPIRICAL VALUES). Let z be some observed quantity. The following notational distinction is made:

z: observed value, $\hat{z}$: estimated or approximated value. $\diamondsuit$

The above notation agrees with the usual notation for an estimated value used in statistics. Recall that the observation period is $D := [T_*, T^*]$, and the observed point process is given by the sequence of events $(t_1, \ldots, t_n)$, see Definition 1.1. By assumption, the time points t_i are strictly increasing. I will always represent the observed point process using this notation.

Likelihood Function. The likelihood function for the two cases of Hawkes processes are given in Proposition 1.18. This subsection explain how the different parts of the likelihood function can be calculated efficiently.

1.27 Algorithm (Hawkes likelihood). Assume one has observed a Hawkes process on D and the points are represented as in Definition 1.1.

(1) *Genuine multivariate.* Define the following estimator:

$$\log \hat{L} = \sum_{m=1}^{n} \log \hat{\lambda}_{d_m}(t_m) + \sum_{m=1}^{n} \log f_{d_m}(x_m) - \sum_{j=1}^{d} \hat{\Lambda}_j(T^*).$$

(2) *Pseudo multivariate.* Define the following estimator:

$$\log \hat{L} = \sum_{m=1}^{n} \log \hat{\lambda}(t_m) + \sum_{m=1}^{n} \sum_{k=1}^{d} \log f_k(x_m) - \hat{\Lambda}(T^*). \qquad \diamondsuit$$

Intensity Process. Without any changes, the algorithm for the calculation of the intensity process would have quadratic complexity, i.e. the computational time would increase at the order of n^2, where n is the number of observed events. To avoid this, the sum in the following algorithm is truncated and only the first few terms are calculated. The parameter ε is used to adjust the precision of this approximation.

1.28 Algorithm (Intensity process). The following algorithms can be used to calculate the intensity processes approximatively:

(1) *Genuine multivariate.* According to Definition 1.6, one can assume that

$w_j(\Delta t) \approx 0$, for $\Delta t > q_j$. This leads to the approximation:

$$\hat{\lambda}_j(t) = \eta_j + \sum_{k=1}^{d} \vartheta_{jk} \int_{[t-q_j,t)\times\mathbb{R}} w_j(t-s) g_k(x) N_k(ds \times dx).$$

For the calculation of the likelihood function, one needs to know the following values, see Algorithm 1.27:

$$\hat{\lambda}_{d_1}(t_1), \ldots, \hat{\lambda}_{d_n}(t_n).$$

Given the observed Hawkes process, one therefore has to calculate for $1 \le i \le n$ and $j := d_i$ the values:

$$\hat{\lambda}_j(t_i) = \eta_j + \sum_{m=1}^{i-1} \mathbb{1}_{\{t_i - t_m \le q_j\}} \vartheta_{j,d_m} w_j(t_i - t_m) g_{d_m}(x_m).$$

One remark concerning the indicator $\mathbb{1}_{\{t_i - t_m \le q_j\}}$: An implementation of this function does not really calculate this indicator function. Instead, start the summation at $m := i-1$ and continue with the summation backwards as long as $(t_i - t_m) \le q_j$ is true. As soon as this condition becomes false, stop the summation.

(2) *Pseudo multivariate.* According to Algorithm 1.27, for the calculation of the likelihood function one needs to know the following values:

$$\hat{\lambda}(t_1), \ldots, \hat{\lambda}(t_d).$$

With the same reasoning as in the genuine case, these values can be approximated by:

$$\hat{\lambda}(t_i) = \eta + \vartheta \sum_{m=1}^{i-1} \mathbb{1}_{\{t_i - t_m \le q\}} w(t_i - t_m) \Big[\prod_{k=1}^{d} g_k(x_{k,m}) \Big]. \qquad \diamondsuit$$

For exponential decay functions, the intensity processes can be calculated more efficiently using a recursive scheme:

1.29 Algorithm (Exponential decay function). If the decay functions are exponential, use the following algorithm instead of Algorithm 1.28:

(1) *Genuine multivariate.* Assume the decay functions are exponential func-

tions of the form:

$$w_j(t) := \alpha_j \exp\{-\alpha_j t\}.$$

Assume the values $\lambda_j(r)$ are known, for some $r \in \mathbb{R}$. Then one has for all $t > r$ that:

$$\begin{aligned}
\lambda_j(t) &= \eta_j + \sum_{k=1}^{d} \vartheta_{jk} \int_{(-\infty,t)\times\mathbb{R}} \alpha_j e^{-\alpha_j(t-s)} g_k(x) N_k(ds \times dx) \\
&= \eta_j + e^{-\alpha_j(t-r)} \sum_{k=1}^{d} \vartheta_{jk} \int_{(-\infty,r)\times\mathbb{R}} \alpha_j e^{-\alpha_j(r-s)} g_k(x) N_k(ds \times dx) \\
&\quad + \sum_{k=1}^{d} \vartheta_{jk} \int_{[r,t)\times\mathbb{R}} \alpha_j e^{-\alpha_j(t-s)} g_k(x) N_k(ds \times dx) \\
&= \eta_j + e^{-\alpha_j(t-r)} \big[\lambda_j(r) - \eta_j\big] \\
&\quad + \sum_{k=1}^{d} \vartheta_{jk} \int_{[r,t)\times\mathbb{R}} \alpha_j e^{-\alpha_j(t-s)} g_k(x) N_k(ds \times dx).
\end{aligned}$$

Assume now one has observed a Hawkes process on D. The initial values $\lambda_j(t_1)$ are not known and need to be estimated. To this end, define for $j \in \{1, \ldots, d\}$ the values:

$$\hat{\lambda}_j(t_1) = \eta_j.$$

Then calculate for all $2 \leq i \leq n$ and $1 \leq j \leq d$ the values:

$$\begin{aligned}
\hat{\lambda}_j(t_i) = \eta_j &+ e^{-\alpha_j(t_i - t_{i-1})} \big[\lambda_j(t_{i-i}) - \eta_j\big] \\
&+ \vartheta_{j,d_{i-1}} \alpha_j e^{-\alpha_j(t_i - t_{i-1})} g_{d_{i-1}}(x_{i-1}).
\end{aligned}$$

Note that in order to calculate the likelihood function, one only needs the values $\hat{\lambda}_{d_i}(t_i)$, see Algorithm 1.27. This algorithm however has to calculate all values $\hat{\lambda}_j(t_i)$, due to the recursive nature of the procedure.

(2) *Pseudo multivariate.* Assume the decay function is given by:

$$w(t) := \alpha \exp\{-\alpha t\}.$$

Now assume the value $\lambda(r)$ is known for some $r \in \mathbb{R}$. Then one has for

all $t > r$ that:

$$\lambda(t) = \eta + e^{-\alpha(t-r)}\big[\lambda(r) - \eta\big] + \vartheta \int_{[r,t)\times\mathbb{R}} \alpha e^{-\alpha(t-s)} \Big[\prod_{k=1}^{d} g_k(x_k)\Big] N(ds \times d\boldsymbol{x}).$$

This leads to the following recursive procedure. The first value of the intensity process is estimated by:

$$\hat{\lambda}(t_1) = \eta.$$

Then calculate for all $2 \leq i \leq n$ the values:

$$\hat{\lambda}(t_i) = \eta + e^{-\alpha(t_i - t_{i-1})}\big[\hat{\lambda}(t_{i-i}) - \eta\big] + \vartheta\alpha e^{-\alpha(t_i - t_{i-1})} \prod_{k=1}^{d} g_k(x_{k,i-1}). \quad \diamondsuit$$

For the calculation of the likelihood function one additionally needs to know the compensator values, see Algorithm 1.27. The next algorithm shows how the values $\Lambda_j(T^*)$, $\Lambda(T^*)$ can be calculated efficiently:

1.30 Algorithm (Compensator for likelihood calculation). The following algorithms calculate the compensator values at the terminal time T^* approximatively:

(*1*) *Genuine multivariate.* The value that needs to be calculated is:

$$\Lambda_j(T^*) = \eta_j(T^* - T_*) + \sum_{k=1}^{d} \vartheta_{jk} \int_{(-\infty, T^*)\times\mathbb{R}} \bar{w}_j(T^* - s) g_k(x) N_k(ds \times dx).$$

Now use the approximation $\bar{w}_j(\Delta t) \approx 1$, for $\Delta t > q_j$. Substituting this above leads to:

$$\hat{\Lambda}_j(T^*) = \eta_j(T^* - T_*) + \sum_{k=1}^{d} \vartheta_{jk} \int_{(-\infty, T^* - q_j)\times\mathbb{R}} g_k(x) N_k(ds \times dx) + \sum_{k=1}^{d} \vartheta_{jk} \int_{[T^* - q_j, T^*)\times\mathbb{R}} \bar{w}_j(T^* - s) g_k(x) N_k(ds \times dx).$$

Now assume one has a observed a Hawkes process on D, using the no-

tation from Definition 1.1. The above expression can now be written as:

$$\hat{\Lambda}_j(T^*) = \eta_j(T^* - T_*) + \sum_{m=1}^{n} \mathbb{1}_{\{T^* - t_m \le q_j\}} \vartheta_{j,d_m} \bar{w}_j(T^* - t_m) g_{d_m}(x_m) + \sum_{m=1}^{n} \mathbb{1}_{\{T^* - t_m > q_j\}} \vartheta_{j,d_m} g_{d_m}(x_m).$$

As before, an implementation does not actually need to calculate the two indicator functions. Instead, it starts the summation at $m := n$ and counts backwards until $(T^* - t_m) > q_j$ becomes true. At that point, it assumes $\bar{w}_j(T^* - t_m) = 1$ and continues with the summation.

(2) *Pseudo multivariate.* With the same reasoning as in the genuine multivariate case, one obtains:

$$\hat{\Lambda}(T^*) = \eta(T^* - T_*) + \vartheta \sum_{m=1}^{n} \mathbb{1}_{(q_j,\infty)}(T^* - t_m) \bar{w}(T^* - t_m) \Big[\prod_{k=1}^{d} g_k(x_{k,m}) \Big] + \vartheta \sum_{m=1}^{n} \mathbb{1}_{(0,q_j]}(T^* - t_m) \Big[\prod_{k=1}^{d} g_k(x_{k,m}) \Big].$$

Concerning the indicator functions, the same remark as above applies. ♢

For the calculation of the residual process, one basically needs to calculate the compensator values, see Definition 1.19. One could now simply take the above algorithm and calculate the compensator values at the times t_i instead at the terminal time T^*. But this would be rather inefficient. The following algorithm is more suitable if one wants to calculate all compensator values at all time points and not only at the terminal time..

For the following algorithm it is more convenient if one only has one quantile to deal with instead of d different quantiles, see Definition 1.6. To this end, define in the genuine multivariate case $q := \max\{q_1, \ldots, q_d\}$.

1.31 Algorithm (Residual process). Assume one has observed a Hawkes process on D.

(1) *Genuine multivariate.* Using the same type of approximation as in the

first part of Algorithm 1.30, one obtains:

$$\Lambda_j(t_i) = \eta_j(t_i - T_*) + \sum_{m=1}^{i-1} \mathbb{1}_{\{t_i - t_m \le q\}} \vartheta_{j,d_m} \bar{w}_j(t_i - t_m) g_{d_m}(x_m) + \sum_{m=1}^{i-1} \mathbb{1}_{\{t_i - t_m > q\}} \vartheta_{j,d_m} g_{d_m}(x_m).$$

This formula has to be evaluated for all $1 \le i \le n$. As one can see, the product in the second sum remains the same for all m. To avoid that this term is calculated repeatedly, define a sequence of intermediate values:

$$\Theta_1(t_0), \ldots, \Theta_d(t_0), \ \ldots \ , \Theta_1(t_1), \ldots, \Theta_d(t_1), \ \ldots \ , \Theta_1(t_n), \ldots, \Theta_d(t_n).$$

This sequence is calculated recursively, where one starts with:

$$\Theta_1(t_0) = 0, \quad \ldots \quad , \Theta_d(t_0) = 0.$$

Then calculate for all $1 \le i \le n$ and all $1 \le j \le d$ the values:

$$\Theta_j(t_i) = \Theta_j(t_{i-1}) + \vartheta_{j,d_i} g_{d_i}(x_i).$$

The values $\Lambda_j(t_i)$ can now be calculated as follows: Fix some t_i and take m_0 such that $(t_i - t_{m_0}) \le q < (t_i - t_{m_0 - 1})$. Since the sequence $t_{i=1,\ldots,n}$ is ordered and strictly increasing, one can always find such an index m_0. Then calculate:

$$\Lambda_j(t_i) = \eta_j(t_i - T_*) + \sum_{m=m_0}^{i-1} \vartheta_{j,d_m} \bar{w}_j(t_i - t_m) g_{d_m}(x_m) + \Theta_j(t_{m_0 - 1}).$$

(2) *Pseudo multivariate.* Using the same approximation as in the genuine multivariate case, one obtains:

$$\Lambda(t_i) = \eta(t_i - T_*) + \vartheta \sum_{m=1}^{i-1} \mathbb{1}_{\{t_i - t_m \le q\}} \bar{w}(t_i - t_m) \Big[\prod_{k=1}^{d} g_k(x_{k,m}) \Big] + \vartheta \sum_{m=1}^{i-1} \mathbb{1}_{\{t_i - t_m > q\}} \Big[\prod_{k=1}^{d} g_k(x_{k,m}) \Big].$$

Again calculate first a sequence of intermediate values:

$$\Theta(t_0),\ \Theta(t_1),\quad \ldots\quad ,\Theta(t_n).$$

This sequence is calculated recursively, where one starts with:

$$\Theta(t_0) = 0.$$

The remaining values are calculated for $1 \leq i \leq n$ according to:

$$\Theta(t_i) = \Theta(t_{i-1}) + \vartheta \prod_{k=1}^{d} g_k(x_{k,i}).$$

The values $\Lambda_j(t_i)$ can now be calculated as follows: Fix some t_i and choose m_0 such that $(t_i - t_{m_0}) \leq q < (t_i - t_{m_0-1})$, as before. Then calculate:

$$\Lambda(t_i) = \eta(t_i - T_*) + \vartheta \sum_{m=m_0}^{i-1} \bar{w}(t_i - t_m)\Big[\prod_{k=1}^{d} g_k(x_{k,m})\Big] + \Theta(t_{m_0-1}). \quad \diamondsuit$$

1.5 An Illustrative Case Study

In this section we apply a Hawkes model to financial data and discuss some of the practical issues. The following exposition is not an exhaustive analysis but serves more as an illustration of the theory developed. In an actual study, one has to take more care about how the data are prepared and what kind of Hawkes process is selected.

The Data. We consider daily closing values from the Dow Jones Industrial Average from 1985 to 2005. There is an important point we need to discuss first: This data set is a *time series*, as most other financial data too, and not a point process. The "events" occur at a deterministic rate of exactly one per day, except for weekends and holidays. It would therefore almost certainly lead to nonsensical results if we fitted a Hawkes model directly to this time series. Hence, the raw data is not suitable for further analysis.

We are going to look at extreme values only, and one might motivate this as follows: The market reacts to a variety of external events, which in turn show up in the Dow Jones Index as smaller or bigger value changes. One can now make the simplifying assumption that the majority of small price changes is some sort of noise, whereas only a minority of bigger price changes is caused by relevant external events. It is then plausible that these bigger price changes contain most of the relevant information for future price changes. This motivates to look at extreme values only.

Extracting Extreme Values. For simplicity, we consider values that are above or below two thresholds. In other words, we remove all values in between these two thresholds from the data set. If we then subtract the corresponding thresholds from these values, we obtain the exceedances. Note that this procedure introduces "gaps" into the time series. The time intervals between the events are no longer of fixed length. The transformation is illustrated in Figure 1.3. The lower and upper thresholds are chosen such that there are in each case 10% of the data beyond the threshold. In other words, the lower and upper thresholds are the empirical 10% and 90% quantiles.

Note that a fixed threshold may be over simplistic. We consider a period of 20 years and one should not assume that the time series is stationary over this long time period. Price changes that are considered as extreme in one year may be considered as quite ordinary in another year. Therefore, it would probably be better to take a threshold that varies over time. This issue will

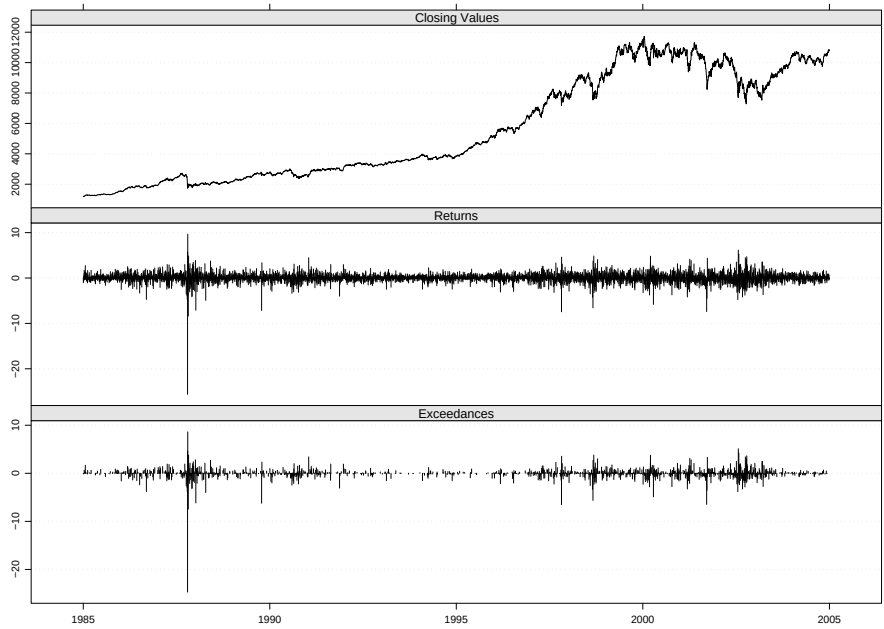

Figure 1.3: Dow Jones Industrial Average from 1985 to 2005

not be pursued further here.

Specification of a Hawkes Model. As explained earlier, it is more convenient for practical applications to take separable transfer functions. Since we want to demonstrate some practical issues, we are going to choose a Hawkes model that is convenient in a pedagogical sense and not necessarily the optimal choice.

Clearly, the Dow Jones Index is a one-dimensional time series. But the number of components of the data set and the number of components of the Hawkes process do not need to coincide. Hence, we do not necessarily need to take a univariate Hawkes model to describe this data. This means that there are many more Hawkes models that can potentially be used. To emphasize this point, we choose specifically a two-dimensional Hawkes model. The first component describes the negative returns and the second component describes the positive returns. In terms of Definition 1.1, we are therefore speaking about

a *genuine* multivariate Hawkes process.

We now specify the Hawkes model used in detail. Every two-dimensional genuine Hawkes model has the following parameters: The two immigration intensities $\eta_1,\ \eta_2$ and the four branching coefficients $\vartheta_{11},\ \vartheta_{12},\ \vartheta_{21},\ \vartheta_{22}$; see Definitions 1.4 and 1.8.

Recall the Definition 1.5, where we have introduce decay functions. A two-dimensional genuine multivariate Hawkes process has two decay functions w_1, w_2. For simplicity we use the same decay function w for both components. We take the exponential function $w(t) = \alpha \exp\{-\alpha t\}$. See also Remark 1.22 that explains why this decay function should be preferred if one has no other preferences.

As mark distribution we choose a Pareto distribution. Each component has its own mark distribution, see Definition 1.11. Hence, let us assume that the negative and positive exceedances both are Pareto distributed, each component with its own set of parameters. We take the parameterization from Section 1.3, which is:

$$f_j(x) = \frac{\rho_j \mu_j^{\rho_j}}{(x+\mu_j)^{\rho_j+1}}.$$

Finally we need to select a suitable impact function; see Definition 1.7. We take two linear impact functions $g_1,\ g_2$, one for each of the two components. From Section 1.3 we obtain that the linear impact function corresponding to a Pareto distribution is:

$$g_j(x) = \frac{(\rho_j - 1)(\rho_j - 2)}{\phi_j(\rho_j - 1)(\rho_j - 2) + \psi_j \mu_j (\rho_j - 2)}(\phi_j + \psi_j x).$$

Note that we have removed the quadratic term and renamed the parameters. In summary, the specified Hawkes model has the following 15 parameters:

$$\eta_1,\ \eta_2,\ \vartheta_{11},\ \vartheta_{12},\ \vartheta_{21},\ \vartheta_{22},\ \alpha,\ \rho_1,\ \rho_2,\ \mu_1,\ \mu_2,\ \phi_1,\ \phi_2,\ \psi_1,\ \psi_2.$$

Parameter Estimates. We estimate these parameters using the maximum likelihood method. The likelihood function is given in Proposition 1.18. The parameters are estimated using a numerical minimization algorithm. For the chosen data set we obtain the following parameter estimates, rounded to a

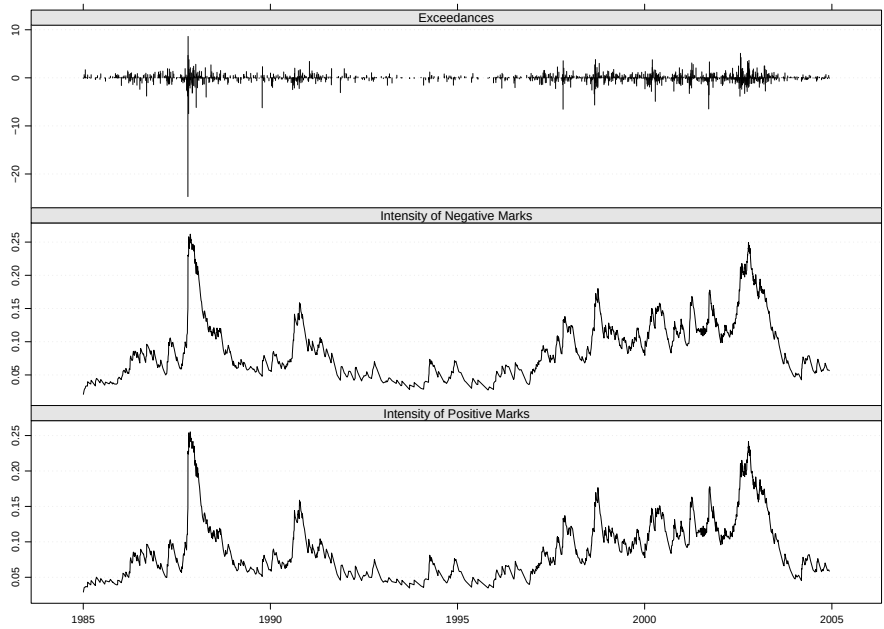

Figure 1.4: Estimated intensity processes

reasonable number of digits:

$$\begin{array}{lllll} \eta_1 = 0.021 & \eta_2 = 0.029 & \vartheta_{11} = 0.61 & \vartheta_{12} = 0.16 & \vartheta_{21} = 0.60 \quad \vartheta_{22} = 0.061 \\ \alpha = 0.015 & \rho_1 = 5.6 & \rho_2 = 7.2 & \mu_1 = 3.6 & \mu_2 = 4.2 \\ \phi_1 = 0.47 & \phi_2 = 1.1 & \psi_1 = 0.22 & \psi_2 = 0.0 & \end{array}$$

We do not give any error bounds for these parameters, as they are not relevant for the discussion below.

First Inspection of Parameter Estimates. The first question we need to answer is whether these parameters specify a well-defined Hawkes model or not. This depends on the following parameters, which specify the global structure of the Hawkes process: The immigration intensities η_1, η_2 and the branching

coefficients ϑ_{11}, ϑ_{12}, ϑ_{21}, ϑ_{22}. Consider first the branching matrix:

$$Q = \begin{pmatrix} 0.61 & 0.16 \\ 0.60 & 0.06 \end{pmatrix}.$$

The branching matrix Q describes the mean number of "children" of an event. For example, an event in component 2 has in the mean $Q_{12} = 0.16$ children in component 1.

Spectral Radius. An important characteristic of a Hawkes model is the spectral radius of its branching matrix, see Definition 1.15. For the above parameters, one obtains $\mathrm{Spr}(Q) = 0.744$. Since $\mathrm{Spr}(Q) < 1$, the Hawkes model is indeed well-defined. Clearly, one needs to be aware that there is some uncertainty associated with the estimate for $\mathrm{Spr}(Q)$. Hence, it might be very well possible that $\mathrm{Spr}(Q) \geq 1$ within some error bounds. Note that the degenerate case where $Q = \mathbb{O}_2$ would mean that the Hawkes process is actually a Poisson process. ($\mathbb{O}_2$ is the 2×2-matrix consisting of zeroes.)

Mean Number of Descendants. The branching matrix gives only the number of direct descendants. The following matrix gives the number of all descendants in all generations:

$$(\mathbb{1}_2 - Q)^{-1} - \mathbb{1}_2 = \begin{pmatrix} 2.40 & 0.57 \\ 2.18 & 0.43 \end{pmatrix}.$$

Note that $(\mathbb{1}_2 - Q)^{-1}$ counts the number of descendants including the originating event. Hence one has to subtract the identity matrix $\mathbb{1}_2$. Consider again a concrete example: The numbers above tell us that every event in component 2 has in the mean 0.57 descendants in component 1. Descendants include all directly or indirectly caused events of the root event.

Mean Intensity. Finally we calculate the mean intensity of the two components. Recall that η_1, η_2 describe the intensities of the immigrants. Define the vector $\boldsymbol{\eta} := (\eta_1, \eta_2)^\intercal$. The mean intensity of the Hawkes process is given by:

$$(\mathbb{1}_2 - Q)^{-1}\boldsymbol{\eta} = \begin{pmatrix} 0.0873 \\ 0.0871 \end{pmatrix}.$$

The two values are almost identical. This is indeed expected, since there are the same amount of negative exceedances than there are positive exceedances.

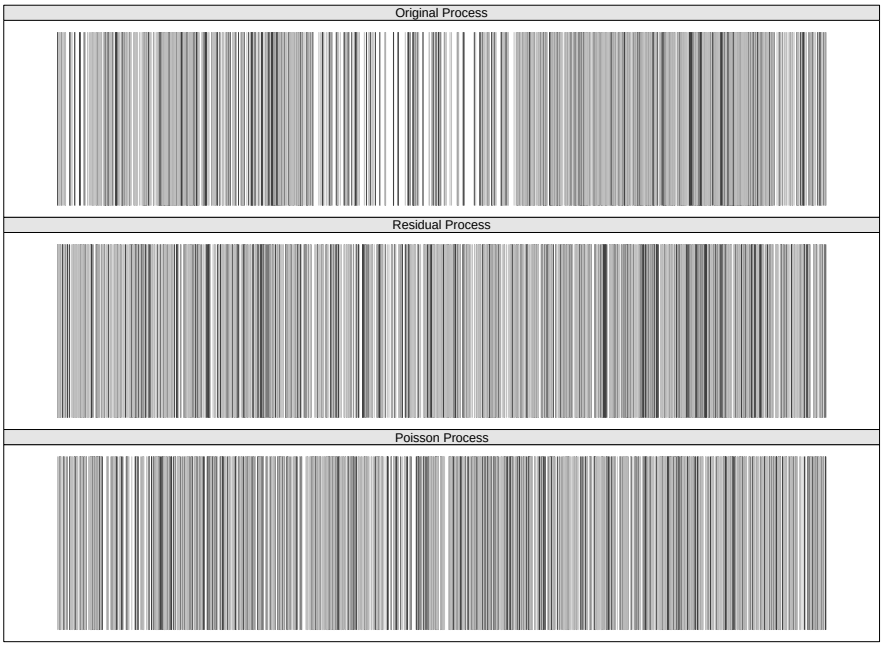

Figure 1.5: Barcode plot

Hence, the mean intensities should be close to each other.

Estimated Intensity Processes. To obtain a first visual idea of the estimated Hawkes process, we consider the estimated intensity processes. The two intensity processes are plotted in Figure 1.4. Note that the processes are almost identical. Again this is not a coincidence, since the two components are strongly influenced by each other.

Goodness-of-Fit Analysis. There are many ways of graphically assessing the goodness-of-fit of the estimated model. The plots we show here are all based on the residual process, see Definition 1.19. Since we have a two-dimensional Hawkes process, there are two residual processes. If the chosen model is appropriate, the residual processes should be two independent compound Poisson processes. One could call the residual processes also the *declustered* processes. One expects that the events are spread out evenly over the time interval, and

there should be no clusters any more.

There is not a unique, best way how one can check whether a given process is a Poisson process. But there are some obvious methods one can use. A Poisson process has several characterizing properties and we may check them in turn.

Barcode Plot. We look only at the ground processes of the residual processes, i.e. the residual processes without their marks. The marks can easily disturb the plots more than they are helping. Consider the plot in Figure 1.5, which one could call a *barcode plot*, for its obvious appearance. The first panel shows the original data set. A vertical bar is drawn for every event, irrespective of the corresponding mark. To distinguish between negative and positive returns, two different shades of gray are used. The second panel shows the residual process. Again, vertical bars are drawn for each event and different shades of gray are used for two components.

The residual processes are both defined on their own time scale. Moreover they are different from the time scale of the original process. Hence, we need to adapt the different time scales in order to compare the original process with his two residual processes. To this end, the time scales of the original process and the two residual processes are all normalized to the unit interval.

It is difficult to assess by eye whether the residual processes are indeed Poisson processes. It might look like the residual processes still exhibit some clustering. To get a better idea of how much irregularities one should expect in a Poisson process, the third panel shows a genuine Poisson process. It has been generated using the estimated mean intensities for the two residual processes.

The purpose of the barcode plot can therefore be summarized as follows: If the model chosen is appropriate, one should see a clear difference between the first and the second panel in terms of clustering. At the same time, the second panel should not exhibit more clustering than the third panel does.

Durations. The lengths of the time intervals between the events are called the *durations*. Since the residual processes should be Poisson processes with unit intensity, one expects that the durations are exponentially distributed with parameter one. We check this property next. Consider the two Q-Q-plots in Figure 1.6, one for each component of the residual process. The theoretical quantiles are plotted on the horizontal axis and the empirical quantiles on the vertical axis. The left Q-Q-plot refers to the residual process associated with the negative returns and the right hand side refers to the positive returns.

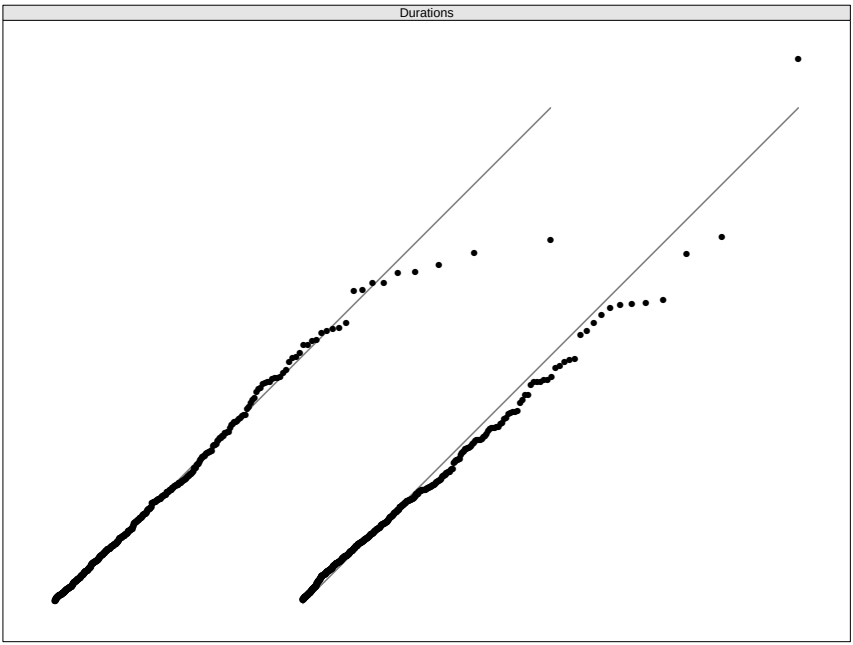

Figure 1.6: Q-Q-plot for durations

Residual Process. Recall the following characterizing property of a Poisson process: Conditioned on the number of events, the location of the events is uniformly distributed in the observation interval. We can check this property again with a Q-Q-plot. To this end, we scale the time domain of the residual processes to the unit interval. Assume therefore without loss of generality that the events of both residual processes are given by a sequence $t_{i=1,\dots,n}$, where $t_0 = 0$ and $t_n = 1$.

Note that the quantile function of the standard uniform distribution is the identity function. The Q-Q-plot is therefore the plot consisting of the points $(\frac{i}{n}, t_i)$, for $i = 1, \dots, n$, where t_i is the i-th event. If one interchanges the horizontal and vertical axis, one obtains a plot consisting of the points $(t_i, \frac{i}{n})$. But this is simply the rescaled counting function C. Recall that the counting function in this case is defined as $C(t) := \sum_{i=1}^{n} \mathbb{1}_{\{t_i \leq t\}}$, for $t \in [0, 1]$. The plot of the function $t \mapsto \frac{1}{n} N(t)$ is shown in Figure 1.7, again for the negative and positive returns separately.

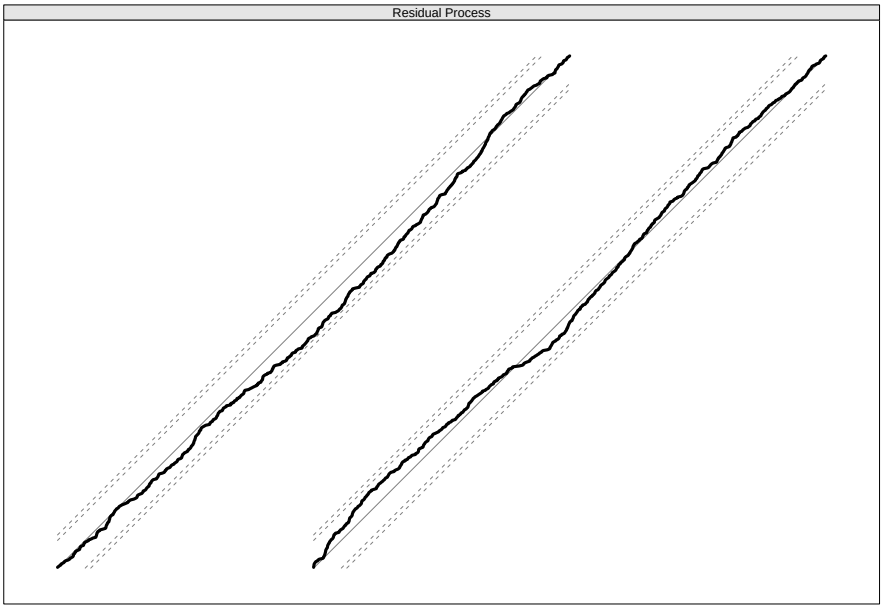

Figure 1.7: Counting function of residual process

Kolmogorov-Smirnov-Test. The counting function C is the also the base of a statistical test. Indeed, the Kolmogorov-Smirnov-Test is based on the value of the maximal deviation from the diagonal $d_n := \sup_{t\in[0,1]}|C(t) - t|$. This connection to the Kolmogorov-Smirnov statistic allows us to add error bounds to the plot. The two pairs of dashed lines correspond to the 95% and 99% error bounds.

We can also perform a formal Kolmogorov-Smirnov-Test using the test statistic d_n from above. For the first component of the residual process, which refers to the negative returns, we obtain $d_n = 0.0496$ with associated p-value 8.9%. For the second component we obtain $d_n = 0.0392$ with associated p-value 28.5%. Hence, neither component rejects the null hypothesis on a 5% confidence level.

Chapter 2

Counting Measures

This chapter is not directly related to Hawkes processes. Instead, it introduces a few elementary concepts, most notably reduced measures, point configurations and an extended notion of Kronecker's delta function. The presented definitions and related results are a preparation for the next chapter which deals with higher order moment measures of point processes.

2.1 Motivation and Objectives

The book [DVJ03] has served as a general source for this chapter. The following lines give an overview over the approach chosen and explain in what sense the presentation differs and extends the material from the mentioned reference.

Reduced Measures. The first part deals with reduced measures, see Definition 2.13. Reduced measures appear naturally if one looks at moment measures of stationary point processes. The main idea behind reduced measures is to isolate the one-dimensional symmetry of a translation-invariant measure, see Definition 2.9. Basically, a translation-invariant measure can be represented as the product measure of the reduced measure and the Lebesgue measure. The precise statement is formulated in Theorem 2.14.

The reason why reduced measures are useful, and even needed, is explained in Section 8.1 of [DVJ03]. The authors look at translation-invariant measures from the more general perspective of measures on topological groups. In this setting, translation-invariant measures satisfy an invariance property under a group of transformation. One can then use some of the powerful results from

the theory of measures on topological groups. More about this theory can be found in Appendix A2.7 of the above reference. The main result is a factorization theorem, see Lemma A2.7.II and Lemma A2.7.III. For further background one can consult also [Bou63] or [Kri74], Section 2.6.2 about disintegration of invariant measures under a group of transformations.

I decided not to embed translation-invariant measures in the more general theory of measures on topological groups. Instead, I give a direct, self-contained proof of the factorization result, see Theorem 2.14, using only elementary mathematics. In this way, the abstract theory of measures on topological groups can be avoided, and a explicit definition can be given, see Definition 2.13.

Factorial Product Measures. The next part of this chapter deals with so-called *factorial products* of point configurations. Although the idea behind factorial products is quite simple, a formal definition is not that accessible. In [DVJ03], factorial products are first defined on a specific family of rectangular sets, where factorial products of the number of points are taken. And in a second step, these partially defined measures are then extended to factorial product measures. For details consult Section 5.4 and Proposition 5.4.I in [DVJ03].

Results concerning factorial product measures are essentially combinatorial statements. Moreover, there is an inherent relation between factorial products and partitions of sets of the form $\{1, \ldots, n\}$. That partitions of this form can be quite helpful if one deals with moment measures has already been observed by [Kri74], see especially Section 2.6.4 therein. This means that factorial product measures exhibit a certain underlying complexity that can not be avoided and one has to deal with sometimes complicated expressions.

I present an alternative definition for factorial product measures: The idea is to define an extended notion of Kronecker's delta-function, see Definition 2.28. Moreover, to exhibit the relation to partitions more explicitly, I introduce a second extension of Kronecker's delta-function, see Definition 2.29. Using this notation, results concerning factorial product measures can then be formulated in compact form.

I think this approach offers an interesting alternative. To my knowledge, delta-functions of the kind introduced here have not been used systematically in the context of point processes. The notation chosen has the advantage that combinatorial statements can now be formulated in standard algebraic notation, see e.g. the decomposition given in Theorem 2.32.

Decomposition of Product Measures. The last part of this chapter applies the results concerning the delta-functions in the context of point configurations in order to derive a series of decomposition formulas, see Corollary 2.41, Theorem 2.42 and Corollary 2.43. These decompositions become important as soon as one is dealing with moment measures of point processes, the reason being that factorial moment measures are often easier to deal with than ordinary product measures. Point processes and moment measures are the topic of the next chapter.

2.2 Event Spaces

We call the space in which the points of a point process lie the *event space*. Often the event space is simply $\mathbb{R}$ and is interpreted as the time line. If one has a marked point process, that is each of the events has a mark attached to it, then these marks lie in a different space, the so-called *mark space*. This section defines these two spaces and points out in which way they differ.

We will use below several times disjoint unions. For completeness, recall the following standard definition. Note that the actual representation of the disjoint union is not relevant. Important is only that a disjoint union strictly distinguishes between elements from different sets. Clearly, one could think of other, conceptually equivalent definitions:

2.1 Definition (Disjoint union). Let $\mathscr{X}_1, \ldots, \mathscr{X}_d$ be a family of arbitrary sets. The *disjoint union* of this family is defined as:

$$\mathscr{X}_1 \sqcup \ldots \sqcup \mathscr{X}_d := \Big\{(j,x) \in \{1,\ldots,d\} \times \bigcup_{k=1}^{d} \mathscr{X}_k : x \in \mathscr{X}_j\Big\}. \qquad \diamondsuit$$

The next three definitions introduce event, mark and state spaces. The definitions are almost identical. But since these spaces are not used in the same way, it is justified to given them separate names.

The event space is the space where the points live, most often this is the time axis. The mark space contains additional values which are attached to these points. If one combines the event space and the mark space, one obtains the state space. An event is completely described once its location in the state space is known.

There are several reasons why we introduce a specific notation for event spaces. One reason is that we want that the notation for event and mark

spaces is compatible. Another explanation is given in Remark 2.10.

2.2 Definition (Event space). Let "$\sqcup$" denote the disjoint union.

(*1*) Let $e \geq 1$ be an integer and define the one-dimensional *univariate event space* by $\mathbb{E} := \mathbb{R}^e$.

(*2*) Let $\boldsymbol{e} = (e_1, \ldots, e_d)$ be an index vector, where $e_j \geq 1$ for $j \in \{1, \ldots, d\}$ and define the component spaces $\mathbb{E}_j := \mathbb{R}^{e_j}$. The *multivariate event space* $\mathbb{E}$ is then defined by:

$$\mathbb{E} := \mathbb{E}_1 \sqcup \ldots \sqcup \mathbb{E}_d.$$

We sometimes call $\mathbb{E}$ also the *combined event space*, if we refer specifically to this disjoint union. $\diamondsuit$

It is important to note that the multivariate event space $\mathbb{E}$ is the disjoint union and not the Cartesian product of the component spaces $\mathbb{E}_j$. It is not uncommon that some of the component spaces $\mathbb{X}_j$ coincide, and hence the disjoint union is indeed needed.

2.3 Definition (Mark space). See the remark following this definition for more about regularity conditions.

(*i*) *Univariate mark space.* Let $\mathbb{X}$ be an arbitrary, measurable space. It is called a *mark space* if it is used in connection with marked point configurations.

(*ii*) *Multivariate event space.* Let $\mathbb{X}_j$, for $1 \leq j \leq d$, be univariate mark spaces. Define the disjoint union

$$\mathbb{X} := \mathbb{X}_1 \sqcup \ldots \sqcup \mathbb{X}_d.$$

Sometimes, when we specifically refer to this disjoint union, we call $\mathbb{X}$ also the *combined mark space*. $\diamondsuit$

Note that one cannot define point processes on arbitrary spaces. This means that the underlying event space needs to satisfy some mild technical conditions. Since we use only Euclidean spaces as event spaces, no further conditions are required.

2.4 Remark (Regularity conditions). The following comments explain the minimal assumptions one has to impose on the event and mark spaces:

(*i*) *Regularity conditions for event spaces.* There are several of minimal assumptions for event spaces: If one only wants to define a Poisson process on some space, the event space $\mathbb{E}$ has to be at least a measurable space. See the Poisson process existence theorem by Kingman and Mecke, Proposition 10.4 in [Kal97].

If one wants to define more general processes on $\mathbb{E}$, more regular spaces are required. For example a Cox process cannot be defined on an arbitrary space measurable space, see the existence theorem for Cox processes, Proposition 10.5 in [Kal97].

There are two standard regularity assumptions one can find in the literature: A common assumption is that $\mathbb{X}$ is a complete, separable metric space. This is the assumption made throughout the book [DVJ03]. Another common assumption is that $\mathbb{X}$ is a locally compact, second-countable Hausdorff space. This assumption is made in the book [Kal97].

Note that we define all point processes on the event space $\mathbb{E}$, which is by definition of the form $\mathbb{R}^e$. Hence, the event spaces used in this thesis satisfy certainly the minimal requirements.

(*ii*) *Regularity conditions for mark spaces.* What regularity conditions the mark spaces have to satisfy depends on their exact use. In our case, we do not define any sort of point processes on the mark spaces. Hence, we do not need to impose any regularity conditions on the mark spaces, except that they are measurable spaces.

The above considerations show that the mark spaces can be very general spaces. Hence, obvious choices as e.g. $\mathbb{X} := \mathbb{R}^n$ or $\mathbb{X} := \mathbb{N}^n$ are certainly valid. ♢

Before we define point configurations, we want once more to clarify the difference between event and mark spaces. These two spaces are very closely related and it might seem easier to combine these two spaces and to define only one space, the so-called *state space*. Every point would then takes a value from this state space. But as we see later, especially when one deals with simple point configurations, it is important to clearly distinguish between the event mark and the mark space.

For completeness, we mention the notion of a state space anyhow, although we will not deal with state spaces, and always distinguish between event and mark spaces.

2.5 Remark (State space). Let $\mathbb{E}$, $\mathbb{X}$ and $\mathbb{E}_j$, $\mathbb{X}_j$ be event and mark spaces as in Definitions 2.2 and 2.3.

(1) *Univariate state space.* Define $\mathbb{Y} := \mathbb{E} \times \mathbb{X}$. Consider a point configuration of $\mathbb{E}$ with marks in $\mathbb{X}$. Then every point of this point configuration can be identified with a unique element from $\mathbb{Y}$.

(2a) *Multivariate state space.* The following space has the same interpretation as in the univariate case:

$$\mathbb{Y} := \mathbb{Y}_1 \sqcup \ldots \sqcup \mathbb{Y}_d = (\mathbb{E}_1 \times \mathbb{X}_1) \sqcup \ldots \sqcup (\mathbb{E}_d \times \mathbb{X}_d).$$

(2b) *Common event space.* Consider the special case where all event spaces coincide, i.e. $\mathbb{E}_i = \mathbb{E}_j$, for all i, j. In this case, the above disjoint union can also be written as:

$$\mathbb{Y} = \mathbb{E} \times \mathbb{X} = \mathbb{E} \times (\mathbb{X}_1 \sqcup \ldots \sqcup \mathbb{X}_d). \qquad \diamondsuit$$

Usually the event space is $\mathbb{R}$ and interpreted as time. In the multivariate case, this means that $\mathbb{E} := \mathbb{E}_1 = \ldots = \mathbb{E}_d = \mathbb{R}$. Hence, the case where all event spaces coincide, as in the last part above, is not an exception but the standard situation.

Some Notation. We will often deal with Cartesian products of the component spaces $\mathbb{E}_j$. To this end, we introduce the following notation:

2.6 Notation (Index vector). Let $\mathbb{E}$ be d-dimensional event space. Take a vector $\boldsymbol{j} \in \{1, \ldots, d\}^n$, which in this context is called an *index vector*. Then define:

$$\mathbb{E}^{(\boldsymbol{j})} := \mathbb{E}_{j_1} \times \ldots \times \mathbb{E}_{j_n}. \qquad \diamondsuit$$

Note that the univariate and multivariate event spaces are all based on the real line $\mathbb{R}$. Hence, we will often use the Lebesgue measure:

2.7 Notation (Lebesgue measure). Let $\lambda_{\mathbb{R}^m}$ denote the *Lebesgue measure* on $\mathbb{R}^m$. $\diamondsuit$

For readability, we will not always mention measurability conditions, but nevertheless assume they are satisfied. Hence, we follow the convention:

2.8 Convention (Measurability of sets). Let $\mathscr{E}$ be any of the spaces introduced above, i.e. let $\mathscr{E}$ be one of the spaces $\mathbb{E}$, $\mathbb{E}$ or $\mathbb{X}$, $\mathbb{X}$; or even a Cartesian product of one of these spaces. Always assume that:

(*i*) *Measurability of sets.* Whenever we take a subset $E \subseteq \mathscr{E}$, we will implicitly assume that E is measurable with respect to the Borel-σ-algebra on $\mathscr{E}$.

(*ii*) *Measurability of functions.* Whenever we take a function f on $\mathscr{E}$, we will implicitly assume that f is measurable with respect to the Borel-σ-algebra on $\mathscr{E}$. ♢

Signed Measures. Before we deal with translation-invariant measures, let us restate some common definitions:

2.9 Definition (Properties of signed measures). Let $\mathbb{E}$ be a univariate event space.

(*i*) A signed measure μ on $\mathbb{E}^n$ is called *locally-finite*, if $\mu^+(E) < \infty$ and $\mu^-(E) < \infty$, for all bounded $E \subseteq \mathbb{E}^n$.

(*ii*) A signed measure μ on $\mathbb{E}^n$ is called *finite*, if both, the positive part $\mu^+(\mathbb{E}^n) < \infty$ and the negative part $\mu^-(\mathbb{E}^n) < \infty$, are finite.

(*iii*) A locally-finite, signed measure μ on $\mathbb{E}^n$ is called *translation-invariant* if for all bounded $E \subseteq \mathbb{E}$ of the form $E := E_1 \times \ldots \times E_n$ with $E_k \subseteq \mathbb{E}$ and all $z \in \mathbb{E}$:

$$\mu\big(\{E_1 + z\} \times \ldots \times \{E_n + z\}\big) = \mu(E_1 \times \ldots \times E_n).$$

(*iv*) A locally-finite, signed measure μ on $\mathbb{E}^n$ is called *symmetric* if for all families $\{E_1, \ldots, E_n\}$ of bounded sets and all permutations π of the set $\{1, \ldots, n\}$ of the form $\{\sigma(1), \ldots, \sigma(n)\}$ one has:

$$\mu(E_1 \times \ldots \times E_n) = \mu(E_{\sigma(1)} \times \ldots \times E_{\sigma(n)}).$$ ♢

Note that in last two parts of the above definition, one has to assume that the sets E_k are bounded, since the measures are only locally-finite.

Let us give a short explanation why the notation $\mathbb{E}$ for an event space is used, instead of simply writing $\mathbb{R}^m$: The reason is that one can represent $\mathbb{R}^m = \mathbb{E}^n = (\mathbb{R}^e)^n$ in different ways. Two obvious choices are:

$$e := 1,\ n := m \qquad \text{or} \qquad e := m,\ n := 1.$$

Clearly, in both cases one obtains $\mathbb{E}^n = \mathbb{R}^m$, but as we show in the next remark, they are not equivalent:

2.10 Remark (Reasoning behind event spaces). Let us demonstrate that the notions of *translation-invariance* and *symmetry* do not coincide in these two cases:

(1) *Case $e := 1$, $n := m$.* Clearly, one has $\mathbb{E} = \mathbb{R}$ and $\mathbb{E}^n = \mathbb{R}^m$. We check what translation-invariance and symmetry mean in this situation:

(i) *Translation-invariance.* A measure μ on $\mathbb{E}^n$ is a translation-invariant if and only if

$$\mu\big(\{E_1 + z\} \times \ldots \times \{E_m + z\}\big) = \mu(E_1 \times \ldots \times E_m),$$

for all $E_k \subseteq \mathbb{E} = \mathbb{R}$ and $z \in \mathbb{E} = \mathbb{R}$. Hence, μ is invariant under shifts along the "diagonal" of the Euclidean space $\mathbb{R}^m$.

(ii) *Symmetry.* Now assume μ is a symmetric measure on $\mathbb{E}^n$. Then for all permutations $\{E_{\sigma(1)}, \ldots, E_{\sigma(n)}\}$ of a sequence of bounded sets $E_k \subseteq \mathbb{E} = \mathbb{R}$, where $1 \leq k \leq n$:

$$\mu(E_1 \times \ldots \times E_n) = \mu(E_{\sigma(1)} \times \ldots \times E_{\sigma(n)}).$$

(2) *Case $e := m$, $n := 1$.* Clearly, one has $\mathbb{E} = \mathbb{E}^n = \mathbb{R}^m$. Again we check what translation-invariance and symmetry mean in this situation:

(i) *Translation-invariance.* A measure μ on $\mathbb{E}^n$ is a translation-invariant if and only if

$$\mu(E + z) = \mu(E), \text{ for all } E \subseteq \mathbb{R}^m \text{ and } z \in \mathbb{E} = \mathbb{R}^m.$$

In other words, μ is invariant under all shifts of the Euclidean space $\mathbb{R}^m$. Since only multiples of the Lebesgue measure have this property, μ is therefore a multiple of the Lebesgue measure.

(ii) *Symmetry.* By definition, μ is symmetric on $\mathbb{E}$ if and only if

$$\mu(E) = \mu(E_{\sigma(1)}), \text{ for all } E \subseteq \mathbb{R}^m,$$

for all permutations $\pi \equiv \{\sigma(1)\}$ of the set $\{1\}$. As there is only one such permutation, this condition is trivially satisfied for any measure μ. $\diamondsuit$

In summary, depending on the exact definition of $\mathbb{E}^n$, one obtains different notions of translation-invariance and symmetry.

2.11 Remark (Convention). In the remainder of this chapter, we implicitly assume the following:

(*i*) Whenever we take a measure μ, we assume it is either *non-signed* and *locally-finite* or alternatively it is *signed* and *finite*.

(*ii*) Whenever we take a function f, we implicitly assume f is *bounded* if we integrate f with respect to a finite, signed measure, and we assume f *non-negative* if we integrate f with respect to a locally-finite, signed measure. ♢

Note that the second assumption is not really a restriction. It only guarantees that integrals of the form $\int_{\mathbb{E}} f(x)\mu(dx)$ are always well-defined. In the case where f is non-negative and μ is a non-signed measure, the integral can in fact be $+\infty$, but we consider this case also as acceptable and well-defined.

2.3 Translation Invariance

In this section, we analyze some properties of translation-invariant measures. The main result is the representation in Theorem 2.14, where the relation between translation-invariant measures and the so-called *reduced measures* is established.

2.12 Definition (Linear transformations). Let $\mathbb{E}$ be an event space and take some $1 \leq k \leq n$.

(*1*) *Projections and embeddings.* Define the linear mappings:

$$\begin{aligned} \mathrm{I}^{\langle k\rangle} &: \mathbb{E}^{n-1} \mapsto \mathbb{E}^n \qquad & \mathrm{I}^{\langle k\rangle}(\boldsymbol{y}) &:= (y_1, y_2, \ldots, y_{k-1}, 0, y_k, \ldots, y_{n-1}), \\ (\cdot)^{\langle k\rangle} &: \mathbb{E}^n \mapsto \mathbb{E}^{n-1} & \boldsymbol{x}^{\langle k\rangle} &:= (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n). \end{aligned}$$

(*2*) *Shear of the Euclidean space.* Define the linear mappings $\mathbb{E}^n \mapsto \mathbb{E}^n$:

$$\begin{aligned} \sigma^{\langle k\rangle}(\boldsymbol{y}, z) &:= \mathrm{I}^{\langle k\rangle}(\boldsymbol{y}) + z\mathbf{1}_n, & &\text{for } \boldsymbol{y} \in \mathbb{E}^{n-1} \text{ and } z \in \mathbb{E}, \\ \tau^{\langle k\rangle}(\boldsymbol{x}) &:= \big(\boldsymbol{x}^{\langle k\rangle} - x_k\mathbf{1}_{n-1}, x_k\big), & &\text{for } \boldsymbol{x} \in \mathbb{E}^n, \end{aligned}$$

where $\mathbf{1}_n$ is a vector consisting of ones, of length n. ♢

The above functions have the following interpretation: $\mathrm{I}^{\langle k\rangle}$ takes a vector $\boldsymbol{y} \in \mathbb{E}^{n-1}$ and inserts a 0 at position k, whereas $(\cdot)^{\langle k\rangle}$ takes a vector $\boldsymbol{x} \in \mathbb{E}^n$ and removes the k-th component. Moreover, $\sigma^{\langle k\rangle}$ shifts the space $\mathbb{E}^n$ along its

hyper-diagonal by the amount z. In the same way, $\tau^{\langle k\rangle}$ shifts the space $\mathbb{E}^n$ in the opposite direction. As a consequence, $\tau^{\langle k\rangle}$ is the inverse of $\sigma^{\langle k\rangle}$.

The following definition is motivated by Proposition 8.1.I in [DVJ03]. Let $\mathbb{D} \subseteq \mathbb{E}$ be a fixed set with $\lambda_{\mathbb{E}}(\mathbb{D}) = 1$. The actual choice of $\mathbb{D}$ is not relevant; indeed the unit cube in $\mathbb{E}$ would do.

2.13 Definition (Reduced measure). Let μ be a translation-invariant measure on $\mathbb{E}^n$ and $1 \leq k \leq d$. The k-th *reduced measure* $\breve{\mu}^{\langle k\rangle}$ is a locally-finite measure on $\mathbb{E}^{n-1}$ and is defined by:

$$\begin{aligned}\breve{\mu}^{\langle k\rangle}(B) &:= \int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle k\rangle}\big](\boldsymbol{x})\mu(d\boldsymbol{x}) \\ &:= \int_{\mathbb{E}^n} \mathbb{1}_{B\times\mathbb{D}}\big(\boldsymbol{x}^{\langle k\rangle} - x_k\mathbf{1}_{n-1}, x_k\big)\mu(d\boldsymbol{x}),\end{aligned} \tag{2.1}$$

for bounded sets $B \subseteq \mathbb{E}^{n-1}$. ♢

The above definition may seem rather abstract. It is however difficult to give an intuitive definition for the reduced measure $\breve{\mu}$ without loosing mathematical correctness. Loosely speaking, a translation-invariant measure μ on $\mathbb{E}^n$ satisfies a one-dimensional symmetry. Therefore it should be possible to capture the full information content of μ on the lower-dimensional space $\mathbb{E}^{n-1}$; and this is what the measure $\breve{\mu}$ does.

The following theorem is the main reason why one considers reduced measures, in the first place. It explains how one can isolate the one-dimensional symmetry of a translation-invariant measure. Note that it is a special case of Lemma A2.7.III in [DVJ03], however we will give a self-contained proof.

2.14 Theorem (Reduction formulas). Let $\mathbb{E}$ be an event space and μ a translation-invariant measure on $\mathbb{E}^n$. For two functions f, g on $\mathbb{E}^n$, one has:

$$\int_{\mathbb{E}^n} f(\boldsymbol{x})\mu(d\boldsymbol{x}) = \int_{\mathbb{E}}\int_{\mathbb{E}^{n-1}} [f \circ \sigma^{\langle k\rangle}](\boldsymbol{y}, z)\breve{\mu}^{\langle k\rangle}(d\boldsymbol{y})dz, \tag{2.2}$$

$$\int_{\mathbb{E}^n} [g \circ \tau^{\langle k\rangle}](\boldsymbol{x})\mu(d\boldsymbol{x}) = \int_{\mathbb{E}}\int_{\mathbb{E}^{n-1}} g(\boldsymbol{y}, z)\breve{\mu}^{\langle k\rangle}(d\boldsymbol{y})dz. \tag{2.3}$$

♢

Recall that the definition of the reduced measure $\breve{\mu}^{\langle k\rangle}$, see Equation (2.1), depends a-priori on the set $\mathbb{D}$. But with the above theorem one can easily show that the definition is a-posteriori independent of $\mathbb{D}$:

2.15 Corollary (Consistency of definition). Take a translation-invariant measure μ on $\mathbb{E}^n$ and fix some $1 \leq k \leq d$. Let $\mathbb{D}_1, \mathbb{D}_2 \subseteq \mathbb{E}$ be two sets with $\lambda_{\mathbb{E}}(\mathbb{D}_1) = \lambda_{\mathbb{E}}(\mathbb{D}_2) = 1$. Now denote with $\breve{\mu}_1{}^{\langle k \rangle}$ the reduction with respect to $\mathbb{D}_1$ and with $\breve{\mu}_2{}^{\langle k \rangle}$ the reduction with respect to $\mathbb{D}_2$. Then

$$\breve{\mu}_1^{\langle k \rangle} = \breve{\mu}_2^{\langle k \rangle}.$$

Hence, the definition of the reduced measure is independent of the choice of the set $\mathbb{D}$. ♢

Note that we anticipated this result in Definition 2.13 and used from the beginning on the notation $\breve{\mu}^{\langle k \rangle}$, where no distinction on the set $\mathbb{D}$ is made. The above corollary justifies this notation.

2.16 Proposition (Properties of reduced measures). Let $\mathbb{E}$ be a univariate event space.

(1) *Linearity.* Let μ_1, μ_2 be translation-invariant measures on $\mathbb{E}^n$ and take $\alpha_1, \alpha_2 \in \mathbb{R}$. Then, for all $1 \leq k \leq n$:

$$\left[\alpha_1 \breve{\mu}_1 + \alpha_2 \breve{\mu}_2\right]^{\langle k \rangle} = \alpha_1 \breve{\mu}_1^{\langle k \rangle} + \alpha_2 \breve{\mu}_2^{\langle k \rangle}.$$

(2) *Anti-symmetry in two dimensions.* Let μ be a translation-invariant measure on $\mathbb{E}^2$. Then

$$\breve{\mu}^{\langle 1 \rangle}(dy) = \breve{\mu}^{\langle 2 \rangle}(-dy).$$

(3) *Density of reduced measure.* Let μ be a translation-invariant measure on $\mathbb{E}^n$. Assume it has the density m with respect to the Lebesgue measure $\lambda_{\mathbb{E}^n}$. Then, for all $1 \leq k \leq n$, the reduced measure $\breve{\mu}^{\langle k \rangle}$ has the density

$$\breve{m}^{\langle k \rangle}(\boldsymbol{y}) := \int_{\mathbb{D}} [m \circ \sigma^{\langle k \rangle}](\boldsymbol{y}, z) dz.$$

Moreover, for $(\boldsymbol{y}, z) \in \mathbb{E}^{n-1} \times \mathbb{E}$, one has:

$$\breve{m}^{\langle k \rangle}(\boldsymbol{y}) = [m \circ \sigma^{\langle k \rangle}](\boldsymbol{y}, z), \text{ almost everywhere,}$$

with respect to the Lebesgue measure. ♢

Since the last equation above does not necessarily hold for all $z \in \mathbb{E}$, it cannot be used as a definition for $\breve{m}^{\langle k \rangle}$. Indeed, this is the reason why the definition of $\breve{m}^{\langle k \rangle}$ contains an integral over the unit interval $\mathbb{D}$.

2.4 Symmetric Measures

As we have seen in the previous section, for a general translation-invariant measure μ on $\mathbb{E}^n$ there exist n different reduced measures. But if the measure μ is additionally symmetric, all reduced measures coincide and one has to deal only with a single reduced measure, which is then called *the* reduced measure. This is the content of the next proposition:

2.17 Proposition (Reduction of symmetric measures). Let μ be a symmetric, translation-invariant measure on $\mathbb{E}^n$. Then the reduced measures coincide, that is $\breve{\mu}^{\langle k\rangle} = \breve{\mu}^{\langle l\rangle}$, for all $1 \le k, l \le n$. Hence, there is only one reduced measure, which is denoted by $\breve{\mu} \equiv \breve{\mu}^{\langle k\rangle}$. ♢

In the case of a symmetric measure μ, some of the previous statements can be formulated in a more compact form. To this end, identify an element $\boldsymbol{x} \in \mathbb{E}^n$ with the pair $(\breve{\boldsymbol{x}}, x_n)$ from the space $\mathbb{E}^{n-1} \times \mathbb{E}$:

2.18 Remark (Alternative definitions). Let μ be a symmetric, translation invariant measure on $\mathbb{E}^n$. An equivalent definition for $\breve{\mu}$ is:

$$\breve{\mu}(B) = \int_{\mathbb{D}} \mu(B + x_n \times dx_n) = \int_{\mathbb{E}^n} \mathbb{1}_{B\times\mathbb{D}}(\breve{\boldsymbol{x}} - x_n, x_n)\mu(d\breve{\boldsymbol{x}} \times dx_n),$$

for bounded sets $B \subseteq \mathbb{E}^{n-1}$. If additionally μ has the density m, then $\breve{\mu}$ has the density

$$\breve{m}(\boldsymbol{y}) = \int_{\mathbb{D}} m(\boldsymbol{y} + z, z)dz, \text{ for } \boldsymbol{y} \in \mathbb{E}^{n-1}. \qquad ♢$$

The reduction formulas can also be restated in a more compact form:

2.19 Corollary (Reduction formulas in the symmetric case). Let μ be a symmetric, translation invariant measure on $\mathbb{E}^n$. For two functions f, g on $\mathbb{E}^n$ one has:

$$\int_{\mathbb{E}^n} f(\boldsymbol{x})\mu(d\boldsymbol{x}) = \int_{\mathbb{E}} \int_{\mathbb{E}^{n-1}} f(\boldsymbol{y} + z, z)\breve{\mu}(d\boldsymbol{y})dz, \tag{2.4}$$

$$\int_{\mathbb{E}^n} g(\breve{\boldsymbol{x}} - x_n, x_n)\mu(d\breve{\boldsymbol{x}} \times dx_n) = \int_{\mathbb{E}} \int_{\mathbb{E}^{n-1}} g(\boldsymbol{y}, z)\breve{\mu}(d\boldsymbol{y})dz. \tag{2.5}$$

♢

One can write the transformation formulas also in a more intuitive way, as we show next. The disadvantage is that the following notation is less suitable for formal calculations:

2.20 Remark (Interpretation & intuitive notation). Let μ be a symmetric, translation-invariant measure on $\mathbb{E}^n$.

(1) *Transformation formulas in terms of sets.* For two sets $E \subseteq \mathbb{E}^{n-1}$, $F \subseteq \mathbb{E}$, one has:

$$\mu(E \times F) = \int_{E\times F} \mu(d\boldsymbol{x} \times dy) = \int_{E\times F} \breve{\mu}(d\boldsymbol{x} - y)dy,$$
$$\breve{\mu}(E)\lambda_{\mathbb{E}}(F) = \int_{E\times F} \breve{\mu}(d\boldsymbol{x})dy = \int_{E\times F} \mu(d\boldsymbol{x} - y \times dy).$$

(2) *Transformation formulas in terms of differentials.* In terms of differentials, the above two equations can be written as:

$$\mu(d\boldsymbol{x} \times dy) = \breve{\mu}(d\boldsymbol{x} - y)dy, \qquad \breve{\mu}(d\boldsymbol{x})dy = \mu(d\boldsymbol{x} - y \times dy). \qquad \diamondsuit$$

Next, we consider some special cases of translation-invariant measures and determine their reduced counterparts:

2.21 Proposition (Reduction of Lebesgue measures). Let $\mathbb{E}$ be a univariate event space.

(1) *Lebesgue measure on the Euclidean space.* Consider the Lebesgue measure on $\mathbb{E}^n$:

$$\mu_1(d\boldsymbol{x}) := \lambda_{\mathbb{E}^n}(d\boldsymbol{x}).$$

The reduced measure is again the Lebesgue measure, i.e.:

$$\breve{\mu}_1(d\boldsymbol{y}) = \lambda_{\mathbb{E}^{n-1}}(d\boldsymbol{y}). \tag{2.6}$$

(2) *Lebesgue measure on the diagonal.* Let $\delta^n_{\boldsymbol{z}}$, for $\boldsymbol{z} \in \mathbb{E}^n$, denote the Dirac-measure on $\mathbb{E}^n$, i.e. the measure with unit mass at the point $\boldsymbol{z}$. Consider the measure

$$\mu_2(d\boldsymbol{x}) := \int_{\mathbb{E}} \delta_{(z,\dots,z)}(d\boldsymbol{x})dz,$$

on $\mathbb{E}^n$. The reduced measure is:

$$\breve{\mu}_2(d\boldsymbol{y}) = \delta^{n-1}_{\boldsymbol{0}}(d\boldsymbol{y}). \tag{2.7}$$
$\diamondsuit$

When we analyze moment measures of Hawkes processes, integrals of a similar form as in the definition of reduced measures will appear. But this time, the underlying measure ν is not translation-invariant. It is nevertheless convenient to introduce a symbol for this type of reduction, in analogy to the reduced measure:

2.22 Definition (Pseudo-reduction of symmetric measures). Let ν be a finite (not necessarily translation-invariant), symmetric measure on $\mathbb{E}^n$. Define the *pseudo-reduced* measure $\mathring{\nu}$ by:

$$\mathring{\nu}(B) := \int_{\mathbb{E}} \nu(B + x_n \times dx_n) = \int_{\mathbb{E}^n} \mathbb{1}_B(\check{\boldsymbol{x}} - x_n)\nu(d\check{\boldsymbol{x}} \times dx_n),$$

for bounded sets $B \subseteq \mathbb{E}^{n-1}$. ♢

Note that in contrast to the reduced measure $\breve{\mu}$, the integration domain is $\mathbb{E}$ and not the unit cube $\mathbb{D}$. Also note that ν has additionally to be symmetric, whereas for the regular reduced measure, μ does not need to be symmetric.

When we deal with moment measures of Hawkes processes, translation-invariant measures with a special form will show up. In the next proposition, we take a closer look at some of these special cases:

2.23 Proposition (Reduction of derived measures). Let $\mathbb{E}$ be a univariate event space.

(1) *Multivariate underlying measure.* Let ν_1 be a finite, symmetric measure on $\mathbb{E}^n$. Define the translation-invariant measure μ_1 on $\mathbb{E}^n$ by:

$$\mu_1(d\boldsymbol{x}) := \int_{\mathbb{E}} \nu_1(d\boldsymbol{x} - w)dw.$$

The associated reduced measure is

$$\breve{\mu}_1(d\boldsymbol{y}) = \mathring{\nu}_1(d\boldsymbol{y}). \tag{2.8}$$

(2) *Product of univariate underlying measures.* Let ν_2 be a finite, symmetric measure on $\mathbb{E}$. Define the translation-invariant measure μ_2 on $\mathbb{E}^n$ by:

$$\mu_2(d\boldsymbol{x}) := \prod_{k=1}^{n} \Big[\int_{\mathbb{E}} \nu_2(dx_k - w_k)dw_k \Big].$$

The associated reduced measure is

$$\breve{\mu}_2(d\boldsymbol{y}) = \big[\nu_2(\mathbb{E})\big]^n d\boldsymbol{y}. \tag{2.9}$$

♢

2.5 Partitions and Delta-Functions

In the next definition, we introduce *ordered partitions* and three different ways to represent them: As a family $\mathscr{P}$ of sets; as a function φ; and as a vector $\boldsymbol{p}$.

2.24 Definition (Ordered partitions). The following three representations of a partition of the set $\{1,\ldots,n\}$ will be used interchangeably. The ordering of the sets is part of the definition, hence they are called *ordered* partitions.

(*i*) *Family of sets.* If $\mathscr{P}$ is a partition of $\{1,\ldots,n\}$, then there exists an $m \geq 1$, such that one has the representation:

$$\mathscr{P} = \{S_1,\ldots,S_m\}, \text{ where each } S_r \subseteq \{1,\ldots,n\}.$$

(*ii*) *Index function.* Let $\mathscr{P}$ be a partition as in the first part. Define for $i = 1,\ldots,n$ the function φ by:

$$\varphi(i) := r, \text{ where } r \text{ is such that } i \in S_r.$$

In other words, φ assigns element i to the set $S_{\varphi(i)}$. Hence, $\mathscr{P}$ can equivalently be represented as a function of the form:

$$\varphi : \{1,\ldots,n\} \to \{1,\ldots,m\}.$$

If φ is given, then the notation $\mathscr{P}_\varphi$ is used to denote this partition.

(*iii*) *Index vector.* Let $\mathscr{P}$ be a partition and φ its representation as an index function. Define the associated index vector

$$\boldsymbol{p} := \Big(\varphi(1),\ldots,\varphi(n)\Big) \in \{0,\ldots,m\}^n.$$

In other words, $p_i = r$ if and only if $i \in S_r$, for all $i = 1,\ldots,n$. If the vector $\boldsymbol{p}$ is given, the associated partition is denoted by $\mathscr{P}_{\boldsymbol{p}}$. ♢

In case that either φ or $\boldsymbol{p}$ are used to define a partition, we call $\mathscr{P}_\varphi$ or $\mathscr{P}_{\boldsymbol{p}}$ the *induced partition*. We add some more useful notation related to partitions:

2.25 Notation (Partitions). The following notation refers always to ordered partitions:

(1) *Size of a partition.* The number of sets of a partition is called its *size*. Hence, for a partition $\mathscr{P} = \{S_1, \ldots, S_m\}$, one writes $|\mathscr{P}| = m$.

(2) *Family of all partitions.* Let $\mathfrak{P}^n$ denote all ordered partitions of $\{1, \ldots, n\}$ and let $\mathfrak{P}^n_m$ denote all ordered partitions of size m of $\{1, \ldots, n\}$, i.e.:

$$\mathfrak{P}^n_m := \Big\{ \mathscr{P} \in \mathfrak{P}^n : |\mathscr{P}| = m \Big\}.$$

(3) *Refinements.* For two partitions $\mathscr{P}, \mathscr{Q} \in \mathfrak{P}^n$ we say $\mathscr{Q}$ is a *refinement* of $\mathscr{P}$ if all sets of $\mathscr{P}$ can be written as a union of sets in $\mathscr{Q}$. In this case, the notation $\mathscr{Q} \ll \mathscr{P}$ is used. ♢

The following remark focuses on the index function φ but is true also for an index vector $\boldsymbol{p}$:

2.26 Remark (Induced Partition). Take an index function φ and consider the induced ordered partition $\mathscr{P}_\varphi$. Clearly, for general φ, the mapping

$$\varphi : \{1, \ldots, n\} \to \{1, \ldots, m\}.$$

does not need to be surjective. As a consequence, the size $\tilde{m} := |\mathscr{P}_\varphi|$ can be strictly smaller than m. If it happens that $\tilde{m} < m$, one can always choose an equivalent index function $\tilde{\varphi}$, i.e. an index function $\tilde{\varphi}$ with $\mathscr{P}_{\tilde{\varphi}} = \mathscr{P}_\varphi$, such that

$$\varphi : \{1, \ldots, n\} \to \{1, \ldots, \tilde{m}\}.$$

Hence, one can assume without loss of generality that $\tilde{m} = m$. ♢

Next we introduce some notation that helps to deal with sub-vectors and other manipulations of the components of vectors. We will use this notation for all vector-like objects for which it makes sense to refer to single components.

2.27 Notation (Index notation for vectors). Let $\mathbb{E}_k$, for $1 \le k \le d$, be a sequence of univariate event spaces. Consider the product space

$$\mathbb{E}_{j_1} \times \ldots \times \mathbb{E}_{j_n}, \text{ where } j_k \in \{1, \ldots, d\}, \text{ for } 1 \le k \le n.$$

Take an element $\boldsymbol{y}$ from this product space, i.e.

$$\boldsymbol{y} = (y_1, \ldots, y_n), \text{ where } y_k \in \mathbb{E}_{j_k}, \text{ for } 1 \le k \le n,$$

and introduce the following notation:

(*i*) Let $S \subseteq \{1, \ldots, n\}$ be given and assume that:

$$y_i = y_j, \text{ for all } i, j \in S.$$

In terms of delta functions, see Definition 2.28, this condition is equivalent to $\delta^{(|S|)}(\boldsymbol{y}_S) = 1$. Now choose an arbitrary (which one is irrelevant) $i \in S$ and define:

$$y_S := y_i.$$

(*ii*) Let $S \subseteq \{1, \ldots, n\}$, where $S = \{i_1, \ldots, i_m\}$, for some $m \geq 1$, be given. Then define the vector:

$$\boldsymbol{y}_S := (y_{i_1}, \ldots, y_{i_m}).$$

(*iii*) Let $\mathscr{P} \in \mathfrak{P}_m^n$ be given and assume that:

$$y_i = y_j, \text{ for all } i, j \in S_r, \text{ all } 1 \leq r \leq m.$$

In terms of delta-functions, this condition is equivalent to $\delta^{(|S_r|)}(\boldsymbol{y}_{S_r}) = 1$, all $1 \leq r \leq m$. Then define the vector:

$$\boldsymbol{y}_{\mathscr{P}} := (y_{S_1}, \ldots, y_{S_m}).$$

(*iv*) Let $\boldsymbol{j} \in \{1, \ldots, d\}^m$ be given and define the vector:

$$\boldsymbol{y}^{(\boldsymbol{j})} := (y_{j_1}, \ldots, y_{j_m}). \qquad \diamondsuit$$

It should be emphasized that y_S and $\boldsymbol{y}_{\mathscr{P}}$ are only well-defined in a specific context, i.e. only if $\boldsymbol{y}$ satisfies the corresponding condition. Also note that in the definition of $\boldsymbol{y}_S$, one always has $i_j \neq i_k$, for $j \neq k$. But in the definition of $\boldsymbol{y}^{(\boldsymbol{j})}$ it can happen that $j_k = j_l$, for $j \neq k$. Moreover, y_S is a scalar but $\boldsymbol{y}_S$ is a vector.

For the next definition recall Notation 2.27:

2.28 Definition (Delta functions). Let $\mathbb{E}$ be a univariate and $\mathbb{E}$ be a multivariate event space.

(1a) *Univariate ordinary delta function.* For $\boldsymbol{x} \in \mathbb{E}^n$ define:

$$\delta^{(n)}(\boldsymbol{x}) = \prod_{k,l} \mathbb{1}_{\{x_k = x_l\}} = \begin{cases} 1 & \text{if } x_k = x_l \text{ for all } 1 \le k, l \le n, \\ 0 & \text{otherwise.} \end{cases}$$

(1b) *Multivariate ordinary delta function.* For $\boldsymbol{j} \in \{1, \dots, d\}^n$ and $\boldsymbol{y} \in \mathbb{E}^{(\boldsymbol{j})}$ define:

$$\delta^{(n)}(\boldsymbol{j}, \boldsymbol{y}) := \prod_{k,l} \mathbb{1}_{\{j_k = j_l, y_k = y_l\}} = \delta^{(n)}(\boldsymbol{j}) \delta^{(n)}(\boldsymbol{y}).$$

(2a) *Univariate factorial delta function.* For $\boldsymbol{x} \in \mathbb{E}^n$ define:

$$\delta^{[n]}(\boldsymbol{x}) = \prod_{k \ne l} \mathbb{1}_{\{x_k \ne x_l\}} = \begin{cases} 1 & \text{if } x_k \ne x_l \text{ for all } 1 \le k < l \le n, \\ 0 & \text{otherwise.} \end{cases}$$

(2b) *Multivariate factorial delta function.* For $\boldsymbol{j} \in \{1, \dots, d\}^n$ and $\boldsymbol{y} \in \mathbb{E}^{(\boldsymbol{j})}$ define:

$$\delta^{[n]}(\boldsymbol{j}, \boldsymbol{y}) := \prod_{k \ne l} \mathbb{1}_{\{(j_k, y_k) \ne (j_l, y_l)\}} = 1 - \delta^{(n)}(\boldsymbol{j}, \boldsymbol{y}). \qquad \diamondsuit$$

As an extension of the above delta-functions, we consider the case where we have given an ordered partition and the components are only compared on the sets of this partition:

2.29 Definition (Delta functions linked to partitions). Let $\mathbb{E}$ be a univariate and $\mathbb{E}$ be a multivariate event space.

(1) *Ordinary partition delta function.* For a partition $\mathscr{P} \in \mathfrak{P}_m^n$ define:

$$\delta^{(\mathscr{P})}(\boldsymbol{x}) := \prod_{r=1}^{m} \delta^{(|S_r|)}(\boldsymbol{x}_{S_r}).$$

For $\boldsymbol{j} \in \{1, \dots, d\}^n$, a vector $\boldsymbol{y} \in \mathbb{E}^{(\boldsymbol{j})}$ and a partition $\mathscr{P} \in \mathfrak{P}_m^n$ define:

$$\begin{aligned} \delta^{(\mathscr{P})}(\boldsymbol{j}, \boldsymbol{y}) &= \prod_{r=1}^{m} \delta^{(|S_r|)}(\boldsymbol{j}_{S_r}, \boldsymbol{y}_{S_r}) \\ &= \begin{cases} 1 & \text{if } j_k = j_l,\ y_k = y_l, \text{ for all } k, l \in S_r, \text{ for all } r, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(2) *Factorial partition delta function.* For a partition $\mathscr{P} \in \mathfrak{P}_m^n$ define:

$$\delta^{[\mathscr{P}]}(\boldsymbol{x}) := \delta^{[m]}(\boldsymbol{x}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{x}) = \delta^{[m]}(x_{S_1}, \ldots, x_{S_m}) \prod_{r=1}^{m} \delta^{(|S_r|)}(\boldsymbol{x}_{S_r})$$

For $\boldsymbol{j} \in \{1, \ldots, d\}^n$, a vector $\boldsymbol{y} \in \mathbb{E}^{(\boldsymbol{j})}$ and a partition $\mathscr{P} \in \mathfrak{P}_m^n$ define:

$$\begin{aligned}
\delta^{[\mathscr{P}]}(\boldsymbol{j}, \boldsymbol{y}) &= \delta^{[m]}(\boldsymbol{j}_{\mathscr{P}}, \boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{j}, \boldsymbol{y}) \\
&= \delta^{[m]}\Big((j_{S_1}, y_{S_1}), \ldots, (j_{S_m}, y_{S_m})\Big) \prod_{r=1}^{m} \Big[\delta^{(|S_r|)}(\boldsymbol{j}_{S_r})\delta^{(|S_r|)}(\boldsymbol{y}_{S_r})\Big] \\
&= \begin{cases} 1 & \begin{cases} j_k = j_l \text{ and } y_k = y_l, \text{ for } k, l \in S_r, \text{ for all } r, \\ j_k \neq j_l \text{ or } y_k \neq y_l, \text{ for } k \in S_r,\ l \in S_s, \text{ for all } r \neq s, \end{cases} \\ 0 & \text{otherwise.} \end{cases} \quad \diamondsuit
\end{aligned}$$

2.30 Remark (Interpretation). The delta-functions given in the above definition can be interpreted as follows:

(1) *Ordinary partition delta function.* $\delta^{(\mathscr{P})}(\boldsymbol{x})$ indicates whether all component values x_k are identical in each of the groups S_r of the partition. $\delta^{(\mathscr{P})}(\boldsymbol{j}, \boldsymbol{y})$ indicates whether all component indexes j_k and the component values y_k are identical in each of the groups of the partition.

(2) *Factorial partition delta function.* $\delta^{[\mathscr{P}]}(\boldsymbol{x})$ indicates whether in each group S_r all component values x_k are identical and component values from two different groups are distinct. $\delta^{[\mathscr{P}]}(\boldsymbol{j}, \boldsymbol{y})$ indicates whether in each group S_r the component indexes j_k and component values x_k are identical and components from two different groups are distinct. $\diamondsuit$

There is a convenient way to formulate the fact that a partition $\mathscr{Q}$ is a refinement of another partition $\mathscr{P}$:

2.31 Remark (Refinements and index vectors). Let $\mathscr{P}, \mathscr{Q} \in \mathfrak{P}^n$ be two ordered partitions and assume $\boldsymbol{p} \in \{1, \ldots, n\}^n$ is the vector representation of $\mathscr{P}$, i.e. $\mathscr{P} = \mathscr{P}_{\boldsymbol{p}}$. Then:

$$\mathscr{Q} \ll \mathscr{P}_{\boldsymbol{p}} \qquad \Longleftrightarrow \qquad \delta^{(\mathscr{Q})}(\boldsymbol{p}) = 1. \tag{2.10}$$

$\diamondsuit$

The following theorem is the main result of this section. It is the reason why delta-functions have been introduced, in the first place.

2.32 Theorem (Delta-function decomposition). Take a univariate event space $\mathbb{E}$ and a multivariate event space $\mathbb{E}$.

(1) *Univariate decomposition.* For all $\boldsymbol{x} \in \mathbb{E}^n$:

$$1 = \sum_{\mathscr{P} \in \mathfrak{P}^n} \delta^{[\mathscr{P}]}(\boldsymbol{x}) = \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}^n_m} \delta^{[\mathscr{P}]}(\boldsymbol{x}). \tag{2.11}$$

(2) *Multivariate decomposition.* For all $\boldsymbol{j} \in \{1, \ldots, d\}^n$ and $\boldsymbol{y} \in \mathbb{E}^{(\boldsymbol{j})}$:

$$1 = \sum_{\mathscr{P} \in \mathfrak{P}^n} \delta^{[\mathscr{P}]}(\boldsymbol{j}, \boldsymbol{y}) = \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \delta^{[\mathscr{P}]}(\boldsymbol{j}, \boldsymbol{y}). \tag{2.12}$$

♢

As soon as one deals with higher order moment measures of point processes, factorial product measures are often easier to handle than ordinary product measures. Formulas of the kind given above allow one to decompose ordinary moment measures into factorial moment measures. This will be explained in more detail in the next chapter.

2.6 Point Configurations

The next definition is borrowed from Section 6.1 in [DVJ03]. The natural topology on the point configuration space will be introduced in Chapter 3. As long as only fixed, i.e. deterministic point configurations are considered, the topology is not relevant, but it will become important as soon as one considers stochastic point configurations, i.e. point processes.

The general case of multivariate, marked point configurations is introduced stepwise. Note that throughout, we exclusively consider *simple* point configurations and do not bother to introduce notation for general point configurations. We start with the case of univariate, unmarked point configurations and extend the definition stepwise:

2.33 Definition (Point configuration). Let $\mathbb{E}$, $\mathbb{E}$ be event spaces and $\mathbb{X}$, $\mathbb{X}$ be a mark spaces, see Definitions 2.2 and 2.3.

(1a) *Univariate, unmarked point configurations.* Let $\mathscr{N}(\mathbb{E})$ denote the family of all locally-finite, simple counting measures on $\mathbb{E}$. A counting measure ν is *simple* if and only if:

$$\nu(\{u\}) \in \{0, 1\}, \text{ for all } u \in \mathbb{E}.$$

(1b) *Univariate, marked point configurations.* Let $\mathscr{N}(\mathbb{E} \times \mathbb{X})$ be all simple point configurations on $\mathbb{E} \times \mathbb{X}$. Then define:

$$\mathscr{N}_{\mathbb{X}}(\mathbb{E}) := \Big\{ \nu \in \mathscr{N}(\mathbb{E} \times \mathbb{X}) : \nu(du \times \mathbb{X}) \in \mathscr{N}(\mathbb{E}) \Big\}.$$

The notation $\nu(du \times \mathbb{X})$ refers to the projection of ν onto its first component, and is called the *ground process*. Hence, a point configuration from $\mathscr{N}_{\mathbb{X}}(\mathbb{E})$ has a locally-finite, simple ground process. In consequence, $\mathscr{N}_{\mathbb{X}}(\mathbb{E})$ is in general a strict subset of $\mathscr{N}(\mathbb{E} \times \mathbb{X})$.

(2) *Multivariate point configurations.* A multivariate point configuration is a vector of univariate point configurations. Define the space

$$\mathscr{N}_{\mathbb{X}}(\mathbb{E}) := \Big\{ \big(\nu^{(1)}, \ldots, \nu^{(d)}\big) : \nu^{(j)} \in \mathscr{N}_{\mathbb{X}_j}(\mathbb{E}_j), j \in \{1, \ldots, d\} \Big\}. \qquad \diamondsuit$$

We will always use the definition above for the space of simple, multivariate point configurations. But one could also give the following stricter definition of simple, multivariate point configurations:

2.34 Remark (Common event space). Consider the same situation as in the multivariate part of Definition 2.33. But this time, assume that not only the ground processes $\nu^{(j)}(du \times \mathbb{X}_j)$ are simple, but also the joint ground process $\sum_{j=1}^{d} \nu^{(j)}(du \times \mathbb{X}_j)$. This leads to the definition:

$$\mathscr{N}_{\mathbb{X}}(\mathbb{E}) := \Big\{ \nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{E}) : \sum_{j=1}^{d} \nu^{(j)}(du \times \mathbb{X}_j) \in \mathscr{N}(\mathbb{E}) \Big\}.$$

Note that $\mathscr{N}_{\mathbb{X}}(\mathbb{E})$ is in general a true subset of $\mathscr{N}_{\mathbb{X}}(\mathbb{E})$. $\diamondsuit$

It might not be obvious why there are two versions of multivariate point configurations. To clarify this, consider the following example:

2.35 Remark (Weak and strong simplicity). Assume all event spaces coincide and assume for simplicity that $\mathbb{E}_j := \mathbb{R}$, for all $j \in \{1, \ldots, d\}$. One should interpret $\mathbb{R}$ as the time axis. The difference between $\mathscr{N}_{\mathbb{X}}(\mathbb{E})$ and $\mathscr{N}_{\mathbb{X}}(\mathbb{E})$ can now be explained as follows:

(i) *Separate event spaces.* Take some $\nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{E})$ and fix a time $t \in \mathbb{R}$. Since each component $\nu^{(j)}$ is a simple point configuration, $\nu^{(j)}$ has at most one point at time t. But it is not excluded that several components have an event at time t. This shows that although ν consists of simple point

configurations, if one considers all components simultaneously, ν is not any more simple.

(*ii*) *Common event spaces.* This time take some $\nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{E})$ and fix again a time $t \in \mathbb{R}$. As before, each component $\nu^{(j)}$ has at most one point at time t. But this time, at most one component can have a point at time t. This means that if one considers all components simultaneously, ν is still simple.

One could call point configurations from these two spaces also simple in a *weak* sense and simple in a *strong* sense. ♢

2.36 Definition (Product measures). In the univariate case let $\nu \in \mathscr{N}(\mathbb{E})$ and in the multivariate case let $\nu \in \mathscr{N}(\mathbb{E})$.

(*1*) *Univariate product measure.* For $n \geq 1$ define the *ordinary* product measure $\nu^{(n)}$ and the *factorial* product measure $\nu^{[n]}$ by:

$$\nu^{(n)}(d\boldsymbol{x}) := \prod_{k=1}^{n} \nu(dx_k), \qquad \nu^{[n]}(d\boldsymbol{x}) := \delta^{[n]}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}).$$

Note that $\nu^{(n)}, \nu^{[n]} \in \mathscr{N}(\mathbb{E}^n)$.

(*2*) *Multivariate product measure.* Let $n \geq 1$ and $\boldsymbol{j} \in \{1, \ldots, d\}^n$. Define the *ordinary* product measure $\nu^{(\boldsymbol{j})}$ and the *factorial* product measure $\nu^{[\boldsymbol{j}]}$ by:

$$\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) := \prod_{k=1}^{n} \nu^{(j_k)}(dy_k), \qquad \nu^{[\boldsymbol{j}]}(d\boldsymbol{y}) := \delta^{[n]}(\boldsymbol{j}, \boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}).$$

Note that $\nu^{(\boldsymbol{j})}, \nu^{[\boldsymbol{j}]} \in \mathscr{N}(\mathbb{E}^{(\boldsymbol{j})})$. ♢

Under certain conditions, a higher order factorial product measure can be written as a product of lower order factorial product measures. This is the content of the next proposition:

2.37 Proposition (Separation of product measures). Let $\mathbb{E}$ be a univariate and $\mathbb{E} \equiv (\mathbb{E}_1, \ldots, \mathbb{E}_d)$ be a multivariate event space. Take a sequence of integers $n_1, \ldots, n_d \geq 1$ and define $n := \sum_{k=1}^{d} n_k$.

(*1*) Take a sequence of sets of the form:

$$A_{1,1}, \ldots, A_{1,n_1} \quad \ldots \quad A_{d,1}, \ldots, A_{d,n_d},$$

where $A_{k,i} \subseteq \mathbb{E}$, all $1 \leq i \leq n_k$, $1 \leq k \leq d$. Furthermore, assume

$$A_{k,i} \cap B_{l,j} = \emptyset, \text{ for all } 1 \leq i \leq n_k,\ 1 \leq j \leq n_l, \text{ if } k \neq l.$$

Then, for a point configuration $\nu \in \mathscr{N}(\mathbb{E})$, one has:

$$\nu^{[n]}\Big(\bigtimes_{k=1}^{d}\bigtimes_{i=1}^{n_k} A_{k,i}\Big) = \prod_{k=1}^{d} \nu^{[n_k]}\Big(\bigtimes_{i=1}^{n_k} A_{k,i}\Big). \tag{2.13}$$

(2) Define for all $1 \leq k \leq d$ the index vectors

$$\boldsymbol{j}_k := (k, \ldots, k), \text{ where } \boldsymbol{j}_k \in \{1, \ldots, d\}^{n_k}.$$

Let $\boldsymbol{j} := (\boldsymbol{j}_1 \cdots \boldsymbol{j}_d)$ denote the concatenation of the sequence $\boldsymbol{j}_{k=1,\ldots,d}$. Then, for a point configuration $\nu \in \mathscr{N}(\mathbb{E}^{(\boldsymbol{j})})$, one has:

$$\nu^{[\boldsymbol{j}]}(d\boldsymbol{y}) = \nu^{[\boldsymbol{j}_1 \cdots \boldsymbol{j}_d]}(d\boldsymbol{y}_1 \times \ldots \times d\boldsymbol{y}_d) = \prod_{k=1}^{d} \nu^{[\boldsymbol{j}_k]}(d\boldsymbol{y}_k). \tag{2.14}$$

♢

It is possible to enforce the condition on the index vector $\boldsymbol{j}$ given in the second part above with a suitable delta-function. This leads to the following equivalent way to write the second part above:

2.38 Remark (Special case of separation formula). Assume the index vector $\boldsymbol{j}$ is of the form $\boldsymbol{j} = (j_1, \ldots, j_n)$, for scalars $j_k \in \{1, \ldots, d\}$. Then:

$$\delta^{[n]}(\boldsymbol{j})\nu^{[\boldsymbol{j}]}(d\boldsymbol{y}) = \delta^{[n]}(\boldsymbol{j})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}).$$

♢

2.39 Theorem (Complete decay of product measures). Let $\mathbb{E}$ be a univariate and $\mathbb{E} \equiv (\mathbb{E}_1, \ldots, \mathbb{E}_d)$ be a multivariate event space. Fix an integer $n \geq 1$.

(1) *Univariate case.* For a function g on $\mathbb{E}$ one has:

$$\int_{\mathbb{E}^n} g(x_i)\delta^{(n)}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) = \int_{\mathbb{E}} g(x)\nu(dx), \tag{2.15}$$

where $1 \leq i \leq n$ is arbitrary.

(2) *Multivariate case.* For $\boldsymbol{j} \in \{1, \dots, d\}^n$ and a function g on $\mathbb{E}_j$ one has:

$$\int_{\mathbb{E}^{(\boldsymbol{j})}} g(y_i)\delta^{(n)}(\boldsymbol{j}, \boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) = \delta^{(n)}(\boldsymbol{j}) \int_{\mathbb{E}_j} g(y)\nu^{(j)}(dy), \tag{2.16}$$

where $1 \le i \le n$ is arbitrary and $j := j_i$. ♢

Note that $\delta^{(n)}(\boldsymbol{j}) = 1$ if and only if $\boldsymbol{j}$ is of the form $\boldsymbol{j} = (j, \dots, j)$. Indeed, one could have simply assumed that $\boldsymbol{j} = (j, \dots, j)$ in the second part of the above theorem. But the way it is formulated, the formula can be used in all other cases too, and this without case distinction.

Sometimes the formulas in the above theorem appear under a slightly different disguise. Since this might not always be obvious, we state these alternative versions, too:

2.40 Remark. Let $\mathbb{E}$ be a univariate and $\mathbb{E} \equiv (\mathbb{E}_1, \dots, \mathbb{E}_d)$ be a multivariate event space and fix an integer $n \ge 1$. For a function f on $\mathbb{E}^n$ one has:

$$\int_{\mathbb{E}^n} f(\boldsymbol{x})\delta^{(n)}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) = \int_{\mathbb{E}} f(x, \dots, x)\nu(dx).$$

For $\boldsymbol{j} \in \{1, \dots, d\}^n$ and a function g on $\mathbb{E}_j^n$ one has:

$$\int_{\mathbb{E}^{(\boldsymbol{j})}} f(\boldsymbol{x})\delta^{(n)}(\boldsymbol{j}, \boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) = \delta^{(n)}(\boldsymbol{j}) \int_{\mathbb{E}_j} f(x, \dots, x)\nu^{(j)}(dx). \qquad ♢$$

2.41 Corollary (Partial decay of product measures). In the univariate case, let ν be a point configuration on $\mathbb{E}$, $\mathscr{P} \in \mathfrak{P}_m^n$ and $\boldsymbol{p}$ be the vector representation of $\mathscr{P}$.

In the multivariate case, let $\boldsymbol{j} \in \{1, \dots, d\}^n$, $\mathscr{P}_{\boldsymbol{j}}$ be the associated partition, $\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}$ be a refinement, $\boldsymbol{p}$ be the vector representation of $\mathscr{P}$ and ν be a point configuration on $\mathbb{E}$.

(1) *First version, univariate case.* For a function g on $\mathbb{E}^m$ one has:

$$\begin{aligned} \int_{\mathbb{E}^n} g(\boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}) &= \int_{\mathbb{E}^m} g(\boldsymbol{x})\nu^{(m)}(d\boldsymbol{x}), \\ \int_{\mathbb{E}^n} g(\boldsymbol{y}_{\mathscr{P}})\delta^{[\mathscr{P}]}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}) &= \int_{\mathbb{E}^m} g(\boldsymbol{x})\nu^{[m]}(d\boldsymbol{x}). \end{aligned} \tag{2.17}$$

(2) *First version, multivariate case.* For a function g on $\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}$ one has:

$$\begin{aligned}\int_{\mathbb{E}^{(\boldsymbol{j})}} g(\boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) &= \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} g(\boldsymbol{x})\nu^{(\boldsymbol{j}_{\mathscr{P}})}(d\boldsymbol{x}),\\ \int_{\mathbb{E}^{(\boldsymbol{j})}} g(\boldsymbol{y}_{\mathscr{P}})\delta^{[\mathscr{P}]}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) &= \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} g(\boldsymbol{x})\nu^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{x}).\end{aligned} \tag{2.18}$$

(3) *Second version, univariate case.* For a function f on $\mathbb{E}^n$ one has:

$$\begin{aligned}\int_{\mathbb{E}^n} f(\boldsymbol{y})\delta^{(\mathscr{P})}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}) &= \int_{\mathbb{E}^m} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{(m)}(d\boldsymbol{x}),\\ \int_{\mathbb{E}^n} f(\boldsymbol{y})\delta^{[\mathscr{P}]}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}) &= \int_{\mathbb{E}^m} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{[m]}(d\boldsymbol{x}).\end{aligned} \tag{2.19}$$

(4) *Second version, multivariate case.* For a function f on $\mathbb{E}^{(\boldsymbol{j})}$ one has:

$$\begin{aligned}\int_{\mathbb{E}^{(\boldsymbol{j})}} f(\boldsymbol{y})\delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) &= \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{(\boldsymbol{j}_{\mathscr{P}})}(d\boldsymbol{x}),\\ \int_{\mathbb{E}^{(\boldsymbol{j})}} f(\boldsymbol{y})\delta^{[\mathscr{P}]}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) &= \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{x}).\end{aligned} \tag{2.20}$$

♢

The above formulas are used to prove the next theorem. This theorem is motivated by Section 5.4 in [DVJ03], especially by the subsection *Exercises and Complements*, but as formulated below, it is more general:

2.42 Theorem (Factorial decomposition). In the univariate case let $\nu \in \mathscr{N}(\mathbb{E})$ and in the multivariate case let $\nu \in \mathscr{N}(\mathbb{E})$. Fix an integer $n \geq 1$.

(1) *Factorial decomposition.* In the univariate case, let f be a function on $\mathbb{E}^n$. In the multivariate case, let $\boldsymbol{j} \in \{1,\dots,d\}^n$ and f be a function on $\mathbb{E}^{(\boldsymbol{j})}$. Then:

$$\begin{aligned}\int_{\mathbb{E}^n} f(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) &= \sum_{m=1}^{n}\sum_{\mathscr{P}\in\mathfrak{P}_m^n}\int_{\mathbb{E}^m} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{[m]}(d\boldsymbol{x}),\\ \int_{\mathbb{E}^{(\boldsymbol{j})}} f(\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) &= \sum_{\mathscr{P}\ll\mathscr{P}_{\boldsymbol{j}}}\int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} f(\boldsymbol{y}^{(\boldsymbol{p})})\nu^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{y}).\end{aligned} \tag{2.21}$$

(2) *Symmetric, factorial decomposition.* In the univariate case, let f_k be a sequence of functions on $\mathbb{E}$, for $1 \leq k \leq n$. In the multivariate case, let

$\boldsymbol{j} \in \{1, \ldots, d\}^n$ and f_k be a sequence of functions on $\mathbb{E}_{j_k}$. Then:

$$\int_{\mathbb{E}^n} \Big[\prod_{k=1}^{n} f_k(x_k)\Big] \nu^{(n)}(d\boldsymbol{x}) = \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{m} \Big[\prod_{i \in S_r} f_i(x_r)\Big] \nu^{[m]}(d\boldsymbol{x}),$$

$$\int_{\mathbb{E}^{(\boldsymbol{j})}} \Big[\prod_{k=1}^{n} f_k(y_k)\Big] \nu^{(\boldsymbol{j})}(d\boldsymbol{y}) = \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} \prod_{r=1}^{|\mathscr{P}|} \Big[\prod_{i \in S_r} f_i(y_r)\Big] \nu^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{y}). \tag{2.22}$$

♢

In the next corollary, we consider two special cases of the above decomposition formulas: First we evaluate the n-fold product measure on a set of the form $A_1 \times \ldots \times A_n$; and then we take a closer look at the special case $n = 2$.

2.43 Corollary (Common special cases). Let $\mathbb{E}$ be a univariate and $\mathbb{E} \equiv (\mathbb{E}_1, \ldots, \mathbb{E}_d)$ be a multivariate event space.

(1) *Factorial decomposition for sets.* Fix an integer $n \geq 1$. In the univariate case, let $A_k \subseteq \mathbb{E}$, for $1 \leq k \leq n$. In the multivariate case, let $\boldsymbol{j} \in \{1, \ldots, d\}^n$ and $A_k \subseteq \mathbb{E}_{j_k}$, for $1 \leq k \leq n$. Then:

$$\nu^{(n)}(A_1 \times \ldots \times A_n) = \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \nu^{[m]}\Big(\bigcap_{i \in S_1} A_i \times \ldots \times \bigcap_{i \in S_m} A_i\Big),$$

$$\nu^{(\boldsymbol{j})}(A_1 \times \ldots \times A_n) = \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n, \mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \nu^{[\boldsymbol{j}_{\mathscr{P}}]}\Big(\bigcap_{i \in S_1} A_i \times \ldots \times \bigcap_{i \in S_m} A_i\Big).$$

(2) *Two dimensional special case.* In the univariate case, let f be a function on $\mathbb{E}^2$. In the multivariate case, let $1 \leq i, j \leq d$ and f be a function on $\mathbb{E}^{(ij)}$. Then:

$$\int_{\mathbb{E}^2} f(\boldsymbol{y}) \nu^{(2)}(d\boldsymbol{y}) = \int_{\mathbb{E}} f(\boldsymbol{y}) \nu^{[2]}(d\boldsymbol{y}) + \int_{\mathbb{E}} f(y, y) \nu(dy),$$

$$\int_{\mathbb{E}^{(ij)}} f(\boldsymbol{y}) \nu^{(ij)}(d\boldsymbol{y}) = \int_{\mathbb{E}^{(ij)}} f(\boldsymbol{y}) \nu^{[ij]}(d\boldsymbol{y}) + \delta_{ij} \int_{\mathbb{E}^{(i)}} f(y, y) \nu^{(i)}(dy). \quad \diamondsuit$$

Proofs for Chapter 2

Let us state two obvious properties of the linear transformations $\sigma^{\langle k\rangle}$ and $\tau^{\langle k\rangle}$:

2.44 Lemma (Isometric transformations). For all $1 \leq k \leq n$ one has that:

(*1*) *Invertibility.* $\sigma^{\langle k\rangle}$ and $\tau^{\langle k\rangle}$ are inverse functions of each other, i.e.:

$$\tau^{\langle k\rangle} \circ \sigma^{\langle k\rangle} = \sigma^{\langle k\rangle} \circ \tau^{\langle k\rangle} = \mathrm{id}_{\mathbb{E}^n}\,.$$

(*2*) *Isometry.* For the determinants of the Jacobi matrices of $\sigma^{\langle k\rangle}$ and $\tau^{\langle k\rangle}$ one has:

$$\det\Big(\frac{\sigma^{\langle k\rangle}(\boldsymbol{y}, z)}{\partial(\boldsymbol{y}, z)}\Big) = \det\Big(\frac{\tau^{\langle k\rangle}(\boldsymbol{x})}{\partial \boldsymbol{x}}\Big) = 1. \qquad \diamondsuit$$

The above lemma shows that $\sigma^{\langle k\rangle}$ and $\tau^{\langle k\rangle}$ are space preserving transformations of $\mathbb{E}^n$. This fact will be used for parameter substitutions in integrals over $\mathbb{E}^n$.

Proof (Lemma 2.44): Fix some $1 \leq k \leq n$.

(*1a*) *Invertibility of* σ. For $\boldsymbol{y} \in \mathbb{E}^{n-1}$ and $z \in \mathbb{E}$ one has:

$$\begin{aligned}\big[\tau^{\langle k\rangle} \circ \sigma^{\langle k\rangle}\big](\boldsymbol{y}, z) &= \tau^{\langle k\rangle}\big(\mathrm{I}^{\langle k\rangle}(\boldsymbol{y}) + z\mathbf{1}_n\big) = \tau^{\langle k\rangle}\big(\mathrm{I}^{\langle k\rangle}(\boldsymbol{y})\big) + \tau^{\langle k\rangle}\big(z\mathbf{1}_n\big)\\ &= \Big(\big[\mathrm{I}^{\langle k\rangle}(\boldsymbol{y})\big]^{\langle k\rangle} - \big[\mathrm{I}^{\langle k\rangle}(\boldsymbol{y})\big]_k \mathbf{1}_{n-1}, \big[\mathrm{I}^{\langle k\rangle}(\boldsymbol{y})\big]_k\Big)\\ &\qquad + \Big([z\mathbf{1}_n]^{\langle k\rangle} - [z\mathbf{1}_n]_k \mathbf{1}_{n-1}, [z\mathbf{1}_n]_k\Big)\\ &= (\boldsymbol{y}, 0) + (z\mathbf{1}_{n-1} - z\mathbf{1}_{n-1}, z) = (\boldsymbol{y}, z).\end{aligned}$$

(*1b*) *Invertibility of* τ. Let $\boldsymbol{e}_k$ denote the k-th unit vector in $\mathbb{E}^n$. Note that

$$\mathrm{I}^{\langle k\rangle}(\boldsymbol{x}^{\langle k\rangle}) = \boldsymbol{x} - x_k \boldsymbol{e}_k, \text{ as well as } \mathrm{I}^{\langle k\rangle}(x_k \mathbf{1}_{n-1}) = x_k \mathbf{1}_n - x_k \boldsymbol{e}_k.$$

Therefore, $\tau^{\langle k\rangle}$ is invertible, since:

$$\begin{aligned}\big[\sigma^{\langle k\rangle}\circ\tau^{\langle k\rangle}\big](\boldsymbol{x}) &= \sigma^{\langle k\rangle}\big(\boldsymbol{x}^{\langle k\rangle} - x_k\mathbf{1}_{n-1}, x_k\big)\\ &= \sigma^{\langle k\rangle}(\boldsymbol{x}^{\langle k\rangle}, 0) - \sigma^{\langle k\rangle}(x_k\mathbf{1}_{n-1}, 0) + \sigma^{\langle k\rangle}(0, x_k)\\ &= \mathrm{I}^{\langle k\rangle}(\boldsymbol{x}^{\langle k\rangle}) - \mathrm{I}^{\langle k\rangle}(x_k\mathbf{1}_{n-1}) + x_k\mathbf{1}_n\\ &= (\boldsymbol{x} - x_k\boldsymbol{e}_k) - (x_k\mathbf{1}_n - x_k\boldsymbol{e}_k) + x_k\mathbf{1}_n = \boldsymbol{x}.\end{aligned}$$

(2) *Isometry.* The statement is only shown for $\tau^{\langle n\rangle}$. By symmetry, the result then also holds for all other $1 \leq k \leq n$. First note that

$$\begin{aligned}\tau^{\langle n\rangle}(\boldsymbol{x}) &= (\boldsymbol{x}^{\langle n\rangle} - x_n\mathbf{1}_{n-1}, x_n)\\ &= (x_1 - x_n, x_2 - x_n, \ldots, x_{n-1} - x_n, x_n).\end{aligned}$$

Therefore, the Jacobi matrix of $\tau^{\langle n\rangle}$ and its determinant are:

$$\frac{\partial\tau^{\langle n\rangle}(\boldsymbol{x})}{\partial\boldsymbol{x}} = \begin{pmatrix} 1 & & & -1\\ & \ddots & & \vdots\\ & & 1 & -1\\ & & & 1\end{pmatrix}, \qquad \det\Big(\frac{\partial\tau^{\langle n\rangle}(\boldsymbol{x})}{\partial\boldsymbol{x}}\Big) = 1.$$

The statement for $\sigma^{\langle n\rangle}$ follows, since $\sigma^{\langle n\rangle}$ is the inverse of $\tau^{\langle n\rangle}$. □

Proof (Theorem 2.14): The following fact is used below: Since $\sigma^{\langle k\rangle}$ and $\tau^{\langle k\rangle}$ are volume preserving bijections, for any two functions f, g on $\mathbb{E}^n$ one has:

$$\begin{aligned}\int_{\mathbb{E}^n} f(\boldsymbol{x})d\boldsymbol{x} &= \int_{\mathbb{E}^n}\big[f\circ\sigma^{\langle k\rangle}\big](\boldsymbol{y}, z)d\boldsymbol{y}\,dz,\\ \int_{\mathbb{E}^n} g(\boldsymbol{y}, z)d\boldsymbol{y}\,dz &= \int_{\mathbb{E}^n}\big[g\circ\tau^{\langle k\rangle}\big](\boldsymbol{x})d\boldsymbol{x}.\end{aligned}$$

Now fix some $1 \leq k \leq n$.

(1) *First transformation formula.* Take $A \subseteq \mathbb{E}^{n-1}$, $B \subseteq \mathbb{E}$ and let f be of the form:

$$f(\boldsymbol{x}) := \mathbb{1}_A(\boldsymbol{x}^{\langle k\rangle})\mathbb{1}_B(x_k).$$

Due to $\big[\mathrm{I}^{\langle k\rangle}(\boldsymbol{y})\big]^{\langle k\rangle} = \boldsymbol{y}$ and $\big[\mathrm{I}^{\langle k\rangle}(\boldsymbol{y})\big]_k = 0$, one obtains:

$$\begin{aligned}\big[f \circ \sigma^{\langle k\rangle}\big](\boldsymbol{y}, z) &= f\Big(\mathrm{I}^{\langle k\rangle}(\boldsymbol{y}) + z\mathbf{1}_n\Big) \\ &= \mathbb{1}_A\Big(\big[\mathrm{I}^{\langle k\rangle}(\boldsymbol{y}) + z\mathbf{1}_n\big]^{\langle k\rangle}\Big)1_B\Big(\big[\mathrm{I}^{\langle k\rangle}(\boldsymbol{y}) + z\mathbf{1}_n\big]_k\Big) \\ &= \mathbb{1}_A(\boldsymbol{y} + z\mathbf{1}_{n-1})\,\mathbb{1}_B(z).\end{aligned}$$

Now start with the right hand side of Equation (2.2):

$$\begin{aligned}&\int_{\mathbb{E}}\int_{\mathbb{E}^{n-1}} [f \circ \sigma^{\langle k\rangle}](\boldsymbol{y}, z)\breve{\mu}^{\langle k\rangle}(d\boldsymbol{y})dz \\ &\quad= \int_{\mathbb{E}} \mathbb{1}_B(z)\Big[\int_{\mathbb{E}^{n-1}} \mathbb{1}_A(\boldsymbol{y} + z\mathbf{1}_{n-1})\breve{\mu}^{\langle k\rangle}(d\boldsymbol{y})\Big]dz \\ &\quad= \int_{\mathbb{E}} \mathbb{1}_B(z)\big[\breve{\mu}^{\langle k\rangle}(A - z\mathbf{1}_{n-1})\big]dz.\end{aligned}$$

Note that $\{A - z\mathbf{1}_{n-1}\}$ denotes the set A shifted by $(-z\mathbf{1}_{n-1})$. Substituting the definition of $\breve{\mu}^{\langle k\rangle}$ yields:

$$\begin{aligned}(\ldots) &= \int_{\mathbb{E}} \mathbb{1}_B(z)\Big[\int_{\mathbb{E}^n} \big[\mathbb{1}_{(A - z\mathbf{1}_{n-1})\times\mathbb{D}} \circ \tau^{\langle k\rangle}\big](\boldsymbol{x})\mu(d\boldsymbol{x})\Big]dz \\ &= \int_{\mathbb{E}} \mathbb{1}_B(z)\Big[\int_{\mathbb{E}^n} \mathbb{1}_{A\times\mathbb{D}}\big(\boldsymbol{x}^{\langle k\rangle} - x_k\mathbf{1}_{n-1} + z\mathbf{1}_{n-1}, x_k\big)\mu(d\boldsymbol{x})\Big]dz \\ &= \int_{\mathbb{E}^n} \mathbb{1}_{\mathbb{D}}(x_k)\Big[\int_{\mathbb{E}} \mathbb{1}_A\big(\boldsymbol{x}^{\langle k\rangle} - x_k\mathbf{1}_{n-1} + z\mathbf{1}_{n-1}\big)\,\mathbb{1}_B(z)dz\Big]\mu(d\boldsymbol{x}).\end{aligned}$$

After substituting $z \twoheadrightarrow (z + x_k)$ in the inner integral, one gets:

$$\begin{aligned}(\ldots) &= \int_{\mathbb{E}^n} \mathbb{1}_{\mathbb{D}}(x_k)\Big[\int_{\mathbb{E}} \mathbb{1}_A\big(\boldsymbol{x}^{\langle k\rangle} + z\mathbf{1}_{n-1}\big)\,\mathbb{1}_B(x_k + z)dz\Big]\mu(d\boldsymbol{x}) \\ &= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}^n} \mathbb{1}_A\big(\boldsymbol{x}^{\langle k\rangle} + z\mathbf{1}_{n-1}\big)\,\mathbb{1}_B(x_k + z)\,\mathbb{1}_{\mathbb{D}}(x_k)\mu(d\boldsymbol{x})\Big]dz.\end{aligned}$$

Since μ is translation-invariant, the measure $\mu(d\boldsymbol{x})$ and its shifted version $\mu(d\boldsymbol{x} + \mathbf{1}_n z)$ are identical. Hence, one may substitute $\boldsymbol{x} \twoheadrightarrow (\boldsymbol{x} - \mathbf{1}_n z)$ and

then obtains:

$$\begin{aligned}(\ldots) &= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}^n} \mathbb{1}_A(\boldsymbol{x}^{\langle k \rangle}) \mathbb{1}_B(x_k) \mathbb{1}_{\mathbb{D}}(x_k - z) \mu(d\boldsymbol{x}) \Big] dz \\ &= \int_{\mathbb{E}^n} \mathbb{1}_A(\boldsymbol{x}^{\langle k \rangle}) \mathbb{1}_B(x_k) \Big[\int_{\mathbb{E}} \mathbb{1}_{\mathbb{D}}(x_k - z) dz \Big] \mu(d\boldsymbol{x}) \\ &= \int_{\mathbb{E}^n} \mathbb{1}_A(\boldsymbol{x}^{\langle k \rangle}) \mathbb{1}_B(x_k) \mu(d\boldsymbol{x}) = \int_{\mathbb{E}^n} f(\boldsymbol{x}) \mu(d\boldsymbol{x}).\end{aligned}$$

In the one but last equality the fact that $\mathbb{D}$ is a unit cube has been used. The statement for general functions f now follows by approximation with elementary functions and monotone convergence.

(*ii*) *Second transformation formula.* The second formula follows immediately from the first one. To this end, define $f(\boldsymbol{x}) := \big[g \circ \tau^{\langle k \rangle}\big](\boldsymbol{x})$. Now substitute this function into the first transformation formula and use that

$$\big[f \circ \sigma^{\langle k \rangle}\big](\boldsymbol{x}) = \big[g \circ \tau^{\langle k \rangle} \circ \sigma^{\langle k \rangle}\big](\boldsymbol{x}) = g(\boldsymbol{x}). \qquad \square$$

Proof (Corollary 2.15): Let $\mathbb{D} \subseteq \mathbb{E}$ be a third set with $\lambda_{\mathbb{E}}(\mathbb{D}) = 1$. For arbitrary $B \subseteq \mathbb{E}^{n-1}$ define:

$$g(\boldsymbol{y}, z) := \mathbb{1}_B(\boldsymbol{y}) \mathbb{1}_{\mathbb{D}}(z),$$

where $\boldsymbol{y} \in \mathbb{E}^{n-1}$ and $z \in \mathbb{E}$. Due to Equation (2.3), one has:

$$\begin{aligned}\breve{\mu}_1^{\langle k \rangle}(B) = \breve{\mu}_1^{\langle k \rangle}(B) \int_{\mathbb{E}} \mathbb{1}_{\mathbb{D}}(z) dz &= \int_{\mathbb{E}} \mathbb{1}_{\mathbb{D}}(z) \Big[\int_{\mathbb{E}^{n-1}} \mathbb{1}_B(\boldsymbol{y}) \breve{\mu}_1^{\langle k \rangle}(d\boldsymbol{y}) \Big] dz \\ &= \int_{\mathbb{E}} \int_{\mathbb{E}^{n-1}} g(\boldsymbol{y}, z) \breve{\mu}_1^{\langle k \rangle}(d\boldsymbol{y}) = \int_{\mathbb{E}^n} [g \circ \tau^{\langle k \rangle}](\boldsymbol{x}) \mu(d\boldsymbol{x}).\end{aligned}$$

Note that the right hand side is independent of the choice of $\mathbb{D}_1$. Hence, the same result is obtained if the calculation is made for $\breve{\mu}_2{}^{\langle k \rangle}$ instead of $\breve{\mu}_1{}^{\langle k \rangle}$. Therefore, as claimed:

$$\breve{\mu}_1^{\langle k \rangle}(B) = \int_{\mathbb{E}^n} [g \circ \tau^{\langle k \rangle}](\boldsymbol{x}) \mu(d\boldsymbol{x}) = \breve{\mu}_2^{\langle k \rangle}(B). \qquad \square$$

Proof (Proposition 2.16): Let $\mathbb{E}$ be an event space and fix some $n \geq 1$.

(*i*) *Linearity.* The linearity of the reduction operation follows from the linearity of $\sigma^{\langle k \rangle}$ and $\tau^{\langle k \rangle}$.

(2) *Anti-symmetry in two dimensions.* By assumption, μ is invariant under simultaneous shifts of all components. A transformation of this type is given by the following mapping from $\mathbb{E}^2$ onto itself:

$$x_1 \mapsto x_1 + (x_2 - x_1) = x_2, \qquad x_2 \mapsto x_2 + (x_2 - x_1) = 2x_2 - x_1.$$

Therefore, for all functions f on $\mathbb{E}^2$:

$$\int_{\mathbb{E}^2} f(x_1, x_2)\mu(dx_1 \times dx_2) = \int_{\mathbb{E}^2} f(x_2, 2x_2 - x_1)\mu(dx_1 \times dx_2).$$

Applying this formula for $f(x_1, x_2) := \mathbb{1}_{B\times\mathbb{D}}(x_2 - x_1, x_1)$ gives:

$$\begin{aligned}
\breve{\mu}^{\langle 1\rangle}(B) &= \int_{\mathbb{D}} \mu(dx_1 \times B + x_1) = \int_{\mathbb{E}^2} \mathbb{1}_{B\times\mathbb{D}}(x_2 - x_1, x_1)\mu(d\boldsymbol{x}) \\
&= \int_{\mathbb{E}^2} \mathbb{1}_{B\times\mathbb{D}}\big((2x_2 - x_1) - x_2, x_2\big)\mu(d\boldsymbol{x}) \\
&= \int_{\mathbb{E}^2} \mathbb{1}_{\{-B\}\times\mathbb{D}}(x_1 - x_2, x_2)\mu(d\boldsymbol{x}) \\
&= \int_{\mathbb{D}} \mu(-B + x_2 \times dx_2) = \breve{\mu}^{\langle 2\rangle}(-B),
\end{aligned}$$

for all sets $B \subseteq \mathbb{E}$, and this proves the claim.

(3a) *Density of reduced measure.* Take some $A \subseteq \mathbb{E}^{n-1}$. We need to show that the following expression is equal to $\breve{\mu}(A)$:

$$\begin{aligned}
&\int_{\mathbb{E}^{n-1}} \mathbb{1}_A(\boldsymbol{y})\breve{m}^{\langle k\rangle}(\boldsymbol{y})d\boldsymbol{y} \\
&\quad= \int_{\mathbb{E}^{n-1}} \mathbb{1}_A(\boldsymbol{y})\Big[\int_{\mathbb{E}} \mathbb{1}_{\mathbb{D}}(z)[m \circ \sigma^{\langle k\rangle}](\boldsymbol{y}, z)dz\Big]d\boldsymbol{y} \\
&\quad= \int_{\mathbb{E}^{n-1}\times\mathbb{E}} \mathbb{1}_{A\times\mathbb{D}}(\boldsymbol{y}, z)[m \circ \sigma^{\langle k\rangle}](\boldsymbol{y}, z)d\boldsymbol{y}\, dz.
\end{aligned}$$

Now apply the mass-preserving substitution $\boldsymbol{x} \twoheadrightarrow \tau^{\langle k\rangle}(\boldsymbol{x}) \equiv (\boldsymbol{y}, z)$ and use that $\sigma^{\langle k\rangle} \circ \tau^{\langle k\rangle} = \mathrm{id}_{\mathbb{E}^n}$. Hence:

$$(\ldots) = \int_{\mathbb{E}^n} \big[\mathbb{1}_{A\times\mathbb{D}} \circ \tau^{\langle k\rangle}\big](\boldsymbol{x})m(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{E}^n} \big[\mathbb{1}_{A\times\mathbb{D}} \circ \tau^{\langle k\rangle}\big](\boldsymbol{x})\mu(d\boldsymbol{x})$$

Next apply Equation (2.3) for $g(\boldsymbol{y}, z) := \mathbb{1}_{A\times\mathbb{D}}(\boldsymbol{y}, z)$. This leads to

$$(\dots) = \int_{\mathbb{E}} \int_{\mathbb{E}^{n-1}} \mathbb{1}_{A\times\mathbb{D}}(\boldsymbol{y}, z) \breve{\mu}^{\langle k\rangle}(d\boldsymbol{y}) dz = \int_{\mathbb{E}^{n-1}} \mathbb{1}_A(\boldsymbol{y}) \breve{\mu}^{\langle k\rangle}(d\boldsymbol{y}).$$

And this shows that $\breve{m}^{\langle k\rangle}$ is indeed the density of the measure $\breve{\mu}^{\langle k\rangle}$.

(3b) *Relation to sheared density.* Repeat the previous calculation, but take now a general set $B \subseteq \mathbb{E}$ instead of $\mathbb{D}$. This leads to

$$\int_{\mathbb{E}^n} \mathbb{1}_{A\times B}(\boldsymbol{y}, z)[m \circ \sigma^{\langle k\rangle}](\boldsymbol{y}, z) d\boldsymbol{y}\, dz = \int_{\mathbb{E}^n} \mathbb{1}_{A\times B}(\boldsymbol{y}, z) \breve{\mu}^{\langle k\rangle}(d\boldsymbol{y}) dz.$$

Therefore $(m \circ \sigma^{\langle k\rangle})$ is a density for the measure $\breve{\mu}^{\langle k\rangle}(d\boldsymbol{y})dz$. But since $\breve{m}^{\langle k\rangle}$ is also a density, it must hold that:

$$[m \circ \sigma^{\langle k\rangle}](\boldsymbol{y}, z) = \breve{m}^{\langle k\rangle}(\boldsymbol{y}), \text{ for almost all } (\boldsymbol{y}, z) \in \mathbb{E}^{n-1} \times \mathbb{E},$$

with respect to the Lebesgue measure. □

Proof (Proposition 2.17): Consider the substitution that switches the positions of x_k and x_l of a vector $\boldsymbol{x} \in \mathbb{E}^n$, i.e. the substitution given by

$$\big(x_1, \dots, x_k, \dots, x_l, \dots, x_n\big) \twoheadrightarrow \big(x_1, \dots, x_l, \dots, x_k, \dots, x_n\big),$$

for any two $1 \le k, l \le d$. Note that if one switches the positions of x_k and x_l in the expression $\tau^{\langle k\rangle}(\boldsymbol{x})$, this leads to the change $\tau^{\langle k\rangle}(\boldsymbol{x}) \twoheadrightarrow \tau^{\langle l\rangle}(\boldsymbol{x})$. Now take some set $B \subseteq \mathbb{E}^{n-1}$. Since μ is a symmetric measure on $\mathbb{E}^n$, it is invariant with respect to this substitution, and this shows:

$$\begin{aligned}\breve{\mu}^{\langle k\rangle}(B) &= \int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle k\rangle}\big](\boldsymbol{x}) \mu(d\boldsymbol{x}) \\ &= \int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle l\rangle}\big](\boldsymbol{x}) \mu(d\boldsymbol{x}) = \breve{\mu}^{\langle l\rangle}(B).\end{aligned}$$ □

Proof (Remark 2.18): As all reduced measures are identical, one may choose without loss of generality the one with index $k := n$.

(i) The first statement is a reformulation of the definition of the n-th reduced measure.

(ii) We need to check that $\breve{m}$ as defined in Remark 2.18 is the density of $\breve{\mu}$.

Let $B \subseteq \mathbb{E}^{n-1}$ and first note that:

$$\begin{aligned}\breve{\mu}(B) &= \int_{\mathbb{D}} \mu(B + x_n \times dx_n) = \int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}\big](\boldsymbol{x})\mu(d\boldsymbol{x}) \\ &= \int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}\big](\boldsymbol{x})m(\boldsymbol{x})d\boldsymbol{x}.\end{aligned}$$

After applying the measure-preserving substitution $(\boldsymbol{y}, z) \twoheadrightarrow \sigma^{\langle n\rangle}(\boldsymbol{y}, z)$, one obtains:

$$\begin{aligned}&\int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}\big](\boldsymbol{x})m(\boldsymbol{x})d\boldsymbol{x} \\ &\quad= \int_{\mathbb{E}^n} \mathbb{1}_{B\times\mathbb{D}}(\boldsymbol{y}, z)\big[m \circ \sigma^{\langle n\rangle}\big](\boldsymbol{y}, z)d\boldsymbol{y}dz \\ &\quad= \int_{\mathbb{E}^n} \mathbb{1}_{B\times\mathbb{D}}(\boldsymbol{y}, z)m(\boldsymbol{y} + z, z)d\boldsymbol{y}dz \\ &\quad= \int_{\mathbb{E}^{n-1}} \mathbb{1}_B(\boldsymbol{y})\Big[\int_{\mathbb{D}} m(\boldsymbol{y} + z, z)dz\Big]d\boldsymbol{y} = \int_B \breve{m}(\boldsymbol{y})d\boldsymbol{y}.\end{aligned}$$

And this shows that $\breve{m}$ is indeed the intensity of $\breve{\mu}$. □

Proof (Corollary 2.19): Since all reduced measures coincide, one may consider without loss of generality the n-th reduced measure. The two transformation formulas for the symmetric case are then simply reformulations of Equations (2.2) and (2.3). □

Proof (Proposition 2.21): Note that in both cases, the measures are symmetric and therefore one may take without loss of generality the n-th reduced measure.

(1) *Reduced Lebesgue measure.* Since $(\boldsymbol{y}, x) \mapsto \sigma^{\langle n\rangle}(\boldsymbol{y}, x)$ is mass-preserving, one has for $B \subseteq \mathbb{E}^{n-1}$ that:

$$\begin{aligned}\breve{\lambda}_{\mathbb{E}^n}(B) &= \int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}\big](\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{E}} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}\big]\big(\sigma^{\langle n\rangle}(\boldsymbol{y}, z)\big)d\boldsymbol{y}dz \\ &= \int_{\mathbb{E}^n} \mathbb{1}_{B\times\mathbb{D}}(\boldsymbol{y}, z)d\boldsymbol{y}dz = \int_{\mathbb{E}^{n-1}} \mathbb{1}_B(\boldsymbol{y})d\boldsymbol{y} = \lambda_{\mathbb{E}^{n-1}}(B).\end{aligned}$$

(2) *Reduced Lebesgue measure on the diagonal.* First note that for a function

f on $\mathbb{E}^n$ one has:

$$\begin{aligned}\int_{\mathbb{E}^n} f(\boldsymbol{x})\mu_2(d\boldsymbol{x}) &= \int_{\mathbb{E}^n} f(\boldsymbol{x})\Big[\int_{\mathbb{E}} \delta_{(z,\ldots,z)}(d\boldsymbol{x})dz\Big] \\ &= \int_{\mathbb{E}^n\times\mathbb{E}} f(\boldsymbol{x})\delta_{(z,\ldots,z)}(d\boldsymbol{x})dz \\ &= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}^n} f(\boldsymbol{x})\delta_{(z,\ldots,z)}(d\boldsymbol{x})\Big]dz = \int_{\mathbb{E}} f(z,\ldots,z)dz.\end{aligned}$$

Now choose $f(\boldsymbol{x}) := [\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}](\boldsymbol{x})$. Since $\tau^{\langle n\rangle}(z,\ldots,z) = (\mathbf{0}_{n-1}, z)$, where $\mathbf{0}_{n-1} \in \mathbb{E}^{n-1}$ is the zero-vector, the claim follows from:

$$\begin{aligned}\breve{\mu}(B) &= \int_{\mathbb{D}} \mu(B + x_n \times dx_n) = \int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}\big](\boldsymbol{x})\mu(d\boldsymbol{x}) \\ &= \int_{\mathbb{E}} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}\big](z,\ldots,z)dz = \int_{\mathbb{E}} \mathbb{1}_{B\times\mathbb{D}}(\mathbf{0}_{n-1}, z)dz \\ &= \mathbb{1}_B(\mathbf{0}_{n-1})\int_{\mathbb{E}} \mathbb{1}_{\mathbb{D}}(z)dz = \mathbb{1}_B(\mathbf{0}_{n-1}) = \delta_{\mathbf{0}}^{n-1}(B).\end{aligned}$$ □

Proof (Proposition 2.23): Recall the definitions of the regular reduction given in Definition 2.13 and of the pseudo-reduction given in Definition 2.22.

(1) *Multivariate underlying measure.* For all $\boldsymbol{x} \in \mathbb{E}^n$, let $\boldsymbol{x} \equiv (\check{\boldsymbol{x}}, x_n)$ denote the concatenation of the two components $\check{\boldsymbol{x}} \in \mathbb{E}^{n-1}$ and $x_n \in \mathbb{E}$. Then, for a set $B \subseteq \mathbb{E}^{n-1}$ one has:

$$\begin{aligned}\breve{\mu}_1(B) &= \int_{\mathbb{E}^n} \big[\mathbb{1}_{B\times\mathbb{D}} \circ \tau^{\langle n\rangle}\big](\boldsymbol{x})\mu_1(d\boldsymbol{x}) \\ &= \int_{\mathbb{E}^n} \mathbb{1}_{B\times\mathbb{D}}(\check{\boldsymbol{x}} - x_n, x_n)\mu_1(d\check{\boldsymbol{x}} \times dx_n) \\ &= \int_{\mathbb{E}^n} \mathbb{1}_{B\times\mathbb{D}}(\check{\boldsymbol{x}} - x_n, x_n)\Big[\int_{\mathbb{E}} \nu_1\big([d\check{\boldsymbol{x}} - w] \times [dx_n - w]\big)dw\Big] \\ &= \int_{\mathbb{E}^{n+1}} \mathbb{1}_{B\times\mathbb{D}}(\check{\boldsymbol{x}} - x_n, x_n)\nu_1\big([d\check{\boldsymbol{x}} - w] \times [dx_n - w]\big)dw.\end{aligned}$$

Now substitute $\check{\boldsymbol{x}} \twoheadrightarrow \check{\boldsymbol{x}} + w$ and $x_n \twoheadrightarrow x_n + w$, such that:

$$\begin{aligned}(\ldots) &= \int_{\mathbb{E}^{n+1}} \mathbb{1}_{B\times\mathbb{D}}(\check{\boldsymbol{x}} - x_n, x_n + w)\nu_1(d\check{\boldsymbol{x}} \times dx_n)dw \\ &= \int_{\mathbb{E}^n} \mathbb{1}_B(\check{\boldsymbol{x}} - x_n)\Big[\int_{\mathbb{E}} \mathbb{1}_{\mathbb{D}}(x_n + w)dw\Big]\nu_1(d\check{\boldsymbol{x}} \times dx_n) \\ &= \int_{\mathbb{E}^n} \mathbb{1}_B(\check{\boldsymbol{x}} - x_n)\nu_1(d\check{\boldsymbol{x}} \times dx_n) = \mathring{\nu}_1(B).\end{aligned}$$

(2) *Product of univariate underlying measures.* We first show that μ_2 is a multiple of the Lebesgue measure $\lambda_{\mathbb{E}^n}$ on $\mathbb{E}^n$: For a sequence of sets $A_k \subseteq \mathbb{E}$, where $1 \leq k \leq n$, one has:

$$\begin{aligned}\mu_2\Big(\bigtimes_{k=1}^n A_k\Big) &= \prod_{k=1}^n \int_{\mathbb{E}} \nu_2(A_k - w_k)dw_k = \prod_{k=1}^n \big[\nu_2(\mathbb{E})\lambda_{\mathbb{E}}(A_k)\big] \\ &= \big[\nu_2(\mathbb{E})\big]^n \lambda_{\mathbb{E}^n}\Big(\bigtimes_{k=1}^n A_k\Big).\end{aligned}$$

Therefore $\mu_2(d\boldsymbol{x}) = \big[\nu_2(\mathbb{E})\big]^n d\boldsymbol{x}$. But the reduction of the Lebesgue measure on $\mathbb{E}^n$ is given in Equation (2.6). Due to the linearity of the reduction operation, one obtains as claimed:

$$\breve{\mu}_2(d\boldsymbol{y}) = \big[\nu_2(\mathbb{E})\big]^n d\boldsymbol{y}. \qquad \square$$

Proof (Theorem 2.32): Recall that the factorial δ-functions are defined by:

$$\delta^{[\mathscr{P}]}(\boldsymbol{x}) = \delta^{[m]}(\boldsymbol{x}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{x}), \qquad \delta^{[\mathscr{P}]}(\boldsymbol{j},\boldsymbol{y}) = \delta^{[m]}(\boldsymbol{j}_{\mathscr{P}},\boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y}),$$

for $\mathscr{P} \in \mathfrak{P}^n_m$, see also Definition 2.29.

(1) *Univariate decomposition.* Take some $\mathscr{P} \in \mathfrak{P}^n$ with size $|\mathscr{P}| = m$. Clearly, $\delta^{[\mathscr{P}]}(\boldsymbol{x}) = 1$ if and only if both $\delta^{[m]}(\boldsymbol{x}_{\mathscr{P}}) = 1$ and $\delta^{(\mathscr{P})}(\boldsymbol{x}) = 1$. Recall that for $\boldsymbol{x} \in \mathbb{E}^n$ one has:

$$\begin{aligned}\delta^{[m]}(\boldsymbol{x}_{\mathscr{P}}) = 1 \qquad &\Longleftrightarrow \qquad x_k = x_l, \text{ all } k,l \in S_r, \\ \delta^{(\mathscr{P})}(\boldsymbol{x}) = 1 \qquad &\Longleftrightarrow \qquad x_k \neq x_l, \text{ all } k \in S_p, l \in S_q,\end{aligned}$$

for all $1 \leq r \leq m$ and all $1 \leq p \neq q \leq m$. This shows that $\delta^{[\mathscr{P}]}(\boldsymbol{x}) = 1$ if and only if the components of $\boldsymbol{x}$ coincide for indexes that lie in the same set S_r and differ for indexes that lie in different sets S_p, S_q, for $p \neq q$.

For given $\boldsymbol{x} \in \mathbb{E}^n$, there is exactly one partition, say $\mathscr{Q} \in \mathfrak{P}^n$, which groups together all components of $\boldsymbol{x}$ with the same values. Hence, one can find a partition $\mathscr{Q}$ such that:

$$\delta^{[\mathscr{Q}]}(\boldsymbol{x}) = 1 \qquad \text{and} \qquad \delta^{[\mathscr{P}]}(\boldsymbol{x}) = 0, \text{ for all } \mathscr{P} \in \mathfrak{P}^n : \mathscr{P} \neq \mathscr{Q}.$$

Therefore, the sum over all partitions is one, as claimed.

(2) *Multivariate decomposition.* The first equality is shown in the same way as in the univariate case. It remains to check the second equality:

The second equality follows if one can show that $\delta^{[\mathscr{P}]}(\boldsymbol{j}, \boldsymbol{y}) = 0$, for all $\mathscr{P} \in \mathfrak{P}^n$ which are not refinements of $\mathscr{P}_{\boldsymbol{j}}$. But due to Equation (2.10), if $\mathscr{P}$ is not a refinement of $\mathscr{P}_{\boldsymbol{j}}$, then $\delta^{(\mathscr{P})}(\boldsymbol{j}) = 0$, such that:

$$\begin{aligned}\delta^{[\mathscr{P}]}(\boldsymbol{j}, \boldsymbol{y}) &= \delta^{[m]}(\boldsymbol{j}_{\mathscr{P}}, \boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{j}, \boldsymbol{y}) \\ &= \delta^{[m]}(\boldsymbol{j}_{\mathscr{P}}, \boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{j})\delta^{(\mathscr{P})}(\boldsymbol{y}) = 0,\end{aligned}$$

where $m := |\mathscr{P}|$. □

Proof (Proposition 2.37): By definition, the sequence of sets $(A_{k,i})$ consists of the d groups

$$A_{1,1}, \ldots, A_{1,n_1} \quad \ldots \quad A_{d,1}, \ldots, A_{d,n_d}.$$

In the same way as the sequence $(A_{k,i})$ has been grouped, one can split a vector $\boldsymbol{x} \in \mathbb{E}^n$ up in the form:

$$\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d), \text{ where } \boldsymbol{x}_k \in \mathbb{E}^{n_k}, \text{ for } 1 \leq k \leq d.$$

In the multivariate case, one can find an analogue decomposition of a vector $\boldsymbol{y} \in \mathbb{E}^{(\boldsymbol{j})}$ that parallels the decomposition of $\boldsymbol{j}$:

$$\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_d), \text{ where } \boldsymbol{y}_k \in \mathbb{E}^{(\boldsymbol{j}_k)}, \text{ for } 1 \leq k \leq d.$$

(1) First note that

$$\nu^{[n]}\Big(\bigtimes_{k=1}^{d}\bigtimes_{i=1}^{n_k} A_{k,i}\Big) = \int_{\mathbb{E}^n}\Big[\prod_{k=1}^{d}\prod_{i=1}^{n_k}\mathbb{1}_{A_{k,i}}(x_{k,i})\Big]\nu^{[n]}(d\boldsymbol{x})$$
$$= \int_{\mathbb{E}^n}\Big[\prod_{k=1}^{d}\prod_{i=1}^{n_k}\mathbb{1}_{A_{k,i}}(x_{k,i})\Big]\delta^{[n]}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}). \tag{2.23}$$

Recall that $\delta^{[n]}(\boldsymbol{x}) = 1$ if and only if all components of $\boldsymbol{x}$ are distinct. In terms of the sub-vectors $\boldsymbol{x}_k$, one can rewrite the delta-function as:

$$\delta^{[n]}(\boldsymbol{x}) = \Big[\prod_{k\neq l}^{d}\prod_{i,j}^{(n_k,n_l)}\delta^{[2]}(x_{k,i},x_{l,j})\Big]\Big[\prod_{k=1}^{d}\prod_{i\neq j}\delta^{[2]}(x_{k,i,k,j})\Big].$$

For convenience, define:

$$\Xi(\boldsymbol{x}) := \prod_{k=1}^{d}\prod_{i=1}^{n_k}\mathbb{1}_{A_{k,i}}(x_{k,i}).$$

Now assume $\boldsymbol{x}\in\mathbb{E}^n$ is such that $\Xi(\boldsymbol{x}) = 1$. In other words, $x_{k,i}\in A_{k,i}$, for all k,i. Take $k\neq l$ and since $A_{k,i}\cap A_{l,j} = \emptyset$, it must hold that $x_{k,i}\neq x_{l,j}$, and this means that $\delta^{[2]}(x_{k,i},x_{l,j}) = 1$. Therefore, for $\boldsymbol{x}$ with $\Xi(\boldsymbol{x}) = 1$, one finds

$$\delta^{[n]}(\boldsymbol{x}) = \prod_{k=1}^{d}\prod_{i\neq j}\delta^{[2]}(x_{k,i,k,j}) = \prod_{k=1}^{d}\delta^{[n_k]}(\boldsymbol{x}_k).$$

It is now clear that the following equation holds for all $\boldsymbol{x}\in\mathbb{E}^n$, even if $\boldsymbol{x}$ is such that $\Xi(\boldsymbol{x}) = 0$:

$$\Xi(\boldsymbol{x})\delta^{[n]}(\boldsymbol{x}) = \Xi(\boldsymbol{x})\prod_{k=1}^{d}\delta^{[n_k]}(\boldsymbol{x}_k).$$

Substituting this in Equation (2.23) gives:

$$\begin{aligned}\nu^{[n]}\Big(\bigtimes_{k=1}^{d}\bigtimes_{i=1}^{n_k} A_{k,i}\Big) &= \int_{\mathbb{E}^n} \Xi(\boldsymbol{x})\delta^{[n]}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) \\ &= \int_{\mathbb{E}^n} \Xi(\boldsymbol{x})\Big[\prod_{k=1}^{d}\delta^{[n_k]}(\boldsymbol{x}_k)\Big]\nu^{(n)}(d\boldsymbol{x}) \\ &= \prod_{k=1}^{d}\int_{\mathbb{E}^{n_k}}\Big[\prod_{i=1}^{n_k}\mathbb{1}_{A_{k,i}}(x_{k,i})\Big]\delta^{[n_k]}(\boldsymbol{x}_k)\nu^{(n_k)}(d\boldsymbol{x}_k) \\ &= \prod_{k=1}^{d}\nu^{[n_k]}\Big(\bigtimes_{i=1}^{n_k}A_{k,i}\Big).\end{aligned}$$

(2) According to the definition of $\delta^{[n]}(\boldsymbol{j},\boldsymbol{y})$, and since $\boldsymbol{j}_k = (k,\ldots,k)$, one has:

$$\delta^{[n]}(\boldsymbol{j},\boldsymbol{y}) = \delta^{[n]}\Big((\boldsymbol{j}_1,\ldots,\boldsymbol{j}_d),(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_d)\Big) = \prod_{k=1}^{d}\delta^{[n_k]}(\boldsymbol{j}_k,\boldsymbol{y}_k).$$

The statement now follows with:

$$\begin{aligned}\nu^{[\boldsymbol{j}]}(d\boldsymbol{y}) = \delta^{[n]}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) &= \Big[\prod_{k=1}^{d}\delta^{[n_k]}(\boldsymbol{j}_k,\boldsymbol{y}_k)\Big]\Big[\prod_{k=1}^{d}\nu^{(\boldsymbol{j}_k)}(d\boldsymbol{y}_k)\Big] \\ &= \prod_{k=1}^{d}\Big[\delta^{[n_k]}(\boldsymbol{j}_k,\boldsymbol{y}_k)\nu^{(\boldsymbol{j}_k)}(d\boldsymbol{y}_k)\Big] = \prod_{k=1}^{d}\nu^{[\boldsymbol{j}_k]}(d\boldsymbol{y}_k).\end{aligned}$$

□

2.45 Lemma (Equivalent formulations of simplicity). Let $\mathbb{E}$ be an event space and ν a point configuration on $\mathbb{E}$ which does not need to be simple, i.e. ν not necessarily lies in the space $\mathscr{N}(\mathbb{E})$. Then the following statements are equivalent:

(1) $\nu \in \mathscr{N}(\mathbb{E})$, i.e. ν is simple.

(2) For all functions f on $\mathbb{E}$:

$$\int_{\mathbb{E}} f(x)\nu(dx) = \int_{\mathbb{E}} f(x)\nu(\{x\})\nu(dx).$$

(3) For all $n \geq 1$ and $1 \leq i \leq n$ and all functions g on $\mathbb{E}$:

$$\int_{\mathbb{E}^n} g(x_i)\delta^{(n)}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) = \int_{\mathbb{E}} g(x)\nu(dx). \tag{2.24}$$

♢

Proof (Lemma 2.45): Let $\chi_{\mathbb{E}}$ denote the counting measure on $\mathbb{E}$, that is let $\chi_{\mathbb{E}}(A) := \mathrm{card}(A)$ be the number of elements in $A \subseteq \mathbb{E}$.

(1→2) Note first that for any point measure ν on $\mathbb{E}$:

$$\nu(dx) = \nu(\{x\})\chi_{\mathbb{E}}(dx).$$

Since ν is simple by assumption, one has $\nu(\{x\}) \in \{0, 1\}$, for all $x \in \mathbb{E}$, which is equivalent to $\nu(\{x\}) = \big[\nu(\{x\})\big]^2$. Hence, for a function f on $\mathbb{E}$:

$$\begin{aligned}\int_{\mathbb{E}} f(x)\nu(dx) &= \int_{\mathbb{E}} f(x)\nu(\{x\})\chi_{\mathbb{E}}(dx) = \int_{\mathbb{E}} f(x)\big[\nu(\{x\})\big]^2\chi_{\mathbb{E}}(dx) \\ &= \int_{\mathbb{E}} f(x)\nu(\{x\})\nu(dx).\end{aligned}$$

(2→3) Prove by induction: First note that the statement is trivial for $n = 1$. Now assume the statement has already been shown for some $n \geq 1$. Use the notation $\boldsymbol{x} \equiv (\check{\boldsymbol{x}}, x_n)$, for $\boldsymbol{x} \in \mathbb{E}^n$ and $\check{\boldsymbol{x}} \in \mathbb{E}^{n-1}$, $x_n \in \mathbb{E}$. Then, for all $1 \leq i \leq n$, one has:

$$\begin{aligned}&\int_{\mathbb{E}^{n+1}} g(x_i)\delta^{(n+1)}(\boldsymbol{x})\nu^{(n+1)}(d\boldsymbol{x}) \\ &\quad= \int_{\mathbb{E}^n\times\mathbb{E}} g(\check{x}_i)\delta^{(n+1)}(\check{\boldsymbol{x}}, x_{n+1})\nu^{(n)}(d\check{\boldsymbol{x}})\nu(dx_{n+1}) \\ &\quad= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}^n} g(\check{x}_i)\delta^{(2)}(x_{n+1}, \check{x}_i)\delta^{(n)}(\check{\boldsymbol{x}})\nu^{(n)}(d\check{\boldsymbol{x}})\Big]\nu(dx_{n+1}).\end{aligned}$$

Now with the induction assumption applied to $g(\check{x}_i)\delta^{(2)}(x_{n+1}, \check{x}_i)$, one gets:

$$\begin{aligned}(\ldots) &= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}} g(z)\delta^{(2)}(x_{n+1}, z)\nu(dz)\Big]\nu(dx_{n+1}) \\ &= \int_{\mathbb{E}} g(x_{n+1})\nu(\{x_{n+1}\})\nu(dx_{n+1}) = \int_{\mathbb{E}} g(x_{n+1})\nu(dx_{n+1}).\end{aligned}$$

(3→1) Fix some $z \in \mathbb{E}$ and define $g(x) := \delta^{(2)}(z, x)$. Then, due to the assump-

tion applied for the case $n = 2$, one can see that:

$$\begin{aligned}\nu(\{z\}) &= \int_{\mathbb{E}} \delta^{(2)}(z,x)\nu(dx) = \int_{\mathbb{E}^2} \delta^{(2)}(z,x_1)\delta^{(2)}(x_1,x_2)\nu^{(2)}(d\boldsymbol{x}) \\ &= \int_{\mathbb{E}^2} \mathbb{1}_{\{(z,z)\}}(x_1,x_2)\nu^{(2)}(d\boldsymbol{x}) = \nu^{(2)}\big(\{(z,z)\}\big) = \big[\nu(\{z\})\big]^2.\end{aligned}$$

This shows that $\nu(\{z\}) \in \{0,1\}$, i.e. ν is simple. □

Proof (Theorem 2.39): The univariate formula is Equation (2.24), so there is nothing to prove. It remains to show the multivariate case:

Assume without loss of generality that $\delta^{(n)}(\boldsymbol{j}) = 1$, since otherwise both sides vanish. But $\delta^{(n)}(\boldsymbol{j}) = 1$ implies that $\boldsymbol{j}$ is of the form $\boldsymbol{j} = (j,\dots,j)$, such that $\delta^{(n)}(\boldsymbol{j},\boldsymbol{y}) = \delta^{(n)}(\boldsymbol{y})$. Moreover, $\mathbb{E}^{(\boldsymbol{j})} = \mathbb{E}_j^n$ and $\nu^{(\boldsymbol{j})} = [\nu^{(j)}]^n$. The multivariate case now follows due to Equation (2.15) with:

$$\begin{aligned}\int_{\mathbb{E}^{(\boldsymbol{j})}} g(y_i)\delta^{(n)}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) &= \int_{\mathbb{E}_j^n} g(y_i)\delta^{(n)}(\boldsymbol{y})[\nu^{(j)}]^n(d\boldsymbol{y}) \\ = \int_{\mathbb{E}_j} g(y)\nu^{(j)}(dy) &= \delta^{(n)}(\boldsymbol{j})\int_{\mathbb{E}_j} g(y)\nu^{(j)}(dy).\end{aligned}$$ □

Proof (Remark 2.40): For the same reason as in the proof of Theorem 2.39, we only need to check the univariate case. If one defines $g(x) := f(x,\dots,x)$, one obviously has $f(\boldsymbol{x})\delta^{(n)}(\boldsymbol{x}) = g(x_i)\delta^{(n)}(\boldsymbol{x})$, for all $1 \le i \le n$. By Equation (2.15) it then follows:

$$\begin{aligned}\int_{\mathbb{E}^n} f(\boldsymbol{x})\delta^{(n)}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) &= \int_{\mathbb{E}^n} g(x_i)\delta^{(n)}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) = \int_{\mathbb{E}} g(x)\nu(dx) \\ = \int_{\mathbb{E}} f(x,\dots,x)\nu(dx).\end{aligned}$$ □

For the proof of Corollary 2.41, we need an intermediate result that allows us to switch between what we have called the *first* and the *second* version. This is the content of the following lemma:

2.46 Lemma (Conversion between first and second version). Let $n \ge 1$.

(1) *Univariate conversion.* Let $\mathscr{P} \in \mathfrak{P}_m^n$ and $\boldsymbol{p}$ the vector representation of $\mathscr{P}$. Then for every function f on $\mathbb{E}^n$ there exists a function g on $\mathbb{E}^m$,

such that:

$$\delta^{(\mathscr{P})}(\boldsymbol{y}) f(\boldsymbol{y}) = \delta^{(\mathscr{P})}(\boldsymbol{y}) g(\boldsymbol{y}_{\mathscr{P}}) \qquad \text{and} \qquad g(\boldsymbol{x}) = f(\boldsymbol{x}^{(\boldsymbol{p})}). \tag{2.25}$$

(2) *Multivariate conversion.* Let $\boldsymbol{j} \in \{1, \ldots, d\}^n$, $\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}$ a refinement and f a function on $\mathbb{E}^{(\boldsymbol{j})}$. Then there exists a function g on $\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}$ such that Equation (2.25) is satisfied. ♢

Proof (Lemma 2.46): As the univariate case is a special case of the multivariate one, we only show the latter:

We show the statement only for functions of the form $f(\boldsymbol{y}) = \prod_{i=k}^{n} f_k(y_k)$, where f_k are functions $\mathbb{E}_{j_k}$. The general result then follows by approximation and monotone convergence.

(i) *First equality.* Assume $\mathscr{P}$ is of the form $\mathscr{P} = \{S_1, \ldots, S_m\}$. Define:

$$g_r(x_r) := \prod_{i \in S_r} f_i(x_r), \qquad\qquad g(\boldsymbol{x}) := \prod_{r=1}^{m} g_r(x_r).$$

Let $\boldsymbol{y} \in \mathbb{E}^{(\boldsymbol{j})}$ be such that $\delta^{(\mathscr{P})}(\boldsymbol{y}) = 1$. Then $y_i = y_j$, for all $i, j \in S_r$, all $1 \leq r \leq m$, so that one can write $y_{S_r} := y_i$, see also Notation 2.27. The first equality now follows from:

$$\begin{aligned} f(\boldsymbol{y}) &= \prod_{k=1}^{n} f_k(y_k) = \Big[\prod_{r=1}^{m} \prod_{i \in S_r} f_i(y_i)\Big] = \Big[\prod_{r=1}^{m} \prod_{i \in S_r} f_i(y_{S_r})\Big] \\ &= \Big[\prod_{r=1}^{m} g_r(y_{S_r})\Big] = g(y_{S_1}, \ldots, y_{S_r}) = g(\boldsymbol{y}_{\mathscr{P}}). \end{aligned}$$

(ii) *Second equality.* Let $\boldsymbol{p}$ be the vector representation of $\mathscr{P}$. Because $i \in S_r$ if and only if $p_i = r$, one has:

$$\begin{aligned} g(\boldsymbol{x}) &= \Big[\prod_{r=1}^{m} \prod_{i \in S_r} f_i(x_r)\Big] = \Big[\prod_{r=1}^{m} \prod_{i \in S_r} f_i(x_{p_i})\Big] = \Big[\prod_{k=1}^{n} f_k(x_{p_k})\Big] \\ &= f\big(x_{p_1}, \ldots, x_{p_n}\big) = f(\boldsymbol{x}^{(\boldsymbol{p})}). \end{aligned}$$

□

Proof (Corollary 2.41): Recall that an ordered partition $\mathscr{P} \in \mathfrak{P}_m^n$ with corresponding sets $\{S_1, \ldots, S_m\}$ can alternatively be represented as an index vector $\boldsymbol{p} \in \{1, \ldots, m\}^n$, see Definition 2.24.

(1a) *First univariate version, ordinary measure.* The statement is only shown for functions of the form $g(x) = \prod_{k=1}^{m} g_r(x_r)$, where g_r are defined on $\mathbb{E}$. The general case then follows by approximation and monotone convergence. Hence, one has in a first step:

$$\begin{aligned}
&\int_{\mathbb{E}^n} g(\boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}) \\
&\quad = \int_{\mathbb{E}^n} \Big[\prod_{r=1}^{m} g_r(y_{S_r})\Big]\Big[\prod_{r=1}^{m} \delta^{(|S_r|)}(\boldsymbol{y}_{S_r})\Big]\Big[\prod_{r=1}^{m} \nu^{(|S_r|)}(d\boldsymbol{y}_{S_r})\Big] \\
&\quad = \prod_{r=1}^{m}\Big[\int_{\mathbb{E}^{|S_r|}} g_r(y_{S_r})\delta^{(|S_r|)}(\boldsymbol{y}_{S_r})\nu^{(|S_r|)}(d\boldsymbol{y}_{S_r})\Big].
\end{aligned}$$

Applying Equation (2.15), one gets:

$$(\ldots) = \prod_{r=1}^{m}\Big[\int_{\mathbb{E}} g_r(x_r)\nu(dx_r)\Big] = \int_{\mathbb{E}^m} g(\boldsymbol{x})\nu^{(m)}(d\boldsymbol{x}).$$

(1b) *First univariate version, factorial measure.* Recall that one has by definition $\delta^{[\mathscr{P}]}(\boldsymbol{y}) = \delta^{(\mathscr{P})}(\boldsymbol{y})\delta^{[m]}(\boldsymbol{y}_{\mathscr{P}})$. Hence:

$$\begin{aligned}
\int_{\mathbb{E}^n} g(\boldsymbol{y}_{\mathscr{P}})\delta^{[\mathscr{P}]}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}) &= \int_{\mathbb{E}^n} g(\boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{y})\delta^{[m]}(\boldsymbol{y}_{\mathscr{P}})\nu^{(n)}(d\boldsymbol{y}) \\
&= \int_{\mathbb{E}^n} \Big[g(\boldsymbol{y}_{\mathscr{P}})\delta^{[m]}(\boldsymbol{y}_{\mathscr{P}})\Big]\delta^{(\mathscr{P})}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}).
\end{aligned}$$

Applying the first part of Equation (2.17), one gets:

$$(\ldots) = \int_{\mathbb{E}^m} \Big[g(\boldsymbol{x})\delta^{[m]}(\boldsymbol{x})\Big]\nu^{(m)}(d\boldsymbol{x}) = \int_{\mathbb{E}^m} g(\boldsymbol{x})\nu^{[m]}(d\boldsymbol{x}).$$

(2a) *First multivariate version, ordinary measure.* The statement is only shown for functions of the form $g(\boldsymbol{x}) = \prod_{r=1}^{m} g_r(x_r)$, where g_r are defined on $\mathbb{E}_{[\boldsymbol{j}_{\mathscr{P}}]_r} = \mathbb{E}_{j_{S_r}}$. The general case then follows by approximation

and monotone convergence. Hence, one has in a first step:

$$\begin{aligned}\int_{\mathbb{E}^{(\boldsymbol{j})}} & g(\boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) \\ &= \int_{\mathbb{E}^{(\boldsymbol{j})}} \Big[\prod_{r=1}^{m} g_r(y_{S_r})\Big]\Big[\prod_{r=1}^{m}\delta^{(|S_r|)}(\boldsymbol{j}_{S_r},\boldsymbol{y}_{S_r})\Big]\Big[\prod_{r=1}^{m}\nu^{(\boldsymbol{j}_{S_r})}(d\boldsymbol{y}_{S_r})\Big] \\ &= \prod_{r=1}^{m}\Big[\int_{\mathbb{E}_{j_{S_r}}} g_r(y_{S_r})\delta^{(|S_r|)}(\boldsymbol{j}_{S_r},\boldsymbol{y}_{S_r})\nu^{(\boldsymbol{j}_{S_r})}(d\boldsymbol{y}_{S_r})\Big].\end{aligned}$$

Applying Equation (2.16), one gets:

$$(\ldots) = \prod_{r=1}^{m}\Big[\delta^{(|\mathscr{P}|)}(\boldsymbol{j}_{S_r})\int_{\mathbb{E}_{j_{S_r}}} g_r(x_r)\nu^{(j_{S_r})}(dx_r)\Big] = \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} g(\boldsymbol{x})\nu^{(\boldsymbol{j}_{\mathscr{P}})}(d\boldsymbol{x}).$$

Since $\mathscr{P}$ is a refinement of $\mathscr{P}_{\boldsymbol{j}}$, the components of the vector $\boldsymbol{j}_{S_r}$ are all identical. In consequence, $\delta^{(|\mathscr{P}|)}(\boldsymbol{j}_{S_r}) = 1$, and this was used in the last equality above.

(2b) *First multivariate version, factorial measure.* In a first step note that:

$$\begin{aligned}\int_{\mathbb{E}^{(\boldsymbol{j})}} & g(\boldsymbol{y}_{\mathscr{P}})\delta^{[\mathscr{P}]}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) \\ &= \int_{\mathbb{E}^{(\boldsymbol{j})}} \Big[g(\boldsymbol{y}_{\mathscr{P}})\delta^{[m]}(\boldsymbol{j}_{\mathscr{P}},\boldsymbol{y}_{\mathscr{P}})\Big]\delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}).\end{aligned}$$

Applying the first part of Equation (2.18), one gets:

$$(\ldots) = \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} \Big[g(\boldsymbol{x})\delta^{[m]}(\boldsymbol{j}_{\mathscr{P}},\boldsymbol{x})\Big]\nu^{(\boldsymbol{j}_{\mathscr{P}})}(d\boldsymbol{x}) = \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} g(\boldsymbol{x})\nu^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{x}).$$

(3a) *Second univariate version, ordinary measure.* Due to the first part of Equation (2.25), one can find a function g on $\mathbb{E}^m$ such that

$$\int_{\mathbb{E}^n} f(\boldsymbol{y})\delta^{(\mathscr{P})}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}) = \int_{\mathbb{E}^n} g(\boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{y})\nu^{(n)}(d\boldsymbol{y}).$$

Applying now the first part of Equation (2.17) and the second part of Equation (2.25), one gets:

$$(\ldots) = \int_{\mathbb{E}^m} g(\boldsymbol{x})\nu^{(m)}(d\boldsymbol{x}) = \int_{\mathbb{E}^m} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{(m)}(d\boldsymbol{x}).$$

(3b) *Second univariate version, factorial measure.* This follows in the same way as in the univariate case: First replace f with a function g, due to Equation (2.25), and then apply Equation (2.17).

(4a) *Second multivariate second version, ordinary measure.* Due to the first part of Equation (2.25), one can find a function g on $\mathbb{E}^m$ such that for all $\boldsymbol{y} \in \mathbb{E}^n$ one has $\delta^{(\mathscr{P})}(\boldsymbol{y})f(\boldsymbol{y}) = \delta^{(\mathscr{P})}(\boldsymbol{y})g(\boldsymbol{y}_{\mathscr{P}})$. Hence:

$$\begin{aligned}\delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y})f(\boldsymbol{y}) &= \delta^{(\mathscr{P})}(\boldsymbol{j})\Big[\delta^{(\mathscr{P})}(\boldsymbol{y})f(\boldsymbol{y})\Big] = \delta^{(\mathscr{P})}(\boldsymbol{j})\Big[\delta^{(\mathscr{P})}(\boldsymbol{y})g(\boldsymbol{y}_{\mathscr{P}})\Big] \\ &= \delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y})g(\boldsymbol{y}_{\mathscr{P}}).\end{aligned}$$

Together with the first part of Equation (2.18), the claim follows from:

$$\begin{aligned}\int_{\mathbb{E}^{(\boldsymbol{j})}} f(\boldsymbol{y})\delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) &= \int_{\mathbb{E}^{(\boldsymbol{j})}} g(\boldsymbol{y}_{\mathscr{P}})\delta^{(\mathscr{P})}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) \\ = \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} g(\boldsymbol{x})\nu^{(\boldsymbol{j}_{\mathscr{P}})}(d\boldsymbol{x}) &= \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{(\boldsymbol{j}_{\mathscr{P}})}(d\boldsymbol{x}).\end{aligned}$$

(4b) *Second multivariate version, factorial measure.* This is shown in the same way as the statement above but using the second part of Equation (2.18) instead. □

Proof (Theorem 2.42): Recall that an ordered partition $\mathscr{P} \in \mathfrak{P}_m^n$ of size m can be represented equivalently as a sequence of sets $\{S_1, \ldots, S_m\}$ or as an index vector $\boldsymbol{p} \in \{1, \ldots, m\}^n$, see Definition 2.24.

(1a) *Univariate, factorial decomposition.* With the decomposition form Equation (2.11), one has:

$$\int_{\mathbb{E}^n} f(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) = \sum_{m=1}^{n} \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^n} f(\boldsymbol{x})\delta^{[\mathscr{P}]}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}).$$

Due to the second part of Equation (2.19), the statement now follows with

$$(\ldots) = \sum_{m=1}^{n} \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^m} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{[m]}(d\boldsymbol{x}).$$

(1b) *Multivariate, factorial decomposition.* In the same way as above, use the

decomposition form Equation (2.12). One then gets in a first step:

$$\int_{\mathbb{E}^{(\boldsymbol{j})}} f(\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}) = \sum_{\mathscr{P}\ll\mathscr{P}_{\boldsymbol{j}}} \int_{\mathbb{E}^{(\boldsymbol{j})}} f(\boldsymbol{y})\delta^{[\mathscr{P}]}(\boldsymbol{j},\boldsymbol{y})\nu^{(\boldsymbol{j})}(d\boldsymbol{y}).$$

Due to the second part of Equation (2.20), the statement then follows with:

$$(\ldots) = \sum_{\mathscr{P}\ll\mathscr{P}_{\boldsymbol{j}}} \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} f(\boldsymbol{y}^{(\boldsymbol{p})})\nu^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{y}).$$

(2a) *Univariate, symmetric, factorial decomposition.* Define first $f(\boldsymbol{x}) := \prod_{k=1}^n f_k(x_k)$. Then, due to the univariate version of Equation (2.21), one gets:

$$\begin{aligned}\int_{\mathbb{E}^n}\Big[\prod_{k=1}^n f_k(x_k)\Big]\nu^{(n)}(d\boldsymbol{x}) &= \int_{\mathbb{E}^n} f(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) \\ &= \sum_{m=1}^n \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^m} f(\boldsymbol{x}^{(\boldsymbol{p})})\nu^{[m]}(d\boldsymbol{x}).\end{aligned}$$

Now fix a partition $\mathscr{P}\in\mathfrak{P}_m^n$ and denote its associated vector representation with $\boldsymbol{p}$. Since $\boldsymbol{x}^{(\boldsymbol{p})} = (x_{p_1},\ldots,x_{p_n})$, for $\boldsymbol{x}\in\mathbb{E}^n$, the k-th component is $[\boldsymbol{x}^{(\boldsymbol{p})}]_k = x_{p_k}$. Hence:

$$f(\boldsymbol{x}^{(\boldsymbol{p})}) = \prod_{k=1}^n f_k\Big([\boldsymbol{x}^{(\boldsymbol{p})}]_k\Big) = \prod_{k=1}^n f_k(x_{p_k}).$$

Assume the partition $\mathscr{P}$ is of the form $\mathscr{P} = \{S_1,\ldots,S_m\}$. Recall that the index vector $\boldsymbol{p}$ is characterized by the property that $p_i = r$ holds if and only if $i\in S_r$. Therefore:

$$f(\boldsymbol{x}^{(\boldsymbol{p})}) = \prod_{k=1}^n f_k(x_{p_k}) = \prod_{r=1}^m \prod_{i\in S_r} f_i(x_{p_i}) = \prod_{r=1}^m \prod_{i\in S_r} f_i(x_r).$$

After substitution of this expression in the calculation above, the claim follows.

(2b) *Multivariate, symmetric, factorial decomposition.* As in the univariate case, define $f(\boldsymbol{y}) := \prod_{k=1}^n f_k(y_k)$ and then apply the multivariate version

of Equation (2.21). The remaining steps of the proof then follow along the same lines as in the univariate case. □

PROOF (COROLLARY 2.43): Both statements are only shown for the multivariate case, since the proof for the univariate case is almost identical.

(1) *Factorial decomposition for sets.* Take a sequence of sets $A_k \subseteq \mathbb{E}_{j_k}$, $1 \leq k \leq n$, and define the functions $f_k(y_k) := \mathbb{1}_{A_k}(y_k)$. Due to the multivariate version of Equation (2.22), one has:

$$\begin{aligned}\nu^{(\boldsymbol{j})}(A_1 \times \ldots \times A_n) &= \int_{\mathbb{E}^{(\boldsymbol{j})}} \Big[\prod_{k=1}^{n} f_k(y_k)\Big] \nu^{(\boldsymbol{j})}(d\boldsymbol{y}) \\ &= \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} \prod_{r=1}^{|\mathscr{P}|} \Big[\prod_{i \in S_r} f_i(y_r)\Big] \nu^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{y}).\end{aligned}$$

Clearly, $\mathfrak{P}^n$ is the same as the union of all $\mathfrak{P}^n_m$, for $1 \leq m \leq n$. Hence, the previous expression becomes:

$$\begin{aligned}(\ldots) &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}^n_m, \mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} \Big[\prod_{r=1}^{m} \mathbb{1}_{\{\bigcap_{i \in S_r} A_i\}}(y_r)\Big] \nu^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{y}) \\ &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}^n_m, \mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \nu^{[\boldsymbol{j}_{\mathscr{P}}]}\Big(\bigcap_{i \in S_1} A_i \times \ldots \times \bigcap_{i \in S_m} A_i\Big).\end{aligned}$$

(2) *Two dimensional special case.* Let $1 \leq i, j \leq d$ and define the index vector $\boldsymbol{k} := (ij)$. According to the multivariate version of Equation (2.21), one has:

$$\int_{\mathbb{E}^{(ij)}} f(\boldsymbol{y}) \nu^{(ij)}(d\boldsymbol{y}) = \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{k}}} \int_{\mathbb{E}^{(\boldsymbol{k}_{\mathscr{P}})}} f(\boldsymbol{y}^{(\boldsymbol{p})}) \nu^{[\boldsymbol{k}_{\mathscr{P}}]}(d\boldsymbol{y}), \tag{2.26}$$

where $\boldsymbol{p}$ is the vector representation of $\mathscr{P}$. Note that, depending on the values of i, j, $\mathscr{P}_{\boldsymbol{k}}$ is equal to:

$$\mathscr{P}_{\boldsymbol{k}} = \begin{cases} \mathscr{Q}_1 & \text{if } i \neq j, \\ \mathscr{Q}_2 & \text{if } i = j, \end{cases}$$

where the two partitions $\mathscr{Q}_1$ and $\mathscr{Q}_2$ of $\{1,2\}$ are defined as:

$$\mathscr{Q}_1 := \big\{\{1\},\{2\}\big\} \qquad \text{and} \qquad \mathscr{Q}_2 := \big\{\{1,2\}\big\}.$$

Next determine the family of all refinements $\{\mathscr{P} : \mathscr{P} \ll \mathscr{P}_{\boldsymbol{k}}\}$. One needs to distinguish between the same two cases as above:

$$\{\mathscr{P} : \mathscr{P} \ll \mathscr{P}_{\boldsymbol{k}}\} = \begin{cases} \{\mathscr{Q}_1\} & \text{if } i \neq j, \text{ i.e. if } \mathscr{P}_{\boldsymbol{k}} = \mathscr{Q}_1, \\ \{\mathscr{Q}_1, \mathscr{Q}_2\} & \text{if } i = j, \text{ i.e. if } \mathscr{P}_{\boldsymbol{k}} = \mathscr{Q}_2. \end{cases}$$

Continuing with Equation (2.26), one gets:

$$\begin{aligned} \int_{\mathbb{E}^{(ij)}} f(\boldsymbol{y})\nu^{(ij)}(d\boldsymbol{y}) &= \delta_{i\neq j} \sum_{\mathscr{P}\ll\mathscr{P}_{\boldsymbol{k}}} \int_{\mathbb{E}^{(\boldsymbol{k}_{\mathscr{P}})}} f(\boldsymbol{y}^{(\boldsymbol{p})})\nu^{[\boldsymbol{k}_{\mathscr{P}}]}(d\boldsymbol{y}) \\ &\quad + \delta_{i=j} \sum_{\mathscr{P}\ll\mathscr{P}_{\boldsymbol{k}}} \int_{\mathbb{E}^{(\boldsymbol{k}_{\mathscr{P}})}} f(\boldsymbol{y}^{(\boldsymbol{p})})\nu^{[\boldsymbol{k}_{\mathscr{P}}]}(d\boldsymbol{y}) \\ &= \delta_{i\neq j} \sum_{\mathscr{P}\in\{\mathscr{Q}_1\}} \int_{\mathbb{E}^{(\boldsymbol{k}_{\mathscr{P}})}} f(\boldsymbol{y}^{(\boldsymbol{p})})\nu^{[\boldsymbol{k}_{\mathscr{P}}]}(d\boldsymbol{y}) \\ &\quad + \delta_{i=j} \sum_{\mathscr{P}\in\{\mathscr{Q}_1,\mathscr{Q}_2\}} \int_{\mathbb{E}^{(\boldsymbol{k}_{\mathscr{P}})}} f(\boldsymbol{y}^{(\boldsymbol{p})})\nu^{[\boldsymbol{k}_{\mathscr{P}}]}(d\boldsymbol{y}). \end{aligned}$$

Next determine $\boldsymbol{k}_{\mathscr{P}}$ for all possible choices of $\mathscr{P}$:

$$\begin{aligned} &\text{If } i \neq j\text{:} \quad \boldsymbol{k}_{\mathscr{Q}_1} = \boldsymbol{k}, \\ &\text{If } i = j\text{:} \quad \boldsymbol{k}_{\mathscr{Q}_1} = \boldsymbol{k} \text{ and } \boldsymbol{k}_{\mathscr{Q}_2} = (i). \end{aligned}$$

Moreover, the vector representations of $\mathscr{Q}_1$ and $\mathscr{Q}_2$ are

$$\boldsymbol{p}_1 := (1,2) \qquad \text{and} \qquad \boldsymbol{p}_2 := (1,1)$$

Therefore:

$$\begin{aligned} \int_{\mathbb{E}^{(ij)}} f(\boldsymbol{y})\nu^{(ij)}(d\boldsymbol{y}) &= \delta_{i\neq j} \int_{\mathbb{E}^{(\boldsymbol{k})}} f(\boldsymbol{y}^{(\boldsymbol{p}_1)})\nu^{[\boldsymbol{k}]}(d\boldsymbol{y}) \\ &\quad + \delta_{i=j} \int_{\mathbb{E}^{(\boldsymbol{k})}} f(\boldsymbol{y}^{(\boldsymbol{p}_1)})\nu^{[\boldsymbol{k}]}(d\boldsymbol{y}) + \delta_{i=j} \int_{\mathbb{E}^{(i)}} f(\boldsymbol{y}^{(\boldsymbol{p}_2)})\nu^{[i]}(dy). \end{aligned}$$

Note that one has for $y \in \mathbb{E}$ and $\boldsymbol{y} \in \mathbb{E}^2$:

$$\boldsymbol{y}^{(\boldsymbol{p}_1)} = \boldsymbol{y} \qquad \text{and} \qquad \boldsymbol{y}^{(\boldsymbol{p}_2)} = (y, y).$$

Substituting this above gives:

$$\begin{aligned}\int_{\mathbb{E}^{(ij)}} f(\boldsymbol{y})\nu^{(ij)}(d\boldsymbol{y}) &= \delta_{i\neq j}\int_{\mathbb{E}^{(\boldsymbol{k})}} f(\boldsymbol{y})\nu^{[\boldsymbol{k}]}(d\boldsymbol{y}) \\ &\quad + \delta_{i=j}\int_{\mathbb{E}^{(\boldsymbol{k})}} f(\boldsymbol{y})\nu^{[\boldsymbol{k}]}(d\boldsymbol{y}) + \delta_{i=j}\int_{\mathbb{E}^{(i)}} f(y,y)\nu^{[i]}(dy) \\ &= \int_{\mathbb{E}^{(\boldsymbol{k})}} f(\boldsymbol{y})\nu^{[\boldsymbol{k}]}(d\boldsymbol{y}) + \delta_{i=j}\int_{\mathbb{E}^{(i)}} f(y,y)\nu^{[i]}(dy).\end{aligned}$$

If one uses that $\nu^{[i]} = \nu^{(i)}$, the statement follows, since for the first order measures the ordinary and factorial product measures coincide. □

Chapter 3

Point Processes

This chapter introduces point configuration spaces and point processes, which are random variables with values in a point configuration space. Then, as a first step towards the definition of Hawkes processes, so-called cluster processes are defined and formulas for their moment measures are derived.

3.1 Motivation and Objectives

The content and notation of this chapter is based mainly on Section 6.3 in [DVJ03]. I give a short outline of and explain the reasoning behind the chosen approach.

Cluster Fields. One has essentially two possibilities if one wants to define what are usually called cluster fields or cluster families: Either one restricts oneself to random measures and their distributions alone, i.e. one does not use an underlying probability space. This is the approach taken e.g. in the book [MKM78]. It has the advantage that one does not need to embed the cluster field as a stochastic object in an underlying probability space.

Cluster Families. The alternative approach is to construct a measurable family of point processes $\{K(\cdot|y); y \in \mathbb{E}\}$. This approach has the big advantage that each cluster process $N(\cdot|y)$ can be treated as a random variable, although it takes values not in $\mathbb{R}$ but in a point configuration space. Results can then be derived using standard notation and calculus from probability theory. But it has also a disadvantage: One first needs to specify what the underlying

probability space is and then one needs to give a sensible definition of a cluster family.

Cluster Fields vs. Cluster Families. The difference between the two approaches can be illustrated with the following trivial example: Assume one has two independent random variables X and Y and one is interested in the sum $X + Y$. If the two random variables are defined on an underlying probability, one can naturally speak about the distribution of the sum $X + Y$, which is given by the convolution $F_X * F_Y$. But on the other hand, if one is not using an underlying probability space one can only use the distributions F_X and F_Y as building blocks. It is then not possible to speak about the sum of the two random variables but only about the convolution $F_X * F_Y$.

In the end, both approaches mean the same thing and lead to the same results, only the point of view is different. Let me emphasize that the above example is only an illustration and should demonstrate the implications that the two different approaches entail. If one is actually dealing with point processes and distributions thereof, the situation is far less trivial.

It appears to be better if one defines clusters in terms of a family $K(\cdot|y)_{y\in\mathbb{E}}$ of cluster processes instead of only to calculate with distributions. Unfortunately, there is a small technical problem one cannot simply ignore: Assume for simplicity $\mathbb{E} := \mathbb{R}$, then a cluster family is a family of point processes $K(\cdot|y)_{y\in\mathbb{R}}$. In other words, for each $y \in \mathbb{R}$, there is an associated cluster process $K(\cdot|y)$. Let us call such a family a *full* cluster family, since it defines a cluster for every $y \in \mathbb{E}$. Clearly, one is now speaking of an uncountable number of random variables, and some of the usual problems come with this: For example it is not clear whether the mapping $y \mapsto N(A|y)$ is a measurable function on the space $\mathbb{R} \times \Omega$, for a given bounded $A \subseteq \mathbb{R}$. Without such elementary guarantees, it is impossible to build a reasonable theory.

Full Cluster Fields. Assume one has given a center process, say L, which describes the center points, i.e. the random set of points $y \in \mathbb{E}$ for which a cluster $K(\cdot|y)$ should be generated. Assume such a center point is given by the random variable Y_i. If $K(\cdot|y)_{y\in\mathbb{E}}$ were a full cluster field, as described above, one could now substitute $y \twoheadrightarrow Y_i$ and obtain $K(\cdot|Y_i)$, but the mentioned measurability problems would then arrive.

Sparse Cluster Fields. To avoid the above mentioned problems, I decided to take a different road: Instead of first defining a full cluster family and then

substituting the center points, I directly define a *sparse* cluster family of the form $K(\cdot|Y_i)$. See Definition 3.19 for the exact definition, but the idea is as follows: The cluster family $K(\cdot|y)_{y\in\mathbb{E}}$ is not defined as a separate entity but only in conjunction with the center process L. In fact, $K(\cdot|y)$ is defined for all $y \in \mathbb{E}$, but only if y is chosen as one of the points of L, i.e. only if $y = Y_i$, there is actually an associated cluster. For all other y, the cluster $K(\cdot|y)$ is simply an empty point process. This explains why I call it a *sparse* cluster field.

To implement the above idea, a couple of steps are needed. First of all, a suitable notation needs to be found. In Remark 3.10 and Remark 3.11 and Definition 3.12 it is explained how one can enumerate the points of a point process irrespective of the actual number of point there are. In a next step, so-called *generating clusters* are introduced and then shifted according to the center points of L, see Definition 3.17. Finally, the cluster family is defined by selecting the correct shifted cluster, see Definition 3.19 for a more formal description.

The construction explained above makes one simplifying assumption, namely that all clusters have the same distribution. This allows us to construct the clusters by simply shifting the sequence of generating clusters, see Definition 3.17. But the same construction could also be made in a more general setting, where the cluster distribution depends on the location of the center point y. In this case, one would have to construct the cluster, actually the underlying probability space, using Kolmogorov's Extension Theorem or the Theorem of Ionescu-Tulcea.

3.2 Point Configuration Spaces

Consider the case of a point process on the half line $\mathbb{R}_+$. There are several ways of presenting such a point process:

(i) Enumerate the points and describe the point process as a sequence $(T_k)_{k\geq 1}$ of increasing random variables.

(ii) Consider the counting process $X_t := N([0,t])$, for $t \geq 0$, i.e. the process that counts the number of points in the interval $[0,t]$.

(iii) Interpret the point process as random measures, i.e. as a random variable with values in some point configuration space.

In this thesis, we almost always adopt the last version and treat a point process as a random measure. Recall Definition 2.33 of the point configuration space

$\mathscr{N}(\mathbb{E})$. The next definition treats only the univariate case. The extension to the multivariate case is trivial:

3.1 Definition (Measurable point configuration space). Let $\mathbb{E} := \mathbb{R}^e$, for an integer $e \geq 1$, be a univariate event space.

(*1*) Let $C \subseteq \mathbb{E}$ be a bounded set. The *projection* π_C is a function on $\mathscr{N}(\mathbb{E})$ and defined as:

$$\pi_C : \mathscr{N}(\mathbb{E}) \to \mathbb{N}_0, \qquad \pi_C(\nu) := \nu(C). \tag{3.1}$$

(*2*) Let $n \geq 1$ and $\mathscr{C} = \{C_1, \ldots, C_n\}$ be a family of bounded sets $C_k \subseteq \mathbb{E}$. For a vector $\boldsymbol{r} \equiv (r_1, \ldots, r_n)$ with components $r_k \in \mathbb{N}_0$ define the *cylinder set* $\mathfrak{C}_{\mathscr{C},\boldsymbol{r}}$ generated by $(\mathscr{C}, \boldsymbol{r})$ by:

$$\mathfrak{C}_{\mathscr{C},\boldsymbol{r}} := \Big\{\nu \in \mathscr{N}(\mathbb{E}) : \bigcap_{k=1}^{n} \pi_{C_k}(\nu) = r_k\Big\}. \tag{3.2}$$

Note that $\mathfrak{C}_{\mathscr{C},\boldsymbol{r}} \subseteq \mathscr{N}(\mathbb{E})$. Moreover, for all $\nu \in \mathscr{N}(\mathbb{E})$ one has:

$$\nu \in \mathfrak{C}_{\mathscr{C},\boldsymbol{r}} \qquad \Leftrightarrow \qquad \nu(C_k) = r_k, \text{ all } 1 \leq k \leq n.$$

(*3*) For $n \geq 1$ and a family of sets $\mathscr{C} = \{C_1, \ldots, C_n\}$ as above, define:

$$\pi_{\mathscr{C}} := (\pi_{C_1}, \ldots, \pi_{C_n}).$$

Hence, $\pi_{\mathscr{C}}$ is a vector-valued function with values in $\mathbb{N}_0 \times \ldots \times \mathbb{N}_0$. It is called the *finite dimensional projection* with respect to the family $\mathscr{C}$. $\diamondsuit$

Note that in the definition above, one has to assume that the sets C_k are bounded, so that all projections $\nu(C_k)$ are finite.

In order to define random point measures, we need more structure on the point configuration space $\mathscr{N}(\mathbb{E})$. To this end, we enrich $\mathscr{N}(\mathbb{E})$ with a topology $\mathfrak{T}(\mathbb{E})$ and define the natural σ-algebra $\mathfrak{F}(\mathbb{E})$:

3.2 Definition (Topology and σ-algebra). Let $\mathbb{E} = \mathbb{R}^e$ be a univariate event space.

(*1*) The *topology of vague convergence* $\mathfrak{T}(\mathbb{E})$ on $\mathscr{N}(\mathbb{E})$ is the topology generated by all functions of the form:

$$\mathscr{N}(\mathbb{E}) \to \mathbb{R}_+, \qquad \text{where} \qquad \nu \mapsto \int_{\mathbb{E}} f(x)\nu(dx),$$

where f is a continuous function on $\mathbb{E}$ with compact support.

(2) The natural σ-algebra $\mathfrak{F}(\mathbb{E})$ is the smallest σ-algebra such that all projections of the form given in Equation (3.1) are measurable. In other words, define:

$$\mathfrak{F}(\mathbb{E}) := \sigma\Big\{\pi_C : C \subseteq \mathbb{E}, \text{ bounded and measurable}\Big\}.$$

Equivalently, the natural σ-algebra $\mathfrak{F}(\mathbb{E})$ can be defined as:

$$\mathfrak{F}(\mathbb{E}) := \mathscr{B}\big(\mathfrak{T}(\mathbb{E})\big).$$

Hence, $\mathfrak{F}(\mathbb{E})$ is the Borel-σ-algebra generated by the topology $\mathfrak{T}(\mathbb{E})$. $\diamondsuit$

Note that the functions f in the definition of the topology $\mathfrak{T}(\mathbb{E})$ need to have compact support, since otherwise the integral might not be finite.

In summary, the above definitions lead to a well-defined measurable space $\mathscr{N}(\mathbb{E})$ equipped with σ-algebra $\mathfrak{F}(\mathbb{E})$.

3.3 Convention. Whenever we deal with the point configuration space $\mathscr{N}(\mathbb{E})$, we will always use the associated natural σ-algebra $\mathfrak{F}(\mathbb{E})$. Moreover, when we write $\mathfrak{A} \subseteq \mathscr{N}(\mathbb{E})$, we actually mean $\mathfrak{A} \in \mathfrak{F}(\mathbb{E})$, i.e. we implicitly assume that $\mathfrak{A}$ is a *measurable* set. $\diamondsuit$

The next definition treats only the univariate case. The extension to the multivariate case is trivial. Since $\mathfrak{F}(\mathbb{E})$ is generated by the family of all projections $\pi_C : \mathscr{N}(\mathbb{E}) \to \mathbb{N}_0$, for bounded sets $C \subseteq \mathbb{E}$, one can give two equivalent definitions of a point process:

3.4 Definition (Point process). Let $\mathscr{N}(\mathbb{E})$ with σ-algebra $\mathfrak{F}(\mathbb{E})$ denote the measurable point configuration space on $\mathbb{E}$ and assume $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space. The following two definitions for a point process are equivalent:

(*i*) A function N on Ω with values in $\mathscr{N}(\mathbb{E})$ is a *point process* if and only if it is measurable with respect to $\mathscr{F}$ and $\mathfrak{F}(\mathbb{E})$.

(*ii*) A function N on Ω with values in $\mathscr{N}(\mathbb{E})$ is a *point process* if and only if all real-valued functions of the form $\pi_C \circ N \equiv N(C)$ are measurable with respect to the Borel-σ-algebra $\mathscr{B}(\mathbb{R})$, for all bounded, measurable sets $C \subseteq \mathbb{E}$. $\diamondsuit$

3.5 Definition (Distribution of a point process). Let N be a point process with values in $\mathscr{N}(\mathbb{E})$. The *distribution*, say Φ_N, of N is the image probability measure induced by N on the measurable space $\mathscr{N}(\mathbb{E})$. In other words, the distribution Φ_N is defined by:

$$\Phi_N(\mathfrak{A}) := \mathbb{P}\big[N^{-1}(\mathfrak{A})\big] = \mathbb{P}\Big[\big\{\omega \in \Omega : N(\omega) \in \mathfrak{A}\big\}\Big],$$

where $\mathfrak{A} \subseteq \mathscr{N}(\mathbb{E})$. ♢

Note that the family of cylinder sets, see Equation (3.2), is an intersection stable generator for the σ-algebra $\mathfrak{F}(\mathbb{E})$. Hence, one only needs to know all finite-dimensional distributions of a point process N, in order to uniquely determine its distribution:

3.6 Remark (Finite-dimensional distributions). The distribution Φ_N of a point process N is uniquely determined if for all $n \geq 1$, the distributions of all n-dimensional random variables of the following form are known:

$$\pi_{\mathscr{C}} \circ N \equiv \Big(N(C_1), \ldots, N(C_n)\Big),$$

where $\mathscr{C} = \{C_1, \ldots, C_n\}$ and bounded $C_k \subseteq \mathbb{E}$, for $1 \leq k \leq n$. An alternative way to look at this statement is as follows: Let Φ be a point process distribution and N be a point process. Then N has distribution Φ if and only if for all $n \geq 1$, all families of sets $\mathscr{C} = \{C_1, \ldots, C_n\}$ and all vectors $\boldsymbol{r} \in \mathbb{N}_0^n$ it holds that:

$$\mathbb{P}\Big[N(C_1) = r_1, \ldots, N(C_n) = r_n\Big] = \Phi\big(\mathfrak{C}_{\mathscr{C},\boldsymbol{r}}\big).$$

For the definition of $\mathfrak{C}_{\mathscr{C},\boldsymbol{r}}$ see Equation (3.2). ♢

The following definition of moment measures is based on the Definition 2.36 of product measures:

3.7 Definition (Moment measure). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and N either a univariate point process with values in $\mathscr{N}(\mathbb{E})$ or a multivariate point process with values in $\mathscr{N}(\mathbb{E})$. Furthermore, let $\boldsymbol{j} \in \{1, \ldots, d\}^n$.

(*1*) The *ordinary moment measure* of order n in the univariate and multivariate case are defined by:

$$M_N^{(n)}(d\boldsymbol{x}) := \mathbb{E}\big[N^{(n)}(d\boldsymbol{x})\big], \qquad M_N^{(\boldsymbol{j})}(d\boldsymbol{y}) := \mathbb{E}\big[N^{(\boldsymbol{j})}(d\boldsymbol{y})\big].$$

(2) The *factorial moment measure* of order n in the univariate and multivariate case are defined by:

$$M_N^{[n]}(d\boldsymbol{x}) := \mathbb{E}\big[N^{[n]}(d\boldsymbol{x})\big], \qquad M_N^{[\boldsymbol{j}]}(d\boldsymbol{y}) := \mathbb{E}\big[N^{[\boldsymbol{j}]}(d\boldsymbol{y})\big]. \qquad \diamondsuit$$

Related to factorial moment measures is the following elementary definition. It is taken from Section 5.2 in [DVJ03]:

3.8 Definition (Falling factorial). Let $r, n \geq 0$ be two integers. The *falling factorial* of order n, sometimes also called the *factorial power* of order n, is defined as:

$$r^{[0]} := 1, \qquad r^{[n]} := r(r-1) \cdot \ldots \cdot (r-n+1), \text{ for } n \geq 1.$$

Especially, $r^{[n]} = 0$ whenever $n > r$. $\diamondsuit$

In order to enumerate the points of a point process irrespective of the actual number of total points, we need to introduce some more notation:

3.9 Definition (Extended state space). Let $\mathbb{E} = \mathbb{R}^e$ be a univariate event space and assume $\varnothing \notin \mathbb{E}$ is an arbitrary *auxiliary point*. Define the *extended event space* $\bar{\mathbb{E}}$ by:

$$\bar{\mathbb{E}} := \mathbb{E} \cup \{\varnothing\}. \qquad \diamondsuit$$

The auxiliary point $\varnothing$ allows us to use a consistent notation, independent of whether a point configuration ν has a finite or infinite number of points. The following remark serves as a preparation for Remark 3.11:

3.10 Remark (Enumeration of point configurations). Let ν_* be a finite and ν_∞ be a countably infinite point configuration in $\mathscr{N}(\mathbb{E})$, i.e. assume that $\nu_*(\mathbb{E}) = n$ and $\nu_\infty(\mathbb{E}) = \infty$. Clearly, one can then take a vector $(x_1, \ldots, x_n)$ and a sequence $(y_1, y_2, \ldots)$ with values in $\mathbb{E}$ such that:

$$\nu_*(dz) = \sum_{i=1}^{n} \delta_{x_i}(dz), \qquad \nu_\infty(dz) = \sum_{i=1}^{\infty} \delta_{y_i}(dz). \qquad \diamondsuit$$

The same idea can now be applied to a point process. If N is a point process with $N(\mathbb{E})$ points, we can represent N either as a random point configuration of the form ν_* or of the form ν_∞. This leads to the following non-deterministic version of the above representation:

3.11 Remark (Enumeration of point processes). Let N be a point processes with values in $\mathscr{N}(\mathbb{E})$. For each $n \geq 1$, there exists a vector $(Y_{n,1}, \ldots, Y_{n,n})$ and a sequence $(Y_{\infty,1}, Y_{\infty,2}, \ldots)$ of random variables with values in $\mathbb{E}$ such that:

$$N(dz) = \sum_{n=1}^{\infty} \Big[\mathbb{1}_{\{N(\mathbb{E})=n\}} \sum_{i=1}^{n} \delta_{Y_{n,i}}(dz) \Big] + \mathbb{1}_{\{N(\mathbb{E})=\infty\}} \sum_{i=1}^{\infty} \delta_{Y_{\infty,i}}(dz). \qquad \diamondsuit$$

Obviously this is a fairly complicated way to write the point process N. It would be convenient to use always the same sequence, say $(\bar{Y}_1, \bar{Y}_2, \ldots)$, independently of the number $N(\mathbb{E})$ of points. This is the reason why we introduce the *extended enumeration sequence* $(\bar{Y}_n)_{n\geq 1}$ next:

3.12 Definition (Extended enumeration). Let N be a point process and $(Y_{n,k})$, $(Y_{\infty,k})$ be the sequences of random variables given in Remark 3.11. Define the *extended enumeration* $(\bar{Y}_n)_{n\geq 1}$ by:

$$\bar{Y}_n := \begin{cases} \varnothing & N(\mathbb{E}) < n, \\ Y_{N(\mathbb{E}),n} & N(\mathbb{E}) \geq n \text{ and } N(\mathbb{E}) < \infty, \\ Y_{\infty,n} & N(\mathbb{E}) = \infty. \end{cases}$$

Equivalently, one can define $\bar{Y}_n$ also by:

$$\bar{Y}_n := \mathbb{1}_{\{N(\mathbb{E})<n\}} \varnothing + \mathbb{1}_{\{N(\mathbb{E})\geq n, N(\mathbb{E})<\infty\}} Y_{N(\mathbb{E}),n} + \mathbb{1}_{\{N(\mathbb{E})=\infty\}} Y_{\infty,n},$$

for $n \geq 1$. $\diamondsuit$

The sequence $(\bar{Y}_n)_{n\geq 1}$ allows us to represent the point process N in a consistent way as an enumeration without using indicator functions. Indeed, one has now the more compact representation:

$$N(dz) = \sum_{n=0}^{\infty} \delta_{\bar{Y}_n}(dz).$$

This relies on the following observation: Since $\varnothing$ lies outside of $\mathbb{E}$, it can never happen that $\varnothing \in A$, for sets $A \subseteq \mathbb{E}$. Hence, if there are $N(\mathbb{E}) = n$ points, then $\bar{Y}_{n+k} = \varnothing$, for $k \geq 1$, and so $\delta_{\bar{Y}_{n+k}}(A) = 0$. In this way, the artificial events $\varnothing$ vanish and we actually only sum over the n original events.

For the same reason that we have introduced $\varnothing$ to denote an event that cannot be achieved in the event space $\mathbb{E}$, we now introduce a point configuration that is not part of the point configuration space $\mathscr{N}(\mathbb{E})$:

3.13 Definition (Void point configuration). Let $\nu_\varnothing$ be some (which one is irrelevant) fixed point configuration which is impossible to achieve on $\mathbb{E}$. The following definition would do:

$$\nu_\varnothing(dz) := \delta_\varnothing(dz). \qquad \diamondsuit$$

We can treat $\nu_\varnothing$ as a *void* or *non-existent* point configuration. To end this section, let us clarify the notation concerning shifted sets of point configurations:

3.14 Notation (Shifted point configurations). Let $\mathfrak{A} \subseteq \mathscr{N}(\mathbb{E})$ and $y \in \mathbb{E}$. The following shorter notation is used to refer to a set of shifted point configurations:

$$\{\mathfrak{A} - y\} := \{\nu - y : \nu \in \mathfrak{A}\},$$

where a single point configuration ν is shifted according to:

$$(\nu - y)(dz) := \nu(dz - y).$$

In case one considers point configurations on the extended configuration space $\mathscr{N}(\bar{\mathbb{E}})$ instead, one needs to define for consistency additionally:

$$(\varnothing - y) := \varnothing, \text{ for all } y \in \mathbb{E}. \qquad \diamondsuit$$

3.3 Point Process Clusters

In this section we look not only at one point process but at a whole family of point processes. At least conceptually, the idea is that every point of the event space has an associated point process. We call this point the *center* and the point process the associated *cluster*, and the union of all clusters the *cluster swarm*. In a first step we introduce the necessary notation:

3.15 Definition (Cluster kernel). Let Φ_L, Φ_K be two point process distributions for point processes with values in $\mathscr{N}(\mathbb{E})$, i.e. distributions on the measurable space $\mathscr{N}(\mathbb{E})$ equipped with the natural σ-algebra $\mathfrak{F}(\mathbb{E})$. Call Φ_L the *cluster center* distribution and Φ_K the *cluster distribution*.

Now define the *cluster kernel* $\Phi_K(\mathfrak{B}|y)$ by:

$$\Phi_K(\mathfrak{B}|y) := \Phi_K(\mathfrak{B} - y), \text{ if } y \in \mathbb{E}, \qquad \Phi_K(\mathfrak{B}|\varnothing) := \delta_{\nu_\varnothing}(\mathfrak{B}),$$

where $y \in \bar{\mathbb{E}}$, $\mathfrak{B} \subseteq \mathscr{N}(\bar{\mathbb{E}})$. The family $\Phi_K(\cdot|y)_{y\in\mathbb{E}}$ of measures then constitutes a probability kernel. ♢

We did not mention this explicitly in the above definition, but it is clear that Φ_L and Φ_K can be extended in the natural way to $\mathscr{N}(\bar{\mathbb{E}})$ by assigning zero probability to all point configurations containing the auxiliary point $\varnothing$. For simplicity, we identify the extended distributions with the original distributions and call them by a small abuse of notation again Φ_L and Φ_K.

The above definition is one of the reasons why we have introduced the void point configuration $\nu_\varnothing$. It allows us to give a consistent definition of the cluster kernel $\Phi_K(\mathfrak{B}|y)$, whether there is actually an event at the given location y or not. This can be summarized as follows:

3.16 Remark (Void center point). According to the above definition:

$$\Phi_K\big(\mathscr{N}(\mathbb{E})\big|\varnothing\big) = 0.$$

This has the obvious interpretation: If there is no cluster center at y, i.e. if $y = \varnothing$, then there is no cluster associated to this point either. ♢

In a next step, we construct what we call a *cluster family*. The construction is done in two steps: First we take a point process L, whose points Y_i will serve as the cluster centers. Then, given a concrete realization of L, we construct for each of the center points Y_i a randomly shifted cluster $(J_i + Y_i)$. Moreover, this construction is performed in a way that all shifted clusters $(J_i + Y_i)$ are conditionally independent, given the center process L. We are going to call the collection of all shifted clusters $(J_i + Y_i)$ a *cluster family*:

3.17 Definition (Cluster process). Let Φ_L, Φ_K be given and define the cluster kernel $\Phi_K(\mathfrak{B}|y)_{y\in\mathbb{E}}$ as in Definition 3.15. Assume $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space rich enough such that the following stochastic objects can be defined:

(1) *Center process.* Let L be a point process with distribution Φ_L and $(Y_1, Y_2, \ldots)$ be an enumeration of the points of L, where $Y_i \in \bar{\mathbb{E}} = \mathbb{E} \cup \varnothing$, for $i \geq 1$. The points $Y_{i\geq 1}$ will serve as cluster centers.

(2) *Family of generating cluster.* Assume $(J_1, J_2, \ldots)$ is a sequence of independent and identically distributed point processes with distribution Φ_K. Furthermore, assume L and the sequence $J_{i\geq 1}$ are independent. In other

words, the family $\{L, J_{i\geq 1}\}$ is distributed according to:

$$\mathbb{P}\Big[L \in \mathfrak{A}, J_1 \in \mathfrak{B}_1, \ldots, J_n \in \mathfrak{B}_n\Big] = \Phi_L(\mathfrak{A}) \prod_{i=1}^{n} \Phi_K(\mathfrak{B}_i),$$

for sets $\mathfrak{A}, \mathfrak{B}_i \subseteq \mathscr{N}(\mathbb{E})$ of point configurations, for all $n \geq 1$. Then $J_{i\geq 1}$ is called the *family of generating clusters*.

(3) *Family of shifted clusters.* Consider now the family $(J_i + Y_i)_{i\geq 1}$ of shifted point processes. In order to treat also the case where $Y_i = \varnothing$, define:

$$[J_i + Y_i](dz) = \begin{cases} J_i(dz - Y_i) & \text{if } Y_i \neq \varnothing, \\ \nu_\varnothing & \text{if } Y_i = \varnothing, \end{cases}$$

for all $i \geq 1$. ♢

The shifted point process $(J_i + Y_i)$ in the above definition should be interpreted as a cluster with center in Y_i. But note that Y_i does not need to be a center in a geometrical sense.

Let us point out that this definition is consistent in the sense that the shifted point process $J_i + y$ is distributed according to $\Phi_K(\cdot|y)$. This allows us to describe the distribution of the shifted clusters $(J_i + Y_i)$ with the help of the cluster kernel $\Phi_K(\cdot|\cdot)$. Indeed, we have the following:

3.18 Proposition (Conditional independence). Let $n \geq 1$ and take a sequence of sets of point configurations $\mathfrak{B}_i \subseteq \mathscr{N}(\mathbb{E})$, for $i \geq 1$. The family of randomly displaced clusters $(J_i + Y_i)_{i\geq 1}$ has the joint conditional distribution:

$$\mathbb{P}\Big[\bigcap_{i=1}^{n} (J_i + Y_i) \in \mathfrak{B}_i \Big| L\Big] = \prod_{i=1}^{n} \Phi_K(\mathfrak{B}_i|Y_i), \tag{3.3}$$

for all $n \geq 1$. ♢

Consider in the above proposition the case where $Y_i = \varnothing$, for some $i \geq 1$. We then have $J_i + Y_i = \nu_\varnothing \notin \mathfrak{B}_i$. Hence, $\Phi_K(\mathfrak{B}_i|Y_i) = \Phi_K(\mathfrak{B}_i|\varnothing) = 0$ for all $\mathfrak{B}_i \subseteq \mathscr{N}(\mathbb{E})$. Therefore, Equation (3.3) holds even in the case where $Y_i = \varnothing$. This coincides with the interpretation that $Y_i = \varnothing$ implies that there is no i-th point, i.e. L has less than i points.

The construction above leads to the family $(J_i + Y_i)_{i\geq 1}$ of randomly displaced clusters, but this is only an intermediate step. In a next step we build upon this sequence and define a *cluster swarm* $K(\cdot|y)_{y\in\mathbb{E}}$, which contains all

these clusters in a unified form. This allows us to use a more compact notation and, more importantly, we do not need to enumerate the single clusters any more:

3.19 Definition (Cluster family). Assume L is a cluster center process with enumeration $Y_{i\geq 1}$ and $J_{i\geq 1}$ is a family of generating cluster, as in Definition 3.17.

(1) The *cluster family* $K(\cdot|\cdot)$ with clusters $J_{i\geq 1}$ and centers $Y_{i\geq 1}$ is defined as:

$$K(dz|y) := \sum_{i=1}^{\infty} \mathbb{1}_{\{Y_i=y\}}[J_i + Y_i](dz),$$

for $y \in \mathbb{E}$. Hence, $K(dz|y)_{y\in\mathbb{E}}$ is a family of point processes with values in $\mathscr{N}(\mathbb{E})$.

Now consider the superposition of all clusters in the swarm:

(2) The *compound cluster process* generated by the cluster family $K(dz|y)_{y\in\mathbb{E}}$ and directed by the center process L is defined as:

$$N(dz) := \int_{\mathbb{E}} K(dz|y)L(dy). \qquad \diamondsuit$$

Consider again the definition of a cluster family: Assume there are $n := L(\mathbb{E})$ cluster centers. Then one has $Y_i = \varnothing$, for all $i \geq n$, and therefore $\mathbb{1}_{\{Y_i=y\}} = 0$, for all $i \geq n$. This is due to the fact that $\varnothing \neq y$, for all $y \in \mathbb{E}$. Hence, the cluster family could also be defined as:

$$K(dz|y) := \sum_{i=1}^{L(\mathbb{E})} \mathbb{1}_{\{Y_i=y\}}[J_i + Y_i](dz),$$

Moreover, one can easily check that N is the indeed the superposition of all displaced clusters $(J_i + Y_i)$. Therefore, N could also be written as the sum:

$$N(dz) = \sum_{i=1}^{L(\mathbb{E})} [J_i + Y_i](dz).$$

The definition of a compound cluster process suggests that $K(dz|y)$ is evaluated for all possible center points $y \in \mathbb{E}$. But this is not the case: Indeed, since the integral is with respect to the center process L, the variable y takes only values

out of the sequence $(Y_1, Y_2, \ldots)$. Nevertheless, there is no harm in thinking that $K(dz|y)$ is an independent point process for every $y \in \mathbb{E}$.

As a consequence of the above considerations, it is clear that the cluster family $K(dz|y)_{y\in\mathbb{E}}$ can only be used in combination with its directing process L. It cannot exist as a self-contained entity.

Simple Cluster Process. By definition, each cluster is a simple point process. But we need a slightly stronger assumption: We impose the condition that not only each cluster but also the cluster process $N := \int K(dz|y)L(dy)$, i.e. the superposition of all clusters, is a simple point process. This condition is automatically satisfied in the following common case:

3.20 Remark (Absolutely continuous distributions). Let $Y_{i\geq 1}$ be an enumeration of the points of L and $W_{i\geq 1, l\geq 1}$ an enumeration of the points of the generating clusters J_i. Assume the random variables $Y_{i\geq 1}, W_{i\geq 1,l\geq 1}$ have an absolutely continuous distribution on $\mathbb{E}$. Then, since they are independent by assumption, one has:

$$\mathbb{P}[U \neq V] = 1, \qquad \text{for } U \neq V \in \Big\{Y_i, W_{i,k}, (Y_i + W_{i,k}), \text{ for } i, k \geq 1\Big\}.$$

Hence, two different random variables from the set above take distinct values with probability one. ♢

Note that the above situation is very similar to the one in Definition 2.33 and Remark 2.35, where we have introduced a weaker and a stronger notion of simple, multivariate point configurations.

Formulas for Moment Measures. Next we introduce some notation that simplifies the calculation with cluster processes:

3.21 Notation (Moment measures and clusters). Let $K(dz|y)_{y\in\mathbb{E}}$ be a cluster family directed by a center process $L(dy)$ and consider the compound cluster process $N(dz) = \int K(dz|y)L(dy)$.

(i) *Moments of clusters.* We use the following symbols to denote the moment measures of a compound cluster process or a cluster process with center $y \in \mathbb{E}$:

$$M_N(dz) := \mathbb{E}\big[N(dz)\big], \qquad M_K(dz|y) := \mathbb{E}\big[K(dz|y)\big].$$

(*ii*) *Cluster notation.* Sometimes it is more convenient to refer to a cluster process $K(\cdot|y)$ without using the cumbersome notation $(\cdot)$. To this end, the following notation is used:

$$K_y := K(\cdot|y). \tag{3.4}$$

♢

For simplicity, we consider only the univariate case in the next theorem. The multivariate case is a trivial extension:

3.22 Theorem (Factorial products of center process). Assume $(K_y)_{y\in\mathbb{E}}$ is a cluster family directed by a center process L. Furthermore, assume the assumption in Remark 3.20 is satisfied.

(*1*) *Functions.* Let $n \geq 1$ and ψ_k, for $1 \leq k \leq n$, be a sequence of non-negative functions on $\mathscr{N}(\mathbb{E})$. Assume J is a point process with distribution Φ_K, i.e. assume J has the same distribution as a cluster with center in $0 \in \mathbb{E}$. Then:

$$\mathbb{E}\Big[\int_{\mathbb{E}^n} \prod_{k=1}^{n} \psi_k(K_{y_k}) L^{[n]}(d\boldsymbol{y})\Big] = \int_{\mathbb{E}^n} \prod_{k=1}^{n} \mathbb{E}\big[\psi_k(J+y_k)\big] M_L^{[n]}(d\boldsymbol{y}). \tag{3.5}$$

(*2*) *Kernels.* Let $n \geq 1$ and $m_1, \ldots, m_n$ be a sequence of integers with $m_k \geq 1$. Denote their sum by $m := \sum_{k=1}^{n} m_k$. Assume $\gamma_k(\nu, d\boldsymbol{z}_k)$ is a kernel, where $\nu \in \mathscr{N}(\mathbb{E})$ and $\boldsymbol{z}_k \in \mathbb{E}^{m_k}$. Define the measure Γ on the space $\mathbb{E}^m$ by:

$$\Gamma\Big(\bigtimes_{k=1}^{n} d\boldsymbol{z}_k\Big) = \int_{\mathbb{E}^n} \prod_{k=1}^{n} \gamma_k(K_{y_k}, d\boldsymbol{z}_k) L^{[n]}(d\boldsymbol{y}).$$

Then, for the associated moment measure M_Γ one has:

$$M_\Gamma\Big(\bigtimes_{k=1}^{n} d\boldsymbol{z}_k\Big) = \int_{\mathbb{E}^n} \prod_{k=1}^{n} \mathbb{E}\Big[\gamma_k(J+y_k, d\boldsymbol{z}_k)\Big] M_L^{[n]}(d\boldsymbol{y}). \tag{3.6}$$

(*3*) *Moments.* Consider the compound cluster process $N := \int K_y L(dy)$. The

ordinary and factorial moment measures of N are:

$$\begin{aligned} M_N^{(n)}(d\boldsymbol{z}) &= \sum_{m=1}^{n} \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{m} M_K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r) M_L^{[m]}(d\boldsymbol{y}), \\ M_N^{[n]}(d\boldsymbol{z}) &= \sum_{m=1}^{n} \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{m} M_K^{[|S_r|]}(d\boldsymbol{z}_{S_r}|y_r) M_L^{[m]}(d\boldsymbol{y}). \end{aligned} \tag{3.7}$$

♢

As soon as we are going to deal with second order moment measures of Hawkes processes, we will apply the formulas given above in the special case $n = 2$. Thus, let us take a closer look at this special case:

3.23 Corollary (First and second order moments). Let $K_{y\in\mathbb{E}}$ be a cluster family on the univariate event space $\mathbb{E}$, directed by a center process $L(dy)$.

(1) Assume $\gamma(\nu, d\boldsymbol{z})$ is a kernel, where $\nu \in \mathscr{N}(\mathbb{E})$ and $\boldsymbol{z} \in \mathbb{E}^n$. Define the two random measures Γ_1 and Γ_2 by:

$$\begin{aligned} \Gamma_1(d\boldsymbol{z}) &= \int_{\mathbb{E}} \gamma(K_y, d\boldsymbol{z}) L(dy), \\ \Gamma_2(d\boldsymbol{z}_1 \times d\boldsymbol{z}_2) &= \int_{\mathbb{E}^2} \gamma(K_{y_1}, d\boldsymbol{z}_1)\gamma(K_{y_2}, d\boldsymbol{z}_2) L^{[2]}(d\boldsymbol{y}). \end{aligned}$$

The moment measures of these two random measures are:

$$\begin{aligned} M_{\Gamma_1}(d\boldsymbol{z}) &= \int_{\mathbb{E}} \mathbb{E}\big[\gamma(K_y, d\boldsymbol{z})\big]\, \mathbb{E}\big[L(dy)\big], \\ M_{\Gamma_2}(d\boldsymbol{z}_1 \times d\boldsymbol{z}_2) &= \int_{\mathbb{E}^2} \mathbb{E}\big[\gamma(K_{y_1}, d\boldsymbol{z}_1)\big]\, \mathbb{E}\big[\gamma(K_{y_2}, d\boldsymbol{z}_2)\big]\, \mathbb{E}\big[L^{[2]}(d\boldsymbol{y})\big]. \end{aligned} \tag{3.8}$$

(2) Consider the two point processes Γ_1 and Γ_2 given by:

$$\Gamma_1(dz) = \int_{\mathbb{E}} K_y(dz) L(dy), \quad \Gamma_2(d\boldsymbol{z}) = \int_{\mathbb{E}^2} K_{y_1}(dz_1) K_{y_2}(dz_2) L^{[2]}(d\boldsymbol{y}).$$

The corresponding moment measures of Γ_1 and Γ_2 are:

$$\begin{aligned} \mathbb{E}\big[\Gamma_1(dz)\big] &= \int_{\mathbb{E}} \mathbb{E}\big[K_y(dz)\big]\, \mathbb{E}\big[L(dy)\big], \\ \mathbb{E}\big[\Gamma_2(d\boldsymbol{z})\big] &= \int_{\mathbb{E}^2} \mathbb{E}\big[K_{y_1}(dz_1)\big]\, \mathbb{E}\big[K_{y_2}(dz_2)\big]\, \mathbb{E}\big[L^{[2]}(d\boldsymbol{y})\big]. \end{aligned} \tag{3.9}$$

♢

3.4 Poisson Processes

The Poisson process is a fundamental building block in the construction of Hawkes processes. For the time being, we consider a general multivariate event space $\mathbb{E} = (\mathbb{E}_1, \dots, \mathbb{E}_d)$, where the components $\mathbb{E}_j$ can all be different. For details, see Definition 2.2. But later, as soon as we deal with Hawkes processes, we will only be interested in multivariate event spaces of the form $\mathbb{E} = (\mathbb{E}, \dots, \mathbb{E})$, i.e. we will assume that all component spaces coincide.

3.24 Definition (Poisson process). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $\mathbb{E}$ be a univariate and $\mathbb{E}$ a multivariate event space.

(*1*) *Univariate Poisson process.* Let H be a locally-finite measure on $\mathbb{E}$, which is called the *mean measure*. Assume a point process N with values in $\mathscr{N}(\mathbb{E})$ satisfies the conditions:

(*i*) The random variable $N(A)$ is Poisson distributed with mean $H(A)$, for all bounded sets $A \subseteq \mathbb{E}$.

(*ii*) For all $n \geq 1$ and all sequences of disjoint sets $A_k \subseteq \mathbb{E}$, where $1 \leq k \leq n$, the family of random variables $N(A_1), \dots, N(A_n)$ is independent.

Then N is called a *univariate Poisson process* with mean measure H.

(*2*) *Multivariate Poisson process.* Let $\mathbb{H} \equiv (H_1, \dots, H_d)$ be a family of locally-finite measures, where each H_k is defined on $\mathbb{E}_k$. This family of measures is called the family of *mean measures*. Assume a point process N with values in $\mathscr{N}(\mathbb{E})$ satisfies the conditions:

(*i*) The random variable $N^{(k)}(A)$ is Poisson distributed with mean $H_k(A)$, for all bounded sets $A \subseteq \mathbb{E}_k$, for all $1 \leq k \leq n$.

(*ii*) Let (j_k, A_k) be pairs of indexes $j_k \in \{1 \dots, d\}$ and bounded sets $A_k \subseteq \mathbb{E}_{j_k}$, where $1 \leq k \leq n$. Assume that either $j_k \neq j_l$ or otherwise $A_k \cap A_l = \emptyset$ holds for all $k \neq l$. For all sequences (j_k, A_k) of this form, the family of random variables

$$N^{(j_1)}(A_1), \dots, N^{(j_n)}(A_n)$$

is independent.

Then N is called a *multivariate Poisson process* with mean measure family $\mathbb{H}$. ♢

3.25 Proposition (Moment measures of Poisson processes). In the univariate case let N be a Poisson process with values in $\mathscr{N}(\mathbb{E})$ and mean measure H. In the multivariate case let N be a Poisson process with values in $\mathscr{N}(\mathbb{E})$ and mean measure family $\mathbb{H} \equiv (H_1, \ldots, H_d)$. Furthermore let $\boldsymbol{j} \in \{1, \ldots, d\}^n$.

(1) The ordinary moment measures of order n in the two cases are:

$$M_N^{(n)}\Big(\bigtimes_{k=1}^n A_k\Big) = \sum_{m=1}^n \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \prod_{r=1}^m H\Big(\bigcap_{i \in S_r} A_i\Big),$$

$$M_N^{(\boldsymbol{j})}\Big(\bigtimes_{k=1}^n A_k\Big) = \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \mathbb{H}^{(\boldsymbol{j}_{\mathscr{P}})}\Big(\bigtimes_{r=1}^{|\mathscr{P}|} \bigcap_{i \in S_r} A_i\Big).$$

(2) The factorial moment measures of order n in the two cases are:

$$M_N^{[n]}\Big(\bigtimes_{k=1}^n A_k\Big) = H^{(n)}\Big(\bigtimes_{k=1}^n A_k\Big), \qquad M_N^{[\boldsymbol{j}]}\Big(\bigtimes_{k=1}^n A_k\Big) = \mathbb{H}^{(\boldsymbol{j})}\Big(\bigtimes_{k=1}^n A_k\Big). \tag{3.10}$$

The above statements hold for all bounded sets A_k, i.e. in the univariate case $A_k \subseteq \mathbb{E}$, all $1 \le k \le n$, and in the multivariate case $A_k \subseteq \mathbb{E}_{j_k}$, all $1 \le k \le n$.$\diamondsuit$

Recall that in Definition 3.19, we did not impose any assumptions on the cluster distribution. But in the case of a Hawkes process, the clusters have a Poisson distribution. Thus, we will now analyze this special case in detail. Recall also the notation introduced in Equation (3.4).

We consider only the univariate case. The multivariate case is a trivial extension.

3.26 Definition (Poisson cluster family). Let L be an arbitrary center process with values in $\mathscr{N}(\mathbb{E})$ and let H be a mean measure, i.e. a locally-finite measure on the space $\mathbb{E}$. Define first the translation-covariant family of mean measures on $\mathbb{E}$ by:

$$H_y(dz) \equiv H(dz|y) := H(dz - y),$$

for all $y \in \mathbb{E}$. Next define the cluster kernel Φ_K by:

$$\Phi_K(\mathfrak{B}|y) = \text{Poisson}_{\mathbb{E}}[H_y](\mathfrak{B}),$$

for $\mathfrak{B} \subseteq \mathscr{N}(\mathbb{E})$ and $y \in \mathbb{E}$. In other words, $\Phi_K(\cdot|y)$ is the Poisson distribution on $\mathbb{E}$ with mean measure H_y, where $y \in \mathbb{E}$ is the center of the cluster. Finally

assume $K_{y\in\mathbb{E}}$ is a cluster family with distribution Φ_K, directed by the center process L. The family $K_{y\in\mathbb{E}}$ is then called a *Poisson cluster family.* Clearly, due to the specific choice of the cluster kernel Φ_K, the clusters K_y are conditionally independent Poisson processes given the center process L. ♢

Note that for the definition of a Poisson cluster family, we do not assume that the center process L is a Poisson process. We only assume that the clusters $K_{y\in\mathbb{E}}$ have a Poisson distribution.

Let us formulate the following special case of the univariate version of Equation (3.10):

3.27 Remark (Factorial moments of a Poisson process). If J is a univariate Poisson process on the event space $\mathbb{E}$ with mean measure H, the first two factorial moment measures are:

$$\begin{aligned} M_J(dz) &= H(dz), \\ M_J^{[2]}(d\boldsymbol{z}) &= M_J(dz_1)M_J(dz_2) = H(dz_1)H(dz_2). \end{aligned} \tag{3.11}$$

♢

It is of interest to see whether and how the general moment measure formulas given in Theorem 3.22 simplify in the presence of Poisson distributions. Indeed, we can distinguish three cases: Either the center process; the generating clusters; or both have a Poisson distribution:

3.28 Corollary (Clusters that are Poisson distributed). Let $K_{y\in\mathbb{E}}$ be a cluster family directed by a center process $L(dy)$.

(*i*) *General swarm directed by a Poisson process.* If the center process L is a Poisson process with mean measure G, then ${M_N}^{[2]}(d\boldsymbol{z})$ equals

$$\int_{\mathbb{E}} M_K^{[2]}(d\boldsymbol{z}|y)G(dy) + \int_{\mathbb{E}} M_K(dz_1|y)G(dy) \int_{\mathbb{E}} M_K(dz_2|y)G(dy).$$

(*ii*) *Poisson swarm directed by a general center process.* If the cluster family $K_{y\in\mathbb{E}}$ has a Poisson distribution with mean measure H_y, then ${M_N}^{[2]}(d\boldsymbol{z})$ equals

$$\int_{\mathbb{E}} H(dz_1|y)H(dz_2|y)M_L(dy) + \int_{\mathbb{E}^2} H(dz_2|y_1)H(dz_2|y_2)M_L^{[2]}(d\boldsymbol{y}).$$

(*iii*) *Poisson swarm directed by a Poisson process.* If the cluster family $K_{y\in\mathbb{E}}$ has a Poisson distribution with mean measure H_y and the center process

L is a Poisson process with mean measure G, then ${M_N}^{[2]}(d\boldsymbol{z})$ equals

$$\int_{\mathbb{E}} H(dz_1|y)H(dz_2|y)G(dy) + \int_{\mathbb{E}} H(dz_1|y)G(dy) \int_{\mathbb{E}} H(dz_2|y)G(dy). \quad \diamondsuit$$

We now look at another special case, namely we consider the case where the cluster center process L is a homogeneous Poisson process. Recall that a point process is *homogenous* if its distribution is translation-invariant with respect to shifts of the event space $\mathbb{E}$. It turns out that in this situation the compound cluster process N is also homogeneous:

3.29 Corollary (Homogeneous center process). Assume both $K_{y\in\mathbb{E}}$ and L have a Poisson distribution. Additionally assume that the center process L is a homogeneous Poisson process with mean measure $\lambda_{\mathbb{E}}$, the Lebesgue measure on $\mathbb{E}$. Then one has:

$$M_N^{[2]}(d\boldsymbol{z}) = \int_{\mathbb{E}} H(dz_1|y)H(dz_2|y)dy + \big[H(\mathbb{E})\big]^2 \lambda_{\mathbb{E}^2}(d\boldsymbol{z}). \quad \diamondsuit$$

Note that if L is a homogeneous Poisson process on $\mathbb{E}$, then its mean measure must be of the form $\alpha\lambda_{\mathbb{E}}$, for some constant $\alpha \geq 0$. Therefore, the assumption that L has the mean measure $\lambda_{\mathbb{E}}$ is in fact not a restriction.

Proofs for Chapter 3

For the next lemma, let the assumption given in Remark 3.20 be satisfied.

3.30 Lemma (Intermediate result for factorial decomposition). Let $K(dz|y)_{y\in\mathbb{E}}$ be a cluster family directed by a center process $L(dy)$. Let $\mathscr{P} \in \mathfrak{P}_m^n$ be a partition of the set $\{1,\ldots,n\}$ and assume that $\mathscr{P} = \{S_1,\ldots,S_m\}$. Then, for all $\boldsymbol{y},\boldsymbol{z} \in \mathbb{E}^m$, one has:

$$\delta^{[m]}(\boldsymbol{y})\delta^{[n]}(\boldsymbol{z}) \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r) = \delta^{[m]}(\boldsymbol{y}) \prod_{r=1}^{m} K^{[|S_r|]}(d\boldsymbol{z}_{S_r}|y_r). \tag{3.12}$$

♢

Proof (Lemma 3.30): Let $Y_{i\geq 1}$ be an enumeration of L and $W_{i\geq 1,l\geq 1}$ an enumeration of J_i, see also Remark 3.20 concerning this notation. Now let f be a function on $\mathbb{E}$. One has the following relationship between an integral and a sum:

$$\int_{\mathbb{E}} f(z)K(dz|y) = \sum_{i,k} f(Y_i + W_{i,k})\mathbb{1}_{\{Y_i=y\}}. \tag{3.13}$$

As on other occasions, we will adopt the notation:

$$\boldsymbol{Y}^{(\boldsymbol{i})} \equiv (Y_{i_1},\ldots,Y_{i_m}), \qquad \boldsymbol{W}^{(\boldsymbol{i},\boldsymbol{k})} \equiv (W_{i_1,k_1},\ldots,W_{i_m,k_m}),$$

whenever $\boldsymbol{i},\boldsymbol{k} \in \mathbb{N}^m$.

(i) Let $\boldsymbol{i} \in \mathbb{N}^m$ and $p \neq q$. Since the union of all events forms a simple point process, see Remark 3.20, one has the following implications:

$$i_p \neq i_q \quad \Leftrightarrow \quad Y_{i_p} \neq Y_{i_q} \quad \Rightarrow \quad Y_{i_p} + W_{i_p,k_p} \neq Y_{i_q} + W_{i_q,k_q}.$$

The equivalence on the left hand side can be expressed as $\delta^{[m]}(\boldsymbol{i}) =$

$\delta^{[m]}(\boldsymbol{Y}^{(\boldsymbol{i})})$, and the implication on the right hand side as:

$$\delta^{[m]}(\boldsymbol{Y}^{(\boldsymbol{i})}) = 1 \qquad \Rightarrow \qquad \delta^{[m]}(\boldsymbol{Y}^{(\boldsymbol{i})} + \boldsymbol{W}^{(\boldsymbol{i},\boldsymbol{k})}) = 1. \tag{3.14}$$

(ii) Let $\boldsymbol{y} \in \mathbb{E}^m$ be such that $\delta^{[m]}(\boldsymbol{y}) = 1$. We claim that:

$$\delta^{[m]}(\boldsymbol{z}) \prod_{r=1}^{m} K(dz_r|y_r) = \prod_{r=1}^{m} K(dz_r|y_r). \tag{3.15}$$

This statement is shown as follows: Let f be a function $\mathbb{E}^m$. Due to (3.13), the following integral can be written as a sum:

$$\begin{aligned}\int_{\mathbb{E}^m} & f(\boldsymbol{z})\delta^{[m]}(\boldsymbol{z}) \prod_{r=1}^{m} K(dz_r|y_r) \\ &= \sum_{i_1,\ldots,i_m} \sum_{k_1,\ldots,k_m} f(\boldsymbol{Y}^{(\boldsymbol{i})} + \boldsymbol{W}^{(\boldsymbol{i},\boldsymbol{k})})\delta^{[m]}(\boldsymbol{Y}^{(\boldsymbol{i})} + \boldsymbol{W}^{(\boldsymbol{i},\boldsymbol{k})})\mathbb{1}_{\{\boldsymbol{Y}^{(\boldsymbol{i})}=\boldsymbol{y}\}}.\end{aligned}$$

One clearly has the implication:

$$\mathbb{1}_{\{\boldsymbol{Y}^{(\boldsymbol{i})}=\boldsymbol{y}\}} = 1 \qquad \Rightarrow \qquad \delta^{[m]}(\boldsymbol{Y}^{(\boldsymbol{i})}) = \delta^{[m]}(\boldsymbol{y}) = 1.$$

Now, according to (3.14), this shows that:

$$\mathbb{1}_{\{\boldsymbol{Y}^{(\boldsymbol{i})}=\boldsymbol{y}\}} = 1 \qquad \Rightarrow \qquad \delta^{[m]}(\boldsymbol{Y}^{(\boldsymbol{i})} + \boldsymbol{W}^{(\boldsymbol{i},\boldsymbol{k})}) = 1.$$

As a consequence, the $\delta^{[m]}$-function in the sum above can be left out:

$$\begin{aligned}\int_{\mathbb{E}^m} f(\boldsymbol{z})\delta^{[m]}(\boldsymbol{z}) \prod_{r=1}^{m} K(dz_r|y_r) &= \sum_{\boldsymbol{i},\boldsymbol{k}\in\mathbb{N}^m} f(\boldsymbol{Y}^{(\boldsymbol{i})} + \boldsymbol{W}^{(\boldsymbol{i},\boldsymbol{k})})\mathbb{1}_{\{\boldsymbol{Y}^{(\boldsymbol{i})}=\boldsymbol{y}\}} \\ &= \int_{\mathbb{E}^m} f(\boldsymbol{z}) \prod_{r=1}^{m} K(dz_r|y_r).\end{aligned}$$

Since this equality is satisfied for all functions f, Equation (3.15) follows.

(iii) Now let $\boldsymbol{y} \in \mathbb{E}^m$ such that $\delta^{[m]}(\boldsymbol{y}) = 1$. Moreover, let $\boldsymbol{i} \in \{1,\ldots,d\}^m$ be such that $i_r \in S_r$. We claim the following:

$$\delta^{[m]}(\boldsymbol{z}^{(\boldsymbol{i})}) \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r) = \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r). \tag{3.16}$$

This follows from:

$$\delta^{[m]}(\boldsymbol{z}^{(\boldsymbol{i})}) \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r) = \delta^{[m]}(z_{i_1},\ldots,z_{i_m}) \prod_{r=1}^{m} \prod_{j\in S_r} K(dz_j|y_r)$$

$$= \delta^{[m]}(z_{i_1},\ldots,z_{i_m}) \prod_{r=1}^{m} \Big[K(dz_{i_r}|y_r) \prod_{j\in S_r, j\neq i_r} K(dz_j|y_r) \Big]$$

$$= \Big[\delta^{[m]}(z_{i_1},\ldots,z_{i_m}) \prod_{r=1}^{m} K(dz_{i_r}|y_r) \Big] \Big[\prod_{r=1}^{m} \prod_{j\in S_r, j\neq i_r} K(dz_j|y_r) \Big].$$

Applying Equation (3.15), one gets:

$$(\ldots) = \Big[\prod_{r=1}^{m} K(dz_{i_r}|y_r) \Big] \Big[\prod_{r=1}^{m} \prod_{j\in S_r, j\neq i_r} K(dz_j|y_r) \Big] = \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r).$$

(*iv*) Adopt the notation $\boldsymbol{z}^{(\boldsymbol{i})} := (z_{i_1},\ldots,z_{i_m})$. Recall that $\delta^{[m]}(\boldsymbol{z}) = 1$ if and only if all components of $\boldsymbol{z}$ are distinct. Let the partition $\mathscr{P}$ consist of the sets $\{S_1,\ldots,S_m\}$. One can express $\delta^{[m]}(\boldsymbol{z})$ in the following form, see also Notation 2.27:

$$\delta^{[m]}(\boldsymbol{z}) = \Big[\prod_{r=1}^{m} \delta^{[|S_r|]}(\boldsymbol{z}_{S_r}) \Big] \Big[\prod_{i_1\in S_1} \cdots \prod_{i_m\in S_m} \delta^{[m]}(\boldsymbol{z}^{(\boldsymbol{i})}) \Big].$$

Assume now w.l.o.g. that $\delta^{[m]}(\boldsymbol{y}) = 1$, because otherwise Equation (3.12) is trivially satisfied. Substituting the above expression for $\delta^{[m]}(\boldsymbol{z})$, one obtains:

$$\delta^{[m]}(\boldsymbol{y})\delta^{[n]}(\boldsymbol{z}) \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r)$$

$$= \Big[\prod_{r=1}^{m} \delta^{[|S_r|]}(\boldsymbol{z}_{S_r}) \Big] \Big[\prod_{i_1\in S_1} \cdots \prod_{i_m\in S_m} \delta^{[m]}(\boldsymbol{z}^{(\boldsymbol{i})}) \Big] \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r).$$

Since $\delta^{[m]}(\boldsymbol{y}) = 1$ has been assumed, one can now apply (3.16) once for every delta-function $\delta^{[m]}(\boldsymbol{z}^{(\boldsymbol{i})})$ in the product. Hence, the statement

follows with:

$$(\ldots) = \Big[\prod_{r=1}^{m} \delta^{[|S_r|]}(\boldsymbol{z}_{S_r})\Big] \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r)$$
$$= \prod_{r=1}^{m} \Big[\delta^{[|S_r|]}(\boldsymbol{z}_{S_r}) K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r)\Big] = \delta^{[m]}(\boldsymbol{y}) \prod_{r=1}^{m} K^{[|S_r|]}(d\boldsymbol{z}_{S_r}|y_r).$$

In the last equality $\delta^{[m]}(\boldsymbol{y}) = 1$ has been used. □

3.31 Lemma (Factorial decomposition of products). Let $K_{y\in\mathbb{E}}$ be a cluster family directed by a center process L. Consider the compound cluster process $N = \int K_y L(dy)$. For all $n \geq 1$ one has:

$$N^{(n)}(d\boldsymbol{z}) = \sum_{m=1}^{n} \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r) L^{[m]}(d\boldsymbol{y}),$$
$$N^{[n]}(d\boldsymbol{z}) = \sum_{m=1}^{n} \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{m} K^{[|S_r|]}(d\boldsymbol{z}_{S_r}|y_r) L^{[m]}(d\boldsymbol{y}). \qquad (3.17)$$

♢

Proof (Lemma 3.31): The proof uses the results of Lemma 3.30 and is a simple application of the more general results given in Theorem 2.42.

(*i*) *Ordinary product measure.* First note that:

$$N^{(n)}(d\boldsymbol{z}) = \prod_{k=1}^{n} \Big[\int_{\mathbb{E}} K(dz_k|y_k) L(dy_k)\Big] = \int_{\mathbb{E}^n} \Big[\prod_{k=1}^{n} K(dz_k|y_k)\Big] L^{(n)}(d\boldsymbol{y}).$$

Next define $f_k(x_k) := K(dz_k|x_k)$, for $1 \leq k \leq n$, and apply the univariate version of (2.22):

$$(\ldots) = \sum_{m=1}^{n} \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{m} \Big[\prod_{i\in S_r} K(dz_i|y_r)\Big] L^{[m]}(d\boldsymbol{y})$$
$$= \sum_{m=1}^{n} \sum_{\mathscr{P}\in\mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r}|y_r) L^{[m]}(d\boldsymbol{y}).$$

(*ii*) *Factorial product measure.* According to the definition of the factorial

moment measure:

$$\begin{aligned} N^{[n]}(d\boldsymbol{z}) &= \delta^{[n]}(\boldsymbol{z}) N^{(n)}(d\boldsymbol{z}) \\ &= \delta^{[n]}(\boldsymbol{z}) \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \int_{\mathbb{E}^m} \Big[\prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r} | y_r) \Big] L^{[m]}(d\boldsymbol{y}). \end{aligned}$$

Since $\delta^{[m]}(\boldsymbol{y}) = \big[\delta^{[m]}(\boldsymbol{y})\big]^2$, one has by definition of $L^{[m]}$ that:

$$\begin{aligned} L^{[m]}(d\boldsymbol{y}) &= \delta^{[m]}(\boldsymbol{y}) L^{(n)}(d\boldsymbol{y}) = \big[\delta^{[m]}(\boldsymbol{y})\big]^2 L^{(n)}(d\boldsymbol{y}) \\ &= \delta^{[m]}(\boldsymbol{y}) \big[\delta^{[m]}(\boldsymbol{y}) L^{(n)}(d\boldsymbol{y})\big] = \delta^{[m]}(\boldsymbol{y}) L^{[m]}(d\boldsymbol{y}). \end{aligned}$$

Applying this equation twice below, the claimed decomposition follows due to (3.12) with:

$$\begin{aligned} N^{[n]}(d\boldsymbol{z}) &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \int_{\mathbb{E}^m} \Big[\delta^{[m]}(\boldsymbol{y}) \delta^{[n]}(\boldsymbol{z}) \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r} | y_r) \Big] L^{[m]}(d\boldsymbol{y}) \\ &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \int_{\mathbb{E}^m} \Big[\delta^{[m]}(\boldsymbol{y}) \prod_{r=1}^{m} K^{[|S_r|]}(d\boldsymbol{z}_{S_r} | y_r) \Big] L^{[m]}(d\boldsymbol{y}) \\ &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \int_{\mathbb{E}^m} \Big[\prod_{r=1}^{m} K^{[|S_r|]}(d\boldsymbol{z}_{S_r} | y_r) \Big] L^{[m]}(d\boldsymbol{y}). \end{aligned}$$

□

Proof (Theorem 3.22): Let $J_{i \geq 1}$ be the sequence of generating clusters and $Y_{i \geq 1}$ be an enumeration of the points of the center process L, see Definition 3.17.

(1) *Functions.* It suffices to show the equation for functions of the form $\psi_k(\nu) := \mathbb{1}_{\mathfrak{A}_k}(\nu)$, where $\mathfrak{A}_k \subseteq \mathscr{N}(\mathbb{E})$. First note that:

$$\begin{aligned} \mathbb{E}\Big[\int_{\mathbb{E}^n} \prod_{k=1}^{n} \psi_k(K_{y_k}) L^{[n]}(d\boldsymbol{y}) \Big] &= \mathbb{E}\Big[\int_{\mathbb{E}^n} \prod_{k=1}^{n} \mathbb{1}_{\mathfrak{A}_k}(K_{y_k}) \delta^{[n]}(\boldsymbol{y}) L^{(n)}(d\boldsymbol{y}) \Big] \\ &= \mathbb{E}\Big[\sum_{i_1, \ldots, i_n} \prod_{k=1}^{n} \mathbb{1}_{\mathfrak{A}_k}(J_{i_k} + Y_{i_k}) \delta^{[n]}(Y_{i_1}, \ldots, Y_{i_n}) \Big], \end{aligned}$$

where the sum is taken over all $i_k \geq 1$. Since L is a simple point process,

$Y_{i_p} = Y_{i_q}$ if and only if $i_p = i_q$. Hence, the above expression becomes:

$$(\dots) = \mathbb{E}\Big[\sum_{i_1,\dots,i_n} \prod_{k=1}^{n} \mathbb{1}_{\mathfrak{A}_k}(J_{i_k} + Y_{i_k})\delta^{[n]}(i_1,\dots,i_n)\Big].$$

Next observe that the sum only needs to be taken for tuples $(i_1,\dots,i_n)$ with distinct indexes, since otherwise the $\delta^{[n]}$-function vanishes. Hence, due to (3.3), one has for the previous expression:

$$\begin{aligned}(\dots) &= \sum_{\substack{i_1,\dots,i_n \\ i_k \neq i_l, k \neq l}} \mathbb{E}\Big[\prod_{k=1}^{n} \mathbb{1}_{\mathfrak{A}_k}(J_{i_k} + Y_{i_k})\Big] \\ &= \sum_{\substack{i_1,\dots,i_n \\ i_k \neq i_l, k \neq l}} \mathbb{P}\Big[\bigcap_{k=1}^{n} (J_{i_k} + Y_{i_k}) \in \mathfrak{A}_k\Big] = \sum_{\substack{i_1,\dots,i_n \\ i_k \neq i_l, k \neq l}} \mathbb{E}\Big[\prod_{k=1}^{n} \Phi_K(\mathfrak{A}_k | Y_{i_k})\Big].\end{aligned}$$

Adding again the $\delta^{[n]}$-function, this can be written as:

$$\begin{aligned}(\dots) &= \mathbb{E}\Big[\sum_{i_1,\dots,i_n} \delta^{[n]}(Y_{i_1},\dots,Y_{i_n}) \prod_{k=1}^{n} \Phi_K(\mathfrak{A}_k | Y_{i_k})\Big] \\ &= \mathbb{E}\Big[\int_{\mathbb{E}^n} \prod_{k=1}^{n} \Phi_K(\mathfrak{A}_k | y_k)\delta^{[n]}(\boldsymbol{y}) L^{(n)}(d\boldsymbol{y})\Big].\end{aligned}$$

Since $\delta^{[n]}(\boldsymbol{y})L^{(n)}(d\boldsymbol{y}) = L^{[n]}(d\boldsymbol{y})$, the claim follows with:

$$\begin{aligned}(\dots) &= \int_{\mathbb{E}^n} \Big[\prod_{k=1}^{n} \Phi_K(\mathfrak{A}_k | y_k)\Big] M_L^{[n]}(d\boldsymbol{y}) \\ &= \int_{\mathbb{E}^n} \Big[\prod_{k=1}^{n} \mathbb{E}\big[\mathbb{1}_{\mathfrak{A}_k}(J + y_k)\big]\Big] M_L^{[n]}(d\boldsymbol{y}) \\ &= \int_{\mathbb{E}^n} \prod_{k=1}^{n} \mathbb{E}\big[\psi_k(J + y_k)\big] M_L^{[n]}(d\boldsymbol{y}).\end{aligned}$$

(2) *Kernels.* It suffices to show for a sequence $B_k \subseteq \mathbb{E}^{m_k}$, where $1 \leq k \leq n$, that:

$$\mathbb{E}\Big[\int_{\mathbb{E}^n} \prod_{k=1}^{n} \gamma_k(K_{y_k}, B_k) L^{[n]}(d\boldsymbol{y})\Big] = \int_{\mathbb{E}^n} \prod_{k=1}^{n} \mathbb{E}\Big[\gamma_k(J + y_k, B_k)\Big] M_L^{[n]}(d\boldsymbol{y}).$$

But the above equality follows due to (3.5), if one defines $\psi_k(\nu) := \gamma_k(\nu, B_k)$, for $1 \leq k \leq n$.

(3) *Moments.* We prove only the equation for the ordinary moment measure. The formula for the factorial moment measure follows along the same lines. Taking the expectation on both sides of the first part of Equation (3.17), one has:

$$\begin{aligned} M_N^{(n)}(d\boldsymbol{z}) &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \mathbb{E}\Big[\int_{\mathbb{E}^m} \prod_{r=1}^{m} K^{(|S_r|)}(d\boldsymbol{z}_{S_r} | y_r) L^{[m]}(d\boldsymbol{y}) \Big], \\ &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \mathbb{E}\Big[\int_{\mathbb{E}^m} \prod_{r=1}^{m} \gamma_r(K_{y_r}, d\boldsymbol{z}_{S_r}) L^{[m]}(d\boldsymbol{y}) \Big], \end{aligned}$$

if one defines the kernels γ_r, for $1 \leq r \leq m$, by:

$$\gamma_r(\nu, d\boldsymbol{z}_{S_r}) := \nu^{(|S_r|)}(d\boldsymbol{z}_{S_r}), \text{ for } \nu \in \mathscr{M}(\mathbb{E}) \text{ and } \boldsymbol{z}_{S_r} \in \mathbb{E}^{|S_r|}.$$

Since $(J + y_r)$ is a cluster with center in y_r, the associated moment measure is $M_K(\cdot | y_r)$. This shows:

$$\mathbb{E}\big[\gamma_r(J + y_r, d\boldsymbol{z}_{S_r})\big] = \mathbb{E}\Big[(J + y_r)^{(|S_r|)}(d\boldsymbol{z}_{S_r})\Big] = M_K^{(|S_r|)}(d\boldsymbol{z}_{S_r} | y_r).$$

Due to (3.6), the statement now follows with:

$$\begin{aligned} M_N^{(n)}(d\boldsymbol{z}) &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{m} \mathbb{E}\big[\gamma_r(J + y_r, d\boldsymbol{z}_{S_r})\big] M_L^{[m]}(d\boldsymbol{y}) \\ &= \sum_{m=1}^{n} \sum_{\mathscr{P} \in \mathfrak{P}_m^n} \int_{\mathbb{E}^m} \prod_{r=1}^{n} M_K^{(|S_r|)}(d\boldsymbol{z}_{S_r} | y_r) M_L^{[m]}(d\boldsymbol{y}). \end{aligned}$$ □

Proof (Corollary 3.23): The first statement is a special case of Equation (3.6) and the second statement follows from the first one if one chooses the kernel $\gamma(K_y, dz) := K_y(dz)$. □

For the next lemma, recall Definition 3.8 for the falling factorial $r^{[n]}$, where $r, n \geq 0$ are integers.

3.32 Lemma (Factorial moments of a Poisson distribution). Let $n \geq 0$.

(1) *Poisson random variable.* For a Poisson random variable X with mean $\lambda > 0$, one has:

$$\mathbb{E}\big[X^{[n]}\big] = \mathbb{E}\Big[X(X-1)(X-2)\cdot\ldots\cdot(X-n+1)\Big] = \lambda^n. \tag{3.18}$$

This value is called the *n-th factorial moment* of X.

(2) *Poisson process.* Let $\mathbb{E}$ be a univariate event space and N a Poisson process on $\mathbb{E}$ with locally-finite mean measure H. For any bounded set $A \subseteq \mathbb{E}$ one has:

$$\mathbb{E}\big[N^{[n]}(A \times \ldots \times A)\big] = \mathbb{E}\big[N(A)^{[n]}\big] = \big[H(A)\big]^n. \tag{3.19}$$

♢

Note that $N^{[n]}$ refers to the factorial product measure, see Definition 2.36, whereas $N(A)^{[n]}$ is the n-th falling factorial of the integer $N(A)$, see Definition 3.8.

Proof (Lemma 3.32): The first statement is a well-know fact about the Poisson distribution, see e.g. Section 5.2 in [DVJ03]. It remains to show the second statement, where the two equations are treated in turn:

(i) *First equation.* Let $\nu \in \mathscr{N}(\mathbb{E})$ and assume $m := \nu(A)$. Moreover, let $(y_1, \ldots, y_m)$ be an enumeration of these m points. Since ν is simple, $y_k \neq y_l$ if and only if $k \neq l$. Using that $\nu^{[n]}(d\boldsymbol{x}) = \delta^{[n]}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x})$, one obtains:

$$\begin{aligned}\nu^{[n]}(A \times \ldots \times A) &= \int_{\mathbb{E}^n} \prod_{k=1}^{n} \big[\mathbb{1}_A(x_k)\big]\delta^{[n]}(\boldsymbol{x})\nu^{(n)}(d\boldsymbol{x}) \\ &= \sum_{i_1,\ldots,i_n} \delta^{[n]}(y_{i_1}, \ldots, y_{i_n}) = \sum_{i_1,\ldots,i_n} \delta^{[n]}(i_1, \ldots, i_n),\end{aligned}$$

where the sum is taken over all $1 \leq i_k \leq m$. According to the definition, $\delta^{[n]}(i_1, \ldots, i_n) = 1$ holds if and only if $(i_1, \ldots, i_n)$ is a vector without any duplicated components. Since there are exactly $m^{[n]}$ tuples $(i_1, \ldots, i_n)$ with distinct elements $i_k \in \{1, \ldots, m\}$, it follows that:

$$\nu^{[n]}(A \times \ldots \times A) = m^{[n]} = \nu(A)^{[n]}.$$

Now substitute N instead of ν and the first equality follows.

(ii) *Second equation.* Because $N(A)$ has a Poisson distribution with mean

$H(A)$, the second equality follows with (3.18). □

The following proof treats only the multivariate version. The univariate case is treated along the same lines:

Proof (Proposition 3.25): The formulas for the ordinary moment measures are based on the ones for the factorial moment measures. Thus, the proof of the second statement logically precedes the proof of the first statement:

(1) *Ordinary moment measures.* In a first step, apply the multivariate version of (2.22), such that:

$$M_N^{(\boldsymbol{j})}\Big(\bigtimes_{k=1}^{n} A_k\Big) = \mathbb{E}\Big[\int_{\mathbb{E}^{(\boldsymbol{j})}} \prod_{k=1}^{n} \mathbb{1}_{A_k}(y_k) N^{(\boldsymbol{j})}(d\boldsymbol{y})\Big]$$
$$= \mathbb{E}\Big[\sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} \prod_{r=1}^{|\mathscr{P}|} \Big[\prod_{i \in S_r} \mathbb{1}_{A_i}(y_r)\Big] N^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{y})\Big].$$

Due to Fubini's theorem and the multivariate version of (3.10):

$$(\ldots) = \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} \prod_{r=1}^{|\mathscr{P}|} \Big[\prod_{i \in S_r} \mathbb{1}_{A_i}(y_r)\Big] M_N^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{y})$$
$$= \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \int_{\mathbb{E}^{(\boldsymbol{j}_{\mathscr{P}})}} \prod_{r=1}^{|\mathscr{P}|} \mathbb{1}_{\{\bigcap_{i \in S_r} A_i\}}(y_r) M_N^{[\boldsymbol{j}_{\mathscr{P}}]}(d\boldsymbol{y})$$
$$= \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} M_N^{[\boldsymbol{j}_{\mathscr{P}}]}\Big(\bigtimes_{r=1}^{|\mathscr{P}|} \bigcap_{i \in S_r} A_i\Big) = \sum_{\mathscr{P} \ll \mathscr{P}_{\boldsymbol{j}}} \mathbb{H}^{(\boldsymbol{j}_{\mathscr{P}})}\Big(\bigtimes_{r=1}^{|\mathscr{P}|} \bigcap_{i \in S_r} A_i\Big).$$

(2) *Factorial moment measures.* Let $\boldsymbol{j} \in \{1, \ldots, d\}^n$ and assume that $A_i \subseteq \mathbb{E}_{j_i}$, for $1 \leq i \leq n$.

(i) *First reduction.* By reordering the components of $\boldsymbol{j}$, one can assume w.l.o.g. that $\boldsymbol{j}$ is of the form:

$$\boldsymbol{j} = (1, \ldots, 1\ ,\ 2, \ldots, 2 \quad \ldots \quad d, \ldots, d) = (\boldsymbol{j}_1, \ldots, \boldsymbol{j}_d),$$

where $\boldsymbol{j}_k = (k, \ldots, k)$ is a vector of length n_k and $n = \sum_{k=1}^{d} n_k$. In the same way as $\boldsymbol{j}$ has been reordered and grouped, arrange now the sequence $A_{i=1,\ldots,n}$ in d groups. Assume after reordering and relabeling,

the sequence $A_{i=1,\dots,n}$ is of the form:

$$B_{1,1},\dots,B_{1,n_1} \quad \dots \quad B_{k,1},\dots,B_{1,n_k} \quad \dots \quad B_{d,1},\dots,B_{1,n_d}$$

where $B_{k,i} \subseteq \mathbb{E}_k$, for $1 \le i \le n_k$ and $1 \le k \le n$. In the calculation below, apply (2.14) and use that different components of N are independent:

$$\begin{aligned} M^{[\boldsymbol{j}]}\Big(\bigtimes_{i=1}^{n} A_i\Big) &= M^{[\boldsymbol{j}_1\cdots\boldsymbol{j}_d]}\Big(\bigtimes_{k=1}^{d}\bigtimes_{i=1}^{n_k} B_{k,i}\Big) = \mathbb{E}\Big[N^{[\boldsymbol{j}_1\cdots\boldsymbol{j}_d]}\Big(\bigtimes_{k=1}^{d}\bigtimes_{i=1}^{n_k} B_{k,i}\Big)\Big] \\ &= \mathbb{E}\Big[\prod_{k=1}^{d} N^{[\boldsymbol{j}_k]}\Big(\bigtimes_{i=1}^{n_k} B_{k,i}\Big)\Big] = \prod_{k=1}^{d} \mathbb{E}\Big[N^{[\boldsymbol{j}_k]}\Big(\bigtimes_{i=1}^{n_k} B_{k,i}\Big)\Big] \end{aligned} \tag{3.20}$$

(*ii*) *Second reduction.* Now fix $1 \le k \le d$ and concentrate on the sequence $B_{k,1},\dots,B_{k,n_k}$ only. Note that one can write every product of the form $B_{k,1}\times\ldots\times B_{k,n_k}$ as the union of products $\tilde{B}_{k,1}\times\ldots\times\tilde{B}_{k,n_k}$, where $\tilde{B}_{k,i}$ and $\tilde{B}_{k,j}$, for $i \ne j$, are either identical or disjoint. Hence, assume w.l.o.g. that the sequence $B_{k,i=1,\dots,n_k}$ satisfies:

$$B_{k,i} = B_{k,j} \qquad \text{or} \qquad B_{k,i} \cap B_{k,j} = \emptyset, \qquad \text{if } i \ne j.$$

Once more, reorder and group the sequence $(B_{k,i})$: Assume there are r_k mutually disjoint sets $C_{k,l}$, for $1 \le l \le r_k$, such that the sequence $(B_{k,i})$ is after reordering and relabeling of the form:

$$C_{k,1},\dots,C_{k,1} \quad \dots \quad C_{k,l},\dots,C_{k,l} \quad \dots \quad C_{k,r_k},\dots,C_{k,r_k}$$

Let $m_{k,l}$ be the length of each group $C_{k,l},\dots,C_{k,l}$, so that $n_k = \sum_{l=1}^{r_k} m_{k,l}$. In the same way as the sequence $B_{k,i=1,\dots,n_k}$ has been reordered, arrange $\boldsymbol{j}_k$ into r_k groups:

$$\boldsymbol{j}_k = (\boldsymbol{j}_{k,1},\dots,\boldsymbol{j}_{k,r_k}), \text{ where } \boldsymbol{j}_{k,l} = (k,\dots,k) \text{ with length } m_{k,l}.$$

Now continue with Equation (3.20) and use (2.13). Writing N instead of

$N^{(k)}$ for readability, one gets:

$$\begin{aligned}\mathbb{E}\Big[N^{[\boldsymbol{j}_k]}\Big(\bigtimes_{i=1}^{n_k} B_{k,i}\Big)\Big] &= \mathbb{E}\Big[N^{[n_k]}\Big(\bigtimes_{i=1}^{n_k} B_{k,i}\Big)\Big] = \mathbb{E}\Big[N^{[n_k]}\Big(\bigtimes_{l=1}^{r_k} C_{k,l}^{m_{k,l}}\Big)\Big] \\ &= \mathbb{E}\Big[N^{[n_k]}\Big(\bigtimes_{l=1}^{r_k} C_{k,l}^{m_{k,l}}\Big)\Big] = \mathbb{E}\Big[N^{[n_k]}\Big(\bigtimes_{l=1}^{r_k}\bigtimes_{m_{k,l}} C_{k,l}\Big)\Big] \\ &= \mathbb{E}\Big[\prod_{l=1}^{r_k} N^{[m_{k,l}]}\Big(\bigtimes_{m_{k,l}} C_{k,l}\Big)\Big] = \mathbb{E}\Big[\prod_{l=1}^{r_k} N^{[m_{k,l}]}(C_{k,l}^{m_{k,l}})\Big].\end{aligned}$$

Since $C_{k,l}$ are pairwise disjoint, $N(C_{k,i})$ and $N(C_{k,j})$ are independent, for $i \neq j$. Hence, $N^{[m_{k,i}]}(C_{k,i}{}^{m_{k,i}})$ and $N^{[m_{k,j}]}(C_{k,j}{}^{m_{k,j}})$ are also independent, for $i \neq j$. Due to (3.19), one now obtains:

$$\mathbb{E}\Big[N^{[\boldsymbol{j}_k]}\Big(\bigtimes_{i=1}^{n_k} B_{k,i}\Big)\Big] = \prod_{l=1}^{r_k} \mathbb{E}\Big[N^{[m_{k,l}]}(C_{k,l}^{m_{k,l}})\Big] = \prod_{l=1}^{r_k} \big[H_k(C_{k,l})\big]^{m_{k,l}}. \tag{3.21}$$

(*iii*) *Final step.* Finally, take Equation (3.20), and due to (3.21), the statement follows with:

$$\begin{aligned}M^{[\boldsymbol{j}]}\Big(\bigtimes_{i=1}^{n} A_i\Big) &= \prod_{k=1}^{d} \mathbb{E}\Big[N^{[\boldsymbol{j}_k]}\Big(\bigtimes_{i=1}^{n_k} B_{k,i}\Big)\Big] = \prod_{k=1}^{d}\prod_{l=1}^{r_k} \big[H_k(C_{k,l})\big]^{m_{k,l}} \\ &= \prod_{k=1}^{d}\prod_{l=1}^{r_k} \mathbb{H}^{(\boldsymbol{j}_{k,l})}(C_{k,l}^{m_{k,l}}) = \prod_{k=1}^{d} \mathbb{H}^{(\boldsymbol{j}_k)}\Big(\bigtimes_{l=1}^{r_k} C_{k,l}^{m_{k,l}}\Big) \\ &= \prod_{k=1}^{d} \mathbb{H}^{(\boldsymbol{j}_k)}\Big(\bigtimes_{i=1}^{n_k} B_{k,i}\Big) = \mathbb{H}^{(\boldsymbol{j})}\Big(\bigtimes_{i=1}^{n} A_i\Big).\end{aligned}$$

□

Proof (Corollary 3.28): Let $n = 2$ and consider the following special case of (3.7):

$$M_N^{[2]}(d\boldsymbol{z}) = \int_{\mathbb{E}} M_K^{[2]}(d\boldsymbol{z}|y) M_L(dy) + \int_{\mathbb{E}^2} M_K(dz_1|y_1) M_K(dz_2|y_2) M_L^{[2]}(d\boldsymbol{y}). \tag{3.22}$$

It remains to specialize the above equation in the three cases:

(*i*) *General family directed by a Poisson process.* If L is a Poisson process

with mean measure G, then:

$$M_L(dy) = G(dy), \qquad M_L^{[2]}(d\boldsymbol{y}) = G(dy_1)G(dy_2).$$

Now substitute this in (3.22).

(*ii*) *Poisson family directed by a general center process.* If $K_{y\in\mathbb{E}}$ is a Poisson cluster family with mean measure family $H_{y\in\mathbb{E}}$, then:

$$M_K(dz|y) = H(dz|y), \qquad M_K^{[2]}(d\boldsymbol{z}|y) = H(dz_1|y)H(dz_2|y).$$

Again substitute this in (3.22).

(*iii*) *Poisson cluster family directed by a Poisson process.* This is the combination of the previous two special cases. Hence, the statement follows, after substitution of the four above equalities in (3.22). □

Proof (Corollary 3.29): If L is a homogeneous Poisson process with mean measure $\lambda_{\mathbb{E}}$, then:

$$M_L(dy) = dy \qquad \text{and} \qquad M_L^{[2]}(d\boldsymbol{y}) = d\boldsymbol{y}.$$

Substituting this in (3.22) yields:

$$M_N^{[2]}(d\boldsymbol{z}) = \int_{\mathbb{E}} H(dz_1|y)H(dz_2|y)dy + \int_{\mathbb{E}} H(dz_1|y)dy \int_{\mathbb{E}} H(dz_2|y)dy.$$

For any bounded set $A \subseteq \mathbb{E}$ note that:

$$\begin{aligned}\int_{\mathbb{E}} H(A|y)dy &= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}} \mathbb{1}_A(z)H(dz-y)\Big]dy = \int_{\mathbb{E}}\Big[\int_{\mathbb{E}} \mathbb{1}_A(z+y)H(dz)\Big]dy \\ &= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}} \mathbb{1}_A(z+y)dy\Big]H(dz) = \int_{\mathbb{E}} \lambda_{\mathbb{E}}(A)H(dz) = H(\mathbb{E})\lambda_{\mathbb{E}}(A).\end{aligned}$$

Therefore $\int_{\mathbb{E}} H(dz|y)dy = H(\mathbb{E})\lambda_{\mathbb{E}}(dz)$, and after substitution in the above equation, the claim follows:

$$M_N^{[2]}(d\boldsymbol{z}) = \int_{\mathbb{E}} H(dz_1|y)H(dz_2|y)dy + \big[H(\mathbb{E})\big]^2\lambda_{\mathbb{E}}(dz_1)\lambda_{\mathbb{E}}(dz_2). \qquad \square$$

Chapter 4

Hawkes Processes

There are two conceptually different ways how a Hawkes process can be defined and each one has its advantages and disadvantages. In this chapter, we describe a Hawkes process as a recursive Poisson cluster process. The advantage of this approach is that the results from the previous chapter concerning Poisson cluster families can directly be applied. Hence, this representation is especially suitable to derive the first and second order moment measures. The corresponding results are the topic of the next chapter.

The other common way to describe a Hawkes process is in terms of intensity processes. Chapter 6 is dedicated to that approach. As we will see, specifying a Hawkes process in terms of its intensity process has also its advantages, e.g. if one wants to calculate its likelihood function.

In summary, one can look at Hawkes processes from different perspectives, either within the framework of Poisson cluster processes or within the theory of intensity processes. Each point of view brings with it its own set of powerful tools. I think it is a nice feature of Hawkes processes, that techniques from different areas can be used to analyze them; but one has to take the appropriate representation for the question at hand.

4.1 Motivation and Objectives

I will define Hawkes processes in several steps: First I define *Hawkes trees*, which are basically Hawkes processes with one fixed, deterministic immigrant. This definition is motivated by Section 6.3 of [DVJ03]. But in contrast to this reference, I will put more emphasis on the construction of the underlying

stochastic objects.

A Hawkes tree is obtained with the following recursive procedure: Assume there are finitely many random center points, say $(X_1, \dots, X_n)$. Take a sequence of iid Poisson processes $(K_{X_1}, \dots, K_{X_n})$, where K_{X_k} has mean measure H_{X_k}. The measures $H_{x \in \mathbb{E}}$ are called *transfer measures*. Then take the union of all events in the sequence $(K_{X_1}, \dots, K_{X_n})$ and repeat the recursive procedure until there are no new points, i.e. until $n = 0$. Finally, collect all points. If one starts with $n = 1$, $X_1 = 0$, the above construction yields a Hawkes tree.

Conceptually, one can assume that for every possible location $x \in \mathbb{E}$ there is an associated independent Poisson processes K_x. The sequence $(K_{X_1}, \dots, K_{X_n})$ of Poisson processes is then obtained by selecting the clusters associated to the center points $(X_1, \dots, X_n)$.

One of the purposes of the previous chapter was to explain in detail how one can select a countable sequence of clusters with random indexes out of a conceptually uncountably infinite family of independent clusters. At first, it might seem that one has to construct the underlying probability space by successive conditioning. But one of the nice consequences is that one only needs an iid sequence of point processes, which I call the *driving process*, see Remark 4.12.

Decomposition Formulas. The last part of this chapter serves as a preparation for the next chapter. Specifically, the decomposition formulas given in Lemma 4.24 will be used again later. Let me quickly explain the purpose of these somewhat obscure formulas:

In order to calculate moment measures of Hawkes processes, one basically needs to calculate the expectation of integrals of the form $\int \prod_{k=1}^{n} K_{y_k} L^{(n)}(d\boldsymbol{y})$. Unfortunately, since the events of L are not independent, this is in general not trivial. But if the integral is of the form $\int \prod_{k=1}^{n} K_{y_k} L^{[n]}(d\boldsymbol{y})$, then some sort of distributivity law for the expectation applies; the precise formulation is given in Theorem 3.22.

Hence, the notational small difference between $L^{(n)}$ and $L^{[n]}$ makes a big difference if expectations need to be calculated. The main idea is now to decompose the second order product measures of a Hawkes process in such a way that only integrals with an integrator of the form $L^{[n]}(d\boldsymbol{y})$ appears. Decompositions that satisfy this requirement are given in Lemma 4.24.

4.2 Cluster Kernels

We first summarize some well-known facts from linear algebra:

4.1 Definition (Matrix norms). Consider the space $\mathbb{R}^{d\times d}$ of real-valued, $d \times d$-dimensional matrices.

(*i*) A *matrix norm* $\|\cdot\|$ is a vector norm on the space $\mathbb{R}^{d\times d}$, i.e. a real-valued function on $\mathbb{R}^{d\times d}$ with the following three properties: For all matrices $A, B \in \mathbb{R}^{d\times d}$ and all $\alpha \in \mathbb{R}$ it holds that: $\|A\| \geq 0$; $\|A\| = 0$ if and only if $A = 0$. Moreover, $\|\alpha A\| = |\alpha|\|A\|$ and $\|A + B\| \leq \|A\| + \|B\|$.

(*ii*) A vector-norm $\|\cdot\|$ on the space $\mathbb{R}^{d\times d}$ of square matrices is called *sub-multiplicative* if it satisfies:

$$\|AB\| \leq \|A\| \cdot \|B\|, \text{ for all } A, B \in \mathbb{R}^{d\times d}.$$

(*iii*) Let $|\cdot|$ be a vector-norm on $\mathbb{R}^d$. The *operator-norm* $\|\cdot\|_*$ induced by $|\cdot|$ is the matrix-norm defined by:

$$\|A\|_* := \sup\Big\{|A\boldsymbol{v}|/|\boldsymbol{v}| : \boldsymbol{v} \in \mathbb{R}^d, \boldsymbol{v} \neq \mathbf{0}\Big\},$$

for $A \in \mathbb{R}^{d\times d}$. The operator-norm $\|\cdot\|_*$ is always sub-multiplicative. ♢

In order to analyze geometric convergence of matrices the following definition is very useful. For a text book treatment of the spectral radius see also Definition 10.10 in [Rud91] or Definition 5.6.8 in [HJ85].

4.2 Definition (Spectral radius). Let $A \in \mathbb{R}^{d\times d}$ be a matrix and order the d eigenvalues of A according to:

$$|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_d|.$$

Note that $\lambda_k \in \mathbb{C}$, in general. Define the *spectrum* $\Lambda(A)$ by:

$$\Lambda(A) := \big\{\lambda_k; k = 1, \ldots d\big\}.$$

The *spectral radius* $\mathrm{Spr}(A)$ is then defined as:

$$\mathrm{Spr}(A) := \max\Big\{|\lambda| : \lambda \in \Lambda(A)\Big\} = |\lambda_1|.$$

The spectral radius is a sub-multiplicative matrix-norm. ♢

Later, we deal with the branching matrix Q of a Hawkes process, which has non-negative components. The following specific property of non-negative matrices will not be used later on, but is still mentioned for completeness:

4.3 Remark. Assume A is a $d\times d$-matrix with non-negative components. As a consequence of the Perron-Frobenius theory, it holds that: The sequence $\lambda_{1,\dots,d}$ in Definition 4.2 can be reordered such that:

$$\lambda_1 \geq 0.$$

In other words, A has a non-negative eigenvalue that is as least as large as the absolute value of any other eigenvalue. ♢

Proof: The Perron-Frobenius Theorem is usually formulated for matrices with positive components, but there is a corresponding extension to non-negative matrices. See Theorem 3 in Section XIII.3 of [Gan74] or Theorem 4.2 in Section I.1.4 of [Min88] or also Theorem 8.3.1 in [HJ85]. □

The next proposition shows that any sub-multiplicative matrix norm and the spectral radius are related:

4.4 Proposition (Gelfand's formula). Let $\|\cdot\|$ be a sub-multiplicative matrix norm on $\mathbb{R}^{d\times d}$. Then, for all $A \in \mathbb{R}^{d\times d}$, one has:

$$\mathrm{Spr}(A) = \lim_{n\to\infty} \big(\|A^n\|\big)^{1/n}, \qquad \diamondsuit$$

Proof: See Theorem 10.13 in [Rud91] or Corollary 5.6.14 in [HJ85]. □

Let us make some remarks about the convergence of a sequence of matrices:

4.5 Remark (Convergence of matrices). Fix a matrix norm $\|\cdot\|$ on $\mathbb{R}^{d\times d}$ and consider a sequence of matrices $(A_n)_{n\geq 1}$ and some other matrix A. The canonical definition of convergence is:

$$\lim_{n\to\infty} A_n = A \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \|A_n - A\| = 0.$$

It is important to note that the choice of the matrix norm $\|\cdot\|$ is irrelevant, since on a finite-dimensional vector space all norms are equivalent. In other words, if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two vector norms on $\mathbb{R}^{d\times d}$, then:

$$\lim_{n\to\infty} \|A_n - A\|_1 = 0 \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \|A_n - A\|_2 = 0.$$

As a consequence, one can give the following equivalent definition for convergence: For a sequence of matrices $(A_n)_{n\geq 1}$ and some other matrix A one has:

$$\lim_{n\to\infty} A_n = A \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} (A_n)_{ij} = A_{ij}, \text{ for all } 1 \leq i,j \leq n.$$

Hence, the sequence $(A_n)_{n\geq 1}$ converges to A if and only if all components $A_{n,ij}$ converge to A_{ij}. ♢

A consequence of Gelfand's formula is the following convergence criterium:

4.6 Remark (Geometric convergence). Let A be a matrix in $\mathbb{R}^{d\times d}$ and consider the sequence $A^1, A^2, \ldots$ of n-fold products of A. Then:

$$\mathrm{Spr}(A) < 1 \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} A^n \to \mathbb{O}_d,$$

where $\mathbb{O}_d$ denotes the matrix consisting of zeros. Consider now the sequence $S_n = \sum_{k=0}^{n} A^k$, for $n \geq 0$, where by convention $A^0 := \mathbb{1}_d$. Then, under the assumption that $\mathrm{Spr}(A) < 1$, the sequence $(S_n)_{n\geq 0}$ is convergent and the limit is given by:

$$\lim_{n\to\infty} S_n = \sum_{k=0}^{\infty} A^k = (\mathbb{1}_d - A)^{-1}.$$

The fact that $(\mathbb{1}_d - A)$ is invertible is part of this statement. ♢

Proof: The first part of the above remark corresponds to Theorem 5.6.12 in [HJ85]. The second part corresponds to Corollary 5.6.16 and the exercise following this corollary in the same reference. □

4.7 Definition (Transfer measure matrix). Let $\mathbb{E}$ be a univariate event space and H_{ij} be a family of finite measures on $\mathbb{E}$, for $1 \leq i,j \leq d$.

(i) For brevity of notation, introduce the matrix $\mathbb{H}$ of measures:

$$\mathbb{H} := \big\{H_{ij}, \text{ for } 1 \leq i,j \leq d\big\}.$$

Moreover, define the matrix of shifted measures $\mathbb{H}_y$ by:

$$\mathbb{H}_y(dz) := \mathbb{H}(dz - y), \qquad H_{jk,y}(dz) := H_{jk}(dz - y),$$

for all $y \in \mathbb{E}$. The left hand side is in matrix notation; the right hand side in component-wise notation.

(*ii*) Define the *identity matrix* $\mathbb{1}$ and the *shifted identity matrix* $\mathbb{1}_y$ by:

$$\mathbb{1}_{ij}(dz) := \delta_{ij}\delta_0(dz) \qquad \text{and} \qquad [\mathbb{1}_y]_{ij}(dz) := \delta_{ij}\delta_y(dz),$$

for all $y \in \mathbb{E}$. ♢

Next we introduce the notion of multivariate convolutions. Note that the identity matrix $\mathbb{1}$ is the unit-element, i.e. one has $\mathbb{1} * \mathbb{H} = \mathbb{H}$ and similarly $\mathbb{1}_y * \mathbb{H} = \mathbb{H}_y$, for every matrix $\mathbb{H}$ of measures.

4.8 Definition (Multidimensional convolution). Let $\mathbb{G}$ and $\mathbb{H}$ be two matrices of measures. The *convolution* $\mathbb{G} * \mathbb{H}$ is again a matrix of measures, where the (j,k)-th component, for $1 \le j,k \le n$, is defined as:

$$\begin{aligned}[\mathbb{G} * \mathbb{H}]_{jk}(dz) &:= \sum_{m=1}^{d} \int_{\mathbb{E}} G_{jm}(dy) H_{mk}(dz - y) \\ &:= \sum_{m=1}^{d} \int_{\mathbb{E}} G_{jm}(dz - y) H_{mk}(dy),\end{aligned}$$

Moreover, define recursively the n-th convolution of $\mathbb{H}$ with itself by:

$$\mathbb{H}^{*0} := \mathbb{1} \qquad \text{and} \qquad \mathbb{H}^{*n} := \mathbb{H} * \mathbb{H}^{*(n-1)}. \qquad ♢$$

Later on, $\mathbb{H}$ will be the family of transfer measures of a Hawkes process. Closely related to the transfer measure matrix are the so-called associated *cluster kernels* and the *branching matrix*:

4.9 Definition (Cluster kernel). Let $\mathbb{H}$ be a matrix of finite measures.

(*i*) The *exclusive* $\mathbb{U}^-$ and *inclusive* $\mathbb{U}^+$ cluster kernels generated by $\mathbb{H}$ are:

$$\mathbb{U}^+ := \sum_{n=0}^{\infty} \mathbb{H}^{*n} \qquad \text{and} \qquad \mathbb{U}^- := \sum_{n=1}^{\infty} \mathbb{H}^{*n}. \tag{4.1}$$

Let $y \in \mathbb{E}$ be given. In the same way as for the matrix $\mathbb{H}$, define the shifted versions of the two cluster kernels by:

$$\mathbb{U}_y^+(dz) := \mathbb{U}^+(dz - y), \qquad \mathbb{U}_y^-(dz) := \mathbb{U}^-(dz - y),$$

(*ii*) Define the *branching matrix* Q associated to the measure matrix $\mathbb{H}$ by:

$$Q_{jk} = H_{jk}(\mathbb{E}), \tag{4.2}$$

for $1 \leq j,k \leq d$. Hence, Q is a $d \times d$-matrix and has non-negative components. ♢

Note that without further assumptions, the two cluster kernels $\mathbb{U}^-$ and $\mathbb{U}^+$ are not necessarily finite measures. But there is a simple sufficient condition:

4.10 Proposition (Some formulas for cluster kernels). Let $\mathbb{H}$ be a matrix of finite measures and $\mathbb{U}^-$ and $\mathbb{U}^+$ be the associated cluster kernels. The following statements are given at the same time in matrix notation and in component-wise notation:

(1) The additive relation between $\mathbb{U}^-$ and $\mathbb{U}^+$ is given by:

$$\mathbb{1}_y + \mathbb{U}_y^- = \mathbb{U}_y^+, \qquad \delta_{jk}\delta_y(dz) + U_{jk,y}^-(dz) = U_{jk,y}^+(dz). \tag{4.3}$$

The multiplicative relation between $\mathbb{U}^-$ and $\mathbb{U}^+$ is given by:

$$\mathbb{U}^+ * \mathbb{H}_y = \mathbb{U}_y^-, \qquad \sum_{k=1}^{d} \int_{\mathbb{E}} U_{jk,x}^+(dz) H_{km,y}(dx) = U_{jm,y}^-(dz). \tag{4.4}$$

(2) Assume that $\mathrm{Spr}(Q) < 1$. Then $\mathbb{U}^-$ and $\mathbb{U}^+$ are finite and one has for their total masses:

$$\mathbb{U}^+(\mathbb{E}) = (\mathbb{1}_d - Q)^{-1}, \qquad \mathbb{U}^-(\mathbb{E}) = (\mathbb{1}_d - Q)^{-1} - \mathbb{1}_d. \tag{4.5}$$

♢

4.3 Univariate Hawkes Processes

We could directly give the definition of a multivariate Hawkes process. But it is more convenient to introduce the definitions and notations first in the univariate case and then to extend them to the multivariate case. Thus, take a univariate event space $\mathbb{E}$. We first define so-called Hawkes trees:

4.11 Definition (Univariate Hawkes tree). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Moreover, let H be a finite measure on $\mathbb{E}$ and $H_{y\in\mathbb{E}}$ be the associated family of shifted measures. Let $\Phi_{K,y}$ be the family of Poisson distributions on $\mathscr{N}(\mathbb{E})$ with mean measures H_y.

The construction of a Hawkes tree starts with the *root node* and then adds the *descendant generations* in an inductive procedure:

(*i*) The *process of generation* 0 consists of only one event, the *root node* at location $0 \in \mathbb{E}$:

$$L_{0|0}(dz) := \delta_0(dz).$$

Assume, in the sense of an induction over n, the following processes have already been defined for some $(n-1) \geq 0$:

(*ii*) Let $K_n(dz|x)_{x\in\mathbb{E}}$ be a *cluster family* with distribution Φ_{K+x}, directed by the center process $L_{n-1|0}(dz)$.

(*iii*) Define the point process $L_{n|0}$ with values in $\mathscr{N}(\mathbb{E})$ by:

$$L_{n|0}(dz) := \int_{\mathbb{E}} K_n(dz|x) L_{n-1|0}(dx). \tag{4.6}$$

The process $L_{n|0}$ contains all descendants of generation n.

Repeat the above procedure for all $n \geq 0$, so that one obtains a sequence $(L_{n|0})_{n\geq 0}$ of point processes. A Hawkes tree is now the superposition of all these processes. In fact, we define the following two versions of a Hawkes tree:

(*iv*) One distinguishes between the *exclusive* tree, which does not contain the root node, and the *inclusive* tree, which does contain the root node:

$$N^-_{|0} := \sum_{r=1}^{\infty} L_{r|0} \qquad \text{and} \qquad N^+_{|0} := \sum_{r=0}^{\infty} L_{r|0}. \qquad \diamondsuit$$

Note that in the above procedure, one has to extend the probability space if necessary, in order to accommodate the cluster family $K_n(dz|x)$. But this extension is not based on a sequential conditioning procedure in the sense of Ionesu-Tulcea's Theorem. Indeed, all that the construction needs is an iid sequence of point processes, from which the cluster families and thus the generation processes can be constructed. One could therefore call this underlying sequence the *driving process*. Let us explain this in more detail:

4.12 Remark (Driving process). Assume one has already constructed the process $L_{n-1|0}$. In order to define the process $L_{n|0}$, the following ingredients are needed:

(*i*) *Generating clusters of generation n.* A sequence $(J_{n,i})_{i\geq 1}$ of independent and identically-distributed point processes with distribution Φ_K, independent of all other probabilistic objects defined so far.

(ii) *Cluster swarm of generation n.* With $(J_{n,i})_{i\geq 1}$ one can now construct a cluster family $K_n(dz|x)$, as explained in Definition 3.17. This cluster family is constructed by random displacement of the generating clusters $J_{n,i}$ and involves no other stochastic objects.

In summary, we see that the only random object needed to define a Hawkes process is the double-sequence $(J_{n,i})_{n\geq 1,i\geq 1}$ of independent and identically-distributed generating clusters. This justifies that we call the family $(J_{n,i})$ the underlying *driving process.* ♢

One should note that $N_{|0}{}^-$ and $N_{|0}{}^+$ are not necessarily locally-finite, i.e. the realizations do not need to lie in the space $\mathscr{N}(\mathbb{E})$ of locally-finite point configurations. In consequence, $N^\pm$ are not point processes in the sense used so far. But later will give sufficient conditions, so that this construction leads indeed to well-defined point processes.

Often we are not interested in the full Hawkes tree but in a subtree starting at some given node. The formal definition of a subtree is very similar to the definition of the full tree. And in the same way as for the full Hawkes tree, the definition is again a recursive procedure. But this time we do not need to extend the probability space and add in each iteration an independent cluster family. Instead we reuse the same cluster family $K_{n,x}$ that has been used in the definition of a full Hawkes tree, since otherwise we would only get an independent copy of the subtree.

4.13 Definition (Univariate Hawkes subtree). Fix a generation $n_0 \geq 0$ and the location of the root node $y \in \mathbb{E}$.

(i) Define the *root process* $L_{n_0|n_0}$ of the subtree originating in generation n_0 at location y by:

$$L_{n_0|n_0}(dz|y) := \delta_y(dz). \tag{4.7}$$

In the same way as in Definition 4.11, the generation processes are constructed by induction. Hence, assume the process $L_{n-1|n_0}$ has already been defined, for some $(n-1) \geq n_0$, and continue with:

(ii) Define the process of descendants of order n of the subtree originating in generation n_0 at location y by:

$$L_{n|n_0}(dz|y) := \int_{\mathbb{E}} K_n(dz|x) L_{n-1|n_0}(dx|y). \tag{4.8}$$

(*iii*) Define the *exclusive* and *inclusive* subtrees ${N_{|n_0}}^-$ and ${N_{|n_0}}^+$ originating in generation n_0 at location y by:

$$N^-_{|n_0}(dz|y) := \sum_{r=n_0+1}^{\infty} L_{r|n_0}(dz|y), \quad N^+_{|n_0}(dz|y) := \sum_{r=n_0}^{\infty} L_{r|n_0}(dz|y). \tag{4.9}$$

♢

Note that the origin of a Hawkes subtree is a fixed location $y \in \mathbb{E}$ and not one of the nodes of the Hawkes tree. In this regard, Hawkes subtrees are comparable to cluster families: The originating node of the subtrees ${N_{|n_0}}^-$ and ${N_{|n_0}}^+$ has to be one of the nodes of the generation processes $L_{n|0}$. In fact, we will see below that the subtree processes are only used in situations where y is the integration variable with respect to $L_{n|0}$.

The next property is related to the recursive construction of Hawkes trees and one could call it a *self-similarity* property. Loosely speaking, it says that each branch of a Hawkes tree is again a complete Hawkes tree, conditioned on the location of its root node.

4.14 Proposition (Self-similarity of Hawkes trees). Let the Hawkes tree N and the subtrees $N(\cdot|y)_{y\in\mathbb{E}}$ be given as specified in Definition 4.11 and Definition 4.13.

(*1*) *Complete tree.* For all $n_2 \geq n_1$ one has:

$$\begin{aligned} \int_{\mathbb{E}} L_{n_2|n_1}(dz|x) L_{n_1|0}(dx) &= L_{n_2|0}(dz), \\ \int_{\mathbb{E}} N^+_{|n_1}(dz|x) L_{n_1|0}(dx) &= \sum_{r=n_1}^{\infty} L_{r|0}(dz). \end{aligned} \tag{4.10}$$

(*2*) *Subtree.* For all $n_2 \geq n_1 \geq n_0$ one has:

$$\begin{aligned} \int_{\mathbb{E}} L_{n_2|n_1}(dz|x) L_{n_1|n_0}(dx|y) &= L_{n_2|n_0}(dz|y), \\ \int_{\mathbb{E}} N^+_{|n_1}(dz|x) L_{n_1|n_0}(dx|y) &= \sum_{r=n_1}^{\infty} L_{r|n_0}(dz|y). \end{aligned} \tag{4.11}$$

(3) *Self-similarity.* For all $n_0 \geq 0$ one has:

$$\begin{aligned} N^-_{|0}(dz) &= \int_{\mathbb{E}} N^+_{|1}(dz|x) K_1(dx|0), \\ N^-_{|n_0}(dz|y) &= \int_{\mathbb{E}} N^+_{|n_0+1}(dz|x) L_{n_0+1|n_0}(dx|y). \end{aligned} \tag{4.12}$$

♢

The formulas (4.10) show how the generation processes and subtrees of a complete Hawkes tree are related to each other. Similarly, the formulas (4.11) describe the same relations, but this time for subtrees. Finally, the formulas (4.12) show how higher order subtrees can be aggregated to obtain lower order subtrees.

Now we come to the definition of a Hawkes process itself. Basically, a Hawkes process consists of an immigration process such that each event, which we call an *immigrant*, generates a Hawkes tree and the Hawkes process is then the superposition of all these trees:

The original definition goes back to [Haw71], from where Hawkes processes obtained their name. The following definition is more general and is mainly motivated by [DVJ03], see Example 6.3(c), Example 6.4(c), Example 7.2(b) and Example 8.3(c).

4.15 Definition (Univariate Hawkes process). Let H be a finite and G be a locally-finite measure on $\mathbb{E}$. Assume $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space such that one can define the following objects:

(i) Let J be an inhomogeneous Poisson process on $\mathbb{E}$ with mean measure G.

Now consider the following cluster family directed by J:

(ii) Let $N_{|0}{}^+(dz|x)$ be a family of Hawkes trees directed by the immigration process $J(dx)$.

The Hawkes process is now defined as:

$$N(dz) := \int_{\mathbb{E}} N^+_{|0}(dz|x) J(dx). \qquad \diamondsuit$$

A remark about the naming: There is not an established naming scheme for the two measures G and H describing a Hawkes process. In the literature one can find the following notation, depending on the context in which the Hawkes process is used: The measure H is often called the *offspring measure*, *infectivity measure*, *decay measure* or the *transfer measure*. The measure G

is often called the *immigration measure* but in many applications one simply takes the Lebesgue measure on $\mathbb{E}$. Correspondingly, J is called the *immigration process* and its members the *immigrants*.

As it was already the case for Hawkes trees, without further assumptions, the Hawkes process as defined above is not necessarily a locally-finite point process. But we will see later that if the Hawkes trees are well-defined then so is the associated Hawkes process.

4.16 Remark (Construction of a Hawkes process). Let us clarify two points concerning the above construction:

(1) *Construction of a family of Hawkes trees.* One can define the family ${N_{|0}}^+(dz|x)$ of Hawkes trees in the same way as one defines a cluster family: Start with a sequence $({N_i}^+)_{i\geq 1}$ of iid point processes with the same distribution as the Hawkes tree ${N_{|0}}^+$ with root node in $0 \in \mathbb{E}$. Then construct the cluster family ${N_{|0}}^+(dz|x)$ by shifting each of the generating clusters ${N_i}^+$ according to the points of the immigration process $J(dx)$. This leads to a family of Hawkes trees ${N_{|0}}^+(dz|x)_{x\in\mathbb{E}}$ directed by $J(dx)$. See Definition 3.17 for the details of this construction.

(2) *Definition in terms of exclusive Hawkes trees.* The choice to use inclusive Hawkes trees N^+ is quite arbitrary. Indeed, one can also take the exclusive Hawkes trees N^- instead. This leads to the following equivalent representation of a Hawkes process:

$$N(dz) := J(dz) + \int_{\mathbb{E}} N^-_{|0}(dz|x)J(dx). \qquad \diamondsuit$$

4.4 Multivariate Hawkes Processes

We now turn to the multivariate case. Unfortunately, the notation becomes more cumbersome, as there are now d component processes and not only one. This is the reason why we first considered only the univariate case.

4.17 Remark (Event spaces for Hawkes processes). In the univariate case, the event space is simply $\mathbb{E} := \mathbb{R}^e$. Recall that a general multivariate event space is of the form $\mathbb{E} := \{\mathbb{E}_1, \ldots, \mathbb{E}_d\}$, see also Definition 2.2. But if we want to define a multivariate Hawkes process, we cannot take a general multivariate event space $\mathbb{E}$. The reason is that nodes from one component are used as center points of clusters in other components. Thus, all nodes have to

lie in the same event space. This means that all component event spaces $\mathbb{E}_j$ must coincide. ♢

From now on assume the multivariate event space is of the form $\mathbb{E} := \{\mathbb{E}, \ldots, \mathbb{E}\}$. The construction of a Hawkes process in the multivariate case follows along the same lines as in the univariate case. In the next definition we repeat the same construction as in Definition 4.11, but without the detailed description of the single steps:

4.18 Definition (Multivariate Hawkes tree). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathbb{E} := \{\mathbb{E}, \ldots, \mathbb{E}\}$ be a multivariate event space with a multivariate family of finite transfer measures $\mathbb{H}$.

(*i*) Fix some $1 \leq m \leq d$. The root generation process $L_{0|0}$ is defined as:

$$L^{(j)}_{0|0:m}(dz) := \delta_{jm}\delta_0(dz),$$

for all $1 \leq j \leq d$. Hence, the root node of the Hawkes tree lies at location $0 \in \mathbb{E}$ in component m.

(*ii*) Let ${K_{n|k}}^{(j)}(dz|x)$ be cluster family with a Poisson distribution with mean measure $H_{jk,x}$, for all $1 \leq j, k \leq d$. Define recursively for $n \geq 1$ the generation process of order n by:

$$L^{(j)}_{n|0:m}(dz) := \sum_{k=1}^{d} \int_{\mathbb{E}} K^{(j)}_{n|k}(dz|x) L^{(k)}_{n-1|0:m}(dx). \tag{4.13}$$

(*iii*) Define the *exclusive* and *inclusive* Hawkes trees by:

$$N^{-(j)}_{|0:m} := \sum_{n=1}^{\infty} L^{(j)}_{n|0:m} \qquad \text{and} \qquad N^{+(j)}_{|0:m} := \sum_{n=0}^{\infty} L^{(j)}_{n|0:m}, \tag{4.14}$$

for all $1 \leq j \leq d$. ♢

Note that in the multivariate case, the origin of a Hawkes tree is determined by the pair (m, y), where y is the root location and m is the root component index.

As in the univariate case, we define also subtrees. To this end, one needs to fix a triple (n_0, k, y) indicating the origin of the subtree. The root node of the subtree is then at location y, in component k, in generation n_0.

4.19 Definition (Multivariate Hawkes subtree). Fix a generation $n_0 \geq 0$, a component index $1 \leq k \leq d$ and a location $y \in \mathbb{E}$.

(*i*) Define the root generation $L_{n_0|n_0}$ of the subtree by:

$$L^{(j)}_{n_0|n_0:k}(dz|y) := \delta_{jk}\delta_y(dz).$$

(*ii*) Define the generation processes $L_{n|n_0}$ recursively for $(n-1) \geq n_0$ by:

$$L^{(j)}_{n|n_0:k}(dz|y) := \sum_{l=1}^{d} \int_{\mathbb{E}} K^{(j)}_{n|l}(dz|x) L^{(l)}_{n-1|n_0:k}(dx|y).$$

(*iii*) Define the *exclusive* and *inclusive* subtrees by:

$$N^{-(j)}_{|n_0:k} := \sum_{n=n_0+1}^{\infty} L^{(j)}_{n|n_0:k} \qquad \text{and} \qquad N^{+(j)}_{|n_0:k} := \sum_{n=n_0}^{\infty} L^{(j)}_{n|n_0:k}. \qquad \diamondsuit$$

The definition of a Hawkes process in the multivariate case is analogous to the univariate one:

4.20 Definition (Multivariate Hawkes process). Let $\boldsymbol{G}$ be a vector of locally-finite measures and $\mathbb{H}$ a matrix of finite measures on $\mathbb{E}$. Assume $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space on which are defined the following stochastic objects:

(*i*) Assume $J^{(m)}$ is a family of independent Poisson processes with mean measures G_m, for $1 \leq m \leq d$. The vector of measures $\boldsymbol{G}$ is called the *immigration intensities* and the processes $J^{(m)}$ the *immigration processes.*

(*ii*) Assume $N^{+(j)}_{|0:m}(dz|x)$ is a cluster family of multivariate Hawkes trees generated by the transfer measures $\mathbb{H}$ and directed by the immigration processes $J^{(m)}(dx)$, for all $1 \leq j \leq d$.

(*iii*) The Hawkes process can now be defined based either on the inclusive Hawkes trees N^{+} or based on the exclusive Hawkes trees N^{-}. The two equivalent representations are:

$$\begin{aligned} N^{(j)}(dz) &:= \sum_{m=1}^{d} \int_{\mathbb{E}} N^{+(j)}_{|0:m}(dz|x) J^{(m)}(dx) \\ &:= J^{(j)}(dz) + \sum_{m=1}^{d} \int_{\mathbb{E}} N^{-(j)}_{|0:m}(dz|x) J^{(m)}(dx), \end{aligned} \tag{4.15}$$

where $1 \leq j \leq d$. $\diamondsuit$

Some of the formulas in the following sections will be quite cumbersome. We therefore use a more compact notation, where we take the freedom to omit indexes if they can be deduced from the context:

4.21 Notation (Compact version). Sometimes, the generation index does not influence the distribution of a given process. As soon as we deal with moment measures, only distributions are relevant, so that the additional index specifying the generation does not matter at all. Moreover, it is often clear from the context, which generation is meant. This justifies that we leave out the generation index in some situations. We use the following two styles of notation interchangeably:

$$K_{k,y}^{(j)}(dz) \equiv K_{n|k}^{(j)}(dz|y), \qquad N_{k,y}^{\pm(j)}(dz) \equiv N_{|n_0:k}^{\pm(j)}(dz|y).$$

Let us clarify the notation we use for Hawkes trees and Hawkes processes, and how one can easily distinguish between the two: Note that we distinguish between exclusive Hawkes trees N^- and inclusive Hawkes trees N^+, but there is only one version of a Hawkes process. Hence, if one of the symbols $+$ or $-$ is present, then one knows it is a Hawkes tree and otherwise it is a Hawkes process. ♢

The following corollary is a reformulation of results from the univariate case, now adapted to the multivariate case:

4.22 Corollary (Self-similarity structure of trees). Consider a multivariate Hawkes tree, as specified in Definition 4.18.

(1) Two consecutive generation processes are related by:

$$L_{n|m}^{(j)}(dz) = \sum_{k=1}^{d} \int_{\mathbb{E}} K_{k,x}^{(j)}(dz) L_{n-1|m}^{(k)}(dx). \tag{4.16}$$

(2) A subtree is related to the subtrees of its children by:

$$N_{m,y}^{-(j)}(dz) = \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,x}^{+(j)}(dz) K_{m,y}^{(k)}(dx). \tag{4.17}$$

♢

4.5 Some Intermediate Results

The lengthy expressions given below will appear repeatedly. We introduce some placeholders for these expressions, so we can refer to them more easily later on. In all what follows, we always assume $A, B \subseteq \mathbb{E}$ are two bounded sets.

4.23 Definition (Placeholders for product measures). Consider a multivariate Hawkes process and the associated generation processes and subtree processes. For given $1 \leq i, j \leq d$ and $A, B \subseteq \mathbb{E}$ define the placeholders:

(1a) $$\Upsilon_{\delta\delta:L} := \sum_{n=0}^{\infty} \delta_{ij} \int_{\mathbb{E}} \delta_y(A)\delta_y(B) L^{(i)}_{n|m}(dy).$$

(1b) $$\Upsilon_{\delta N^-:L} := \sum_{n=0}^{\infty} \int_{\mathbb{E}} \delta_y(A) N^{-(j)}_{i,y}(B) L^{(i)}_{n|m}(dy).$$

(1c) $$\Upsilon_{N^-\delta:L} := \sum_{n=0}^{\infty} \int_{\mathbb{E}} \delta_y(B) N^{-(i)}_{j,y}(A) L^{(j)}_{n|m}(dy).$$

(1d) $$\Upsilon_{N^+N^+:[K]:L} := \sum_{n=0}^{\infty} \sum_{r,k,l=1}^{d} \int_{\mathbb{E}^3} N^{+(i)}_{k,y_1}(A) N^{+(j)}_{l,y_2}(B) K^{[kl]}_{r,x}(d\boldsymbol{y}) L^{(r)}_{n|m}(dx).$$

(2a) $$\Upsilon_{\delta\delta} := \delta_{ijm}\delta_0(A)\delta_0(B).$$

(2b) $$\Upsilon_{\delta\delta:N^-} := \delta_{ij} \int_{\mathbb{E}} \delta_y(A)\delta_y(B) N^{-(i)}_m(dy).$$

(2c) $$\Upsilon_{\delta|N^+:K} := \delta_{im}\delta_0(A) \sum_{k=1}^{d} \int_{\mathbb{E}} N^{+(j)}_{k,y}(B) K^{(k)}_m(dy).$$

(2d) $$\Upsilon_{N^+:K|\delta} := \delta_{jm}\delta_0(B) \sum_{k=1}^{d} \int_{\mathbb{E}} N^{+(i)}_{k,y}(A) K^{(k)}_m(dy).$$

(2e) $$\Upsilon_{(N^+):K} := \sum_{k=1}^{d} \int_{\mathbb{E}} N^{+(ij)}_{k,y}(A \times B) K^{(k)}_m(dy).$$

(2f) $$\Upsilon_{[N^+]:K} := \sum_{k=1}^{d} \int_{\mathbb{E}} N^{+[ij]}_{k,y}(A \times B) K^{(k)}_m(dy).$$

(2g) $\Upsilon_{N^+N^+:[K]} := \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N^{+(i)}_{k,y_1}(A) N^{+(j)}_{l,y_2}(B) K^{[kl]}_m(d\boldsymbol{y})$.

(3a) $\Upsilon_{(N^+):J} := \sum_{k=1}^{d} \int_{\mathbb{E}} N^{+(ij)}_{k,x}(A \times B) J^{(k)}(dx)$.

(3b) $\Upsilon_{[N^+]:J} := \sum_{k=1}^{d} \int_{\mathbb{E}} N^{+[ij]}_{k,x}(A \times B) J^{(k)}(dx)$.

(3c) $\Upsilon_{N^+N^+:[J]} := \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N^{+(i)}_{k,x_1}(A) N^{+(j)}_{l,x_2}(B) J^{[kl]}(d\boldsymbol{x})$.

(3d) $\Upsilon_{\delta\delta:(N^+):J} := \sum_{k=1}^{d} \int_{\mathbb{E}^3} \delta_{y_1}(A) \delta_{y_2}(B) \delta^{(2)}(ij, \boldsymbol{y}) N^{+(ij)}_{k,x}(d\boldsymbol{y}) J^{(k)}(dx)$. ♢

The following decompositions are used later to derive the second order moment measures of a Hawkes process. In the case of a Hawkes tree there are actually two representation, which we call the *explicit* and the *implicit* representation:

4.24 Lemma (Second order product measures). Consider a multivariate Hawkes tree and assume it is well-defined. Then the second order product measures can be decomposed as follows:

(1) The *explicit representations for a Hawkes tree* are given by:

$$\begin{aligned} N^{+(ij)}_m(A \times B) &= \Upsilon_{\delta\delta:L} + \Upsilon_{\delta N^-:L} + \Upsilon_{N^-\delta:L} + \Upsilon_{N^+N^+:[K]:L}, \\ N^{+[ij]}_m(A \times B) &= \Upsilon_{\delta N^-:L} + \Upsilon_{N^-\delta:L} + \Upsilon_{N^+N^+:[K]:L}. \end{aligned} \tag{4.18}$$

(2) The *implicit representations for a Hawkes tree* are given by:

$$\begin{aligned} N^{+(ij)}_m(A \times B) &= \Upsilon_{\delta\delta} + \Upsilon_{\delta|N^+:K} + \Upsilon_{N^+:K|\delta} + \Upsilon_{(N^+):K} + \Upsilon_{N^+N^+:[K]}, \\ N^{+[ij]}_m(A \times B) &= \Upsilon_{\delta|N^+:K} + \Upsilon_{N^+:K|\delta} + \Upsilon_{[N^+]:K} + \Upsilon_{N^+N^+:[K]}. \end{aligned} \tag{4.19}$$

(3) The *representations for a Hawkes process* are given by:

$$\begin{aligned} N^{(ij)}(A \times B) &= \Upsilon_{(N^+):J} + \Upsilon_{N^+N^+:[J]}, \\ N^{[ij]}(A \times B) &= \Upsilon_{[N^+]:J} + \Upsilon_{N^+N^+:[J]}. \end{aligned} \tag{4.20}$$

♢

Proofs for Chapter 4

Recall that according to Notation 4.21, we do not always mention explicitly the root node location and write $L_{n|0}(dz)$ instead of $L_{n|0}(dz|0)$, since the root node is always at $0 \in \mathbb{E}$. The same applies for $N_{|0}(dz|0)$, so that we write $N_{|0}(dz)$ instead.

Proof (Proposition 4.14): This is a direct consequence of Definition 4.13.

(1) *Complete tree.* Both statements are special cases of (4.11) if one takes $n_0 := 0$ and $y := 0$.

(2a) *Subtree, first equation.* Let $n_1 \geq n_0$. The statement is shown by induction over n_2. Consider first the base case $n_2 = n_1$: By definition, $L_{n_2|n_1}(dz|x) = L_{n_1|n_1}(dz|x) = \delta_x(dz)$, and since $L_{n_1|n_0}$ is a simple process, one finds:

$$\begin{aligned}\int_{\mathbb{E}} L_{n_2|n_1}(dz|x) L_{n_1|n_0}(dx|y) &= \int_{\mathbb{E}} \delta_x(dz) L_{n_1|n_0}(dx|y) \\ &= L_{n_1|n_0}(dz|y) = L_{n_2|n_0}(dz|y).\end{aligned}$$

Now assume the statement has already been shown for some $(n_2-1) \geq n_1$. Then, due to Equation (4.8):

$$\begin{aligned}&\int_{\mathbb{E}} L_{n_2|n_1}(dz|x) L_{n_1|n_0}(dx|y) \\ &\quad= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}} K_{n_2}(dz|w) L_{n_2-1|n_1}(dw|x) \Big] L_{n_1|n_0}(dx|y) \\ &\quad= \int_{\mathbb{E}} K_{n_2}(dz|w) \Big[\int_{\mathbb{E}} L_{n_2-1|n_1|k}(dw|x) L_{n_1|n_0}(dx|y) \Big].\end{aligned}$$

If one now applies the induction assumption and then Equation (4.8), the

induction step follows with:

$$(\ldots) = \int_{\mathbb{E}} K_{n_2}(dz|w) L_{n_2-1|n_0}(dw|y) = L_{n_2|n_0}(dz|y).$$

(2b) *Subtree, second equation.* This follows due to (4.9) and (4.11) by:

$$\begin{aligned}\int_{\mathbb{E}} N^{+}_{|n_1}(dz|x) L_{n_1|n_0}(dx|y) &= \int_{\mathbb{E}} \Big[\sum_{r=n_1}^{\infty} L_{r|n_1}(dz|x) \Big] L_{n_1|n_0}(dx|y) \\ &= \sum_{r=n_1}^{\infty} \int_{\mathbb{E}} L_{r|n_1}(dz|x) L_{n_1|n_0}(dx|y) = \sum_{r=n_1}^{\infty} L_{r|n_0}(dz|y).\end{aligned}$$

(3a) *Self-similarity, first equation.* The first equation follows if one takes $n_0 := 0$ and $y := 0$ in the second part of (4.12).

(3b) *Self-similarity, second equation.* If one takes $n_1 := (n_0 + 1)$ and $n_2 := r$ in the first part of (4.11), one obtains:

$$\int_{\mathbb{E}} L_{r|n_0+1}(dz|x) L_{n_0+1|n_0}(dx|y) = L_{r|n_0}(dz|y).$$

Now, according to the definition of $N_{|n}{}^{-}$ and $N_{l|n}{}^{+}$, one has:

$$\begin{aligned}N^{-}_{|n_0}(dz|y) &= \sum_{r=n_0+1}^{\infty} L_{r|n_0}(dz|y) \\ &= \sum_{r=n_0+1}^{\infty} \Big[\int_{\mathbb{E}} L_{r|n_0+1}(dz|x) L_{n_0+1|n_0}(dx|y) \Big] \\ &= \int_{\mathbb{E}} \Big[\sum_{r=n_0+1}^{\infty} L_{r|n_0+1}(dz|x) \Big] L_{n_0+1|n_0}(dx|y) \\ &= \int_{\mathbb{E}} N^{+}_{|n_0+1}(dz|x) L_{n_0+1|n_0}(dx|y). \qquad \square\end{aligned}$$

Proof (Corollary 4.22): The first equation follows from (4.13), where the compact Notation 4.21 is used. The second equation is the multivariate version of the first part of (4.12), again in compact notation. $\square$

The proof of Lemma 4.24 is given in three parts, each one corresponding to one of the three parts of the lemma.

Explicit Representations for Hawkes Trees

In this section, we prove the first part of Lemma 4.24. Before we give the proof, we illustrate what lies behind this decomposition. We sketch the proof only in the univariate case:

4.25 Remark (Sketch of proof). In this sketch, we write expressions like $a \in N$, with a slight abuse of notation. This is supposed to mean that a is a point of N, i.e. $N(\{a\}) = 1$. Fix two sets $A, B \subseteq \mathbb{E}$ and consider for $n \geq 0$ the following point process (in informal notation):

$$J_n(A \times B) := \Big\{(a \times b) \in A \times B : \exists y \in L_n : (a \times b) \in N_y^{+(2)}\Big\}.$$

Hence, J_n contains all pairs $(a \times b)$ that have a common ancestor in generation n. The following three observations are important:

(*i*) The sequence $J_{n\geq 0}$ is decreasing. To see this, take some $(a \times b) \in J_{n+1}$. Then both nodes a, b have an common ancestor in generation $n+1$, say $y \in L_{n+1}$. Clearly, every parent of y, say $x \in L_n$, is then also an ancestor of both a and b. Thus $(a \times b) \in J_n$, so that $J_{n+1} \leq J_n$.

(*ii*) The sequence $J_{n\geq 0}$ converges to zero: This is also clear, since the Hawkes tree is well-defined, it has only finitely many nodes. Hence, there are only finitely many generations, which implies that $J_m = 0$, for m large enough.

(*iii*) By definition, a Hawkes tree has only one root node, which is at location 0. Since the root node process is L_0, one has $L_0 = \{0\}$, in informal notation. Hence, all pairs (a, b) have the node $y := 0$ as a common ancestor. This shows $N^+ = J_0$.

The above considerations imply that N^+ can be written as the following telescoping series:

$$N^{+(2)}(A \times B) = \sum_{n=0}^{\infty} \Big[J_n(A \times B) - J_{n+1}(A \times B)\Big] = \sum_{n=0}^{\infty} \hat{J}_n(A \times B),$$

where $\hat{J}_n := J_n - J_{n+1}$, for $n \geq 0$. Since $J_{n\geq 0}$ is a decreasing sequence, $\hat{J}_n$ is indeed non-negative, i.e. a point process. If one now expresses $\hat{J}_n$ in explicit form using the generation processes L_n and the subtree processes N_y^- and N_y^+, the stated decomposition follows.

We will not do this here, since it is done in full generality in the proof below. Note that in the actual proof, the processes J_n are defined more rigorously and the intuitive statements given above are shown formally. Especially, we will use again standard point process notation, which leads to more cumbersome formulas but is certainly more precise. ♢

4.26 Lemma (Intermediate result for explicit decomposition). Consider a multivariate Hawkes process as specified in Definition 4.18.

(*i*) *Auxiliary process.* For $n \geq 0$ define the processes:

$$J_n^{+(ij)}(A \times B) := \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,y}^{+(ij)}(A \times B) L_{n|m}^{(k)}(dy), \tag{4.21}$$

for $A \subseteq \mathbb{E}_i$, $B \subseteq \mathbb{E}_j$ and all $1 \leq i, j \leq d$. Since ${L_{0|m}}^{(k)}(dz) = \delta_{km}\delta_0(dz)$, one has ${J_0}^{+(ij)} = {N_m}^{+(ij)}$.

One has now the following two decompositions:

(*i*) The first decomposition is:

$$\begin{aligned} J_n^{+(ij)}(A \times B) &= \delta_{ij} \int_{\mathbb{E}} \delta_y(A)\delta_y(B) L_{n|m}^{(i)}(dy) \\ &+ \int_{\mathbb{E}} \delta_y(A) N_{i,y}^{-(j)}(B) L_{n|m}^{(i)}(dy) + \int_{\mathbb{E}} \delta_y(B) N_{j,y}^{-(i)}(A) L_{n|m}^{(j)}(dy) \\ &+ \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,y}^{-(i)}(A) N_{k,y}^{-(j)}(B) L_{n|m}^{(k)}(dy). \end{aligned} \tag{4.22}$$

(*ii*) The second decomposition is:

$$\begin{aligned} J_{n+1}^{+(ij)}(A \times B) &= \sum_{r=1}^{d} \int_{\mathbb{E}} N_{r,x}^{-(i)}(A) N_{r,x}^{-(j)}(B) L_{n|m}^{(r)}(dx) \\ &- \sum_{r,k,l=1}^{d} \int_{\mathbb{E}^3} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) K_{r,x}^{[kl]}(d\boldsymbol{y}) L_{n|m}^{(r)}(dx). \end{aligned} \tag{4.23}$$

♢

PROOF (LEMMA 4.26): We will use the results from Corollary 4.22.

(*i*) Since $N^{+(ij)}(d\boldsymbol{y}) = N^{+(i)}(dy_1)N^{+(j)}(dy_2)$, and due to (4.14), the first decomposition follows with:

$$\begin{aligned}
J_n^{+(ij)}(A\times B) &= \sum_{k=1}^{d}\int_{\mathbb{E}} N_{k,y}^{+(i)}(A)N_{k,y}^{+(j)}(B)L_{n|m}^{(k)}(dy)\\
&= \sum_{k=1}^{d}\int_{\mathbb{E}}\Big[\delta_{ik}\delta_y(A)+N_{k,y}^{-(i)}(A)\Big]\Big[\delta_{jk}\delta_y(B)+N_{k,y}^{-(j)}(B)\Big]L_{n|m}^{(k)}(dy)\\
&= \delta_{ij}\int_{\mathbb{E}}\delta_y(A)\delta_y(B)L_{n|m}^{(i)}(dy)+\int_{\mathbb{E}}\delta_y(A)N_{i,y}^{-(j)}(B)L_{n|m}^{(i)}(dy)\\
&\quad+\int_{\mathbb{E}}\delta_y(B)N_{j,y}^{-(i)}(A)L_{n|m}^{(j)}(dy)\\
&\quad+\sum_{k=1}^{d}\int_{\mathbb{E}}N_{k,y}^{-(i)}(A)N_{k,y}^{-(j)}(B)L_{n|m}^{(k)}(dy).
\end{aligned}$$

(*ii*) Due to (4.16), one has in a first step:

$$\begin{aligned}
J_{n+1}^{+(ij)}(A\times B) &= \sum_{k=1}^{d}\int_{\mathbb{E}}N_{k,y}^{+(ij)}(A\times B)L_{n+1|m}^{(k)}(dy)\\
&= \sum_{k=1}^{d}\int_{\mathbb{E}^3}N_{k,y}^{+(ij)}(A\times B)\Big[\sum_{r=1}^{d}\int_{\mathbb{E}}K_{r,x}^{(k)}(dy)L_{n|m}^{(r)}(dx)\Big]\\
&= \sum_{r,k=1}^{d}\int_{\mathbb{E}^2}N_{k,y}^{+(ij)}(A\times B)K_{r,x}^{(k)}(dy)L_{n|m}^{(r)}(dx).
\end{aligned}$$

Next apply Equation (2.16) and use that $\delta^{(2)}(kl, y_1y_2) = 1$ if and only if

$\delta^{(2)}(y_1y_2) = 1$ and $\delta^{(2)}(kl) = 1$. Hence:

$$\begin{aligned}
K^{(k)}_{r,x}(dy_1) &= \delta^{(2)}(kk)K^{(k)}_{r,x}(dy_1) = \int_{\mathbb{E}} \delta^{(2)}(kk, y_1y_2)K^{(kk)}_{r,x}(dy_1 \times dy_2) \\
&= \int_{\mathbb{E}} \sum_{l=1}^{d} \delta^{(2)}(kl, y_1y_2)K^{(kl)}_{r,x}(dy_1 \times dy_2) \\
&= \sum_{r,k=1}^{d} \int_{\mathbb{E}^2} N^{+(ij)}_{k,y_1}(A \times B)\Big[\sum_{l=1}^{d} \int_{\mathbb{E}} \delta^{(2)}(kl, \boldsymbol{y})K^{(kl)}_{r,x}(d\boldsymbol{y})\Big]L^{(r)}_{n|m}(dx) \\
&= \sum_{r,k,l=1}^{d} \int_{\mathbb{E}^3} \delta^{(2)}(kl, \boldsymbol{y})N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{(kl)}_{r,x}(d\boldsymbol{y})L^{(r)}_{n|m}(dx).
\end{aligned}$$

Since $\delta^{(2)}(kl, \boldsymbol{y}) = 1 - \delta^{[2]}(kl, \boldsymbol{y})$, it now follows that:

$$\begin{aligned}
K^{(k)}_{r,x}(dy_1) = &\sum_{r,k,l=1}^{d} \int_{\mathbb{E}^3} N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{(kl)}_{r,x}(d\boldsymbol{y})L^{(r)}_{n|m}(dx) \\
&- \sum_{r,k,l=1}^{d} \int_{\mathbb{E}^3} \delta^{[2]}(kl, \boldsymbol{y})N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{(kl)}_{r,x}(d\boldsymbol{y})L^{(r)}_{n|m}(dx).
\end{aligned}$$

Now use that

$$K^{(kl)}(d\boldsymbol{y}) = K^{(k)}(dy_1)K^{(l)}(dy_2), \quad \delta^{[2]}(kl, \boldsymbol{y})K^{(kl)}_{r,x}(d\boldsymbol{y}) = K^{[kl]}_{r,x}(d\boldsymbol{y}),$$

so that the above expression becomes:

$$\begin{aligned}
K^{(k)}_{r,x}(dy_1) = &\sum_{r=1}^{d} \int_{\mathbb{E}} \Big[\sum_{k=1}^{d} \int_{\mathbb{E}} N^{+(i)}_{k,y_1}(A)K^{(k)}_{r,x}(dy_1)\Big] \\
&\qquad \Big[\sum_{l=1}^{d} \int_{\mathbb{E}} N^{+(j)}_{l,y_2}(B)K^{(l)}_{r,x}(dy_2)\Big]L^{(r)}_{n|m}(dx) \\
&- \sum_{r,k,l=1}^{d} \int_{\mathbb{E}^3} N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{[kl]}_{r,x}(d\boldsymbol{y})L^{(r)}_{n|m}(dx).
\end{aligned}$$

Due to (4.17), the claim follows now with:

$$K_{r,x}^{(k)}(dy_1) = \sum_{r=1}^{d} \int_{\mathbb{E}} N_{r,x}^{-(i)}(A) N_{r,x}^{-(j)}(B) L_{n|m}^{(r)}(dx)$$
$$- \sum_{r,k,l=1}^{d} \int_{\mathbb{E}^3} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) K_{r,x}^{[kl]}(d\boldsymbol{y}) L_{n|m}^{(r)}(dx). \qquad \square$$

Proof (Lemma 4.24): This is the first part of the proof of Lemma 4.24, see also Remark 4.25 for a sketch.

(i) *Ordinary product measure.* Let ${J_n}^{+(ij)}$ be as in Equation (4.21) and define for $n \geq 0$:

$$\hat{J}_n^{+(ij)} := J_n^{+(ij)} - J_{n+1}^{+(ij)}.$$

Now substitute (4.22) and (4.23), so that:

$$\begin{aligned}
\hat{J}_n^{+(ij)}(A \times B) &= J_n^{+(ij)}(A \times B) - J_{n+1}^{+(ij)}(A \times B) \\
&= \delta_{ij} \int_{\mathbb{E}} \delta_y(A)\delta_y(B) L_{n|m}^{(i)}(dy) + \int_{\mathbb{E}} \delta_y(A) N_{i,y}^{-(j)}(B) L_{n|m}^{(i)}(dy) \\
&\quad + \int_{\mathbb{E}} \delta_y(B) N_{j,y}^{-(i)}(A) L_{n|m}^{(j)}(dy) \\
&\quad + \sum_{r,k,l=1}^{d} \int_{\mathbb{E}^3} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) K_{r,x}^{[kl]}(d\boldsymbol{y}) L_{n|m}^{(r)}(dx).
\end{aligned}$$

This shows that $\hat{J}_n$ is indeed non-negative. Therefore, according to the definition, ${J_n}^{+(ij)}$ is a decreasing sequence. Next, fix $n \geq 0$ and write ${N_m}^{+(ij)}$ as a telescoping series:

$$N_m^{+(ij)} = J_0^{+(ij)} = \left[J_0^{+(ij)} - J_n^{+(ij)} \right] + J_n^{+(ij)} = \sum_{r=0}^{n-1} \hat{J}_r^{+(ij)} + J_n^{+(ij)}.$$

Since the Hawkes tree is by assumption well-defined, all realizations are locally-finite point configurations. Hence, the limit $\lim_{n\to\infty} {J_n}^{+(ij)}$ vanishes on every bounded set, so that:

$$N_m^{+(ij)}(A \times B) = \lim_{n\to\infty} \left[\sum_{r=0}^{n-1} \hat{J}_r^{+(ij)} - J_n^{+(ij)} \right] = \sum_{r=0}^{\infty} \hat{J}_r^{+(ij)}(A \times B).$$

The decomposition of the ordinary product measure follows if one replaces $\hat{J}_n$ with the explicit expression from above.

(*ii*) *Factorial product measure.* Due to Equation (2.16) and Definition 4.18, one has in a first step:

$$\begin{aligned}\int_{A\times B} \delta^{(2)}(ij,\boldsymbol{y}) N_m^{+(ij)}(d\boldsymbol{y}) &= \int_{\mathbb{E}^2} \mathbb{1}_{\{A\cap B\}}(y_1)\delta^{(2)}(ij,\boldsymbol{y}) N_m^{+(ij)}(d\boldsymbol{y}) \\ &= \delta_{ij}\int_{\mathbb{E}} \mathbb{1}_{\{A\cap B\}}(y) N^{+(i)}(dy) = \delta_{ij}\int_{\mathbb{E}} \delta_y(A)\delta_y(B) N_m^{+(i)}(dy) \\ &= \delta_{ij}\int_{\mathbb{E}} \delta_y(A)\delta_y(B)\Big[\sum_{n=0}^{\infty} L_{n|m}^{(i)}(dy)\Big] = \Upsilon_{\delta\delta:L}.\end{aligned}$$

Due to $\delta^{[2]} = 1 - \delta^{(2)}$, one finds for the factorial product measure:

$$\begin{aligned}N_m^{+[ij]}(A\times B) &= \int_{A\times B} \delta^{[2]}(ij,\boldsymbol{y}) N_m^{+(ij)}(d\boldsymbol{y}) \\ &= \int_{A\times B} N_m^{+(ij)}(d\boldsymbol{y}) - \int_{A\times B} \delta^{(2)}(ij,\boldsymbol{y}) N_m^{+(ij)}(d\boldsymbol{y}).\end{aligned}$$

Now apply the first part of Equation (4.18) and replace the second integral with $\Upsilon_{\delta\delta:L}$, so that:

$$N_m^{+[ij]}(A\times B) = \Big[\Upsilon_{\delta\delta:L} + \Upsilon_{\delta N^-:L} + \Upsilon_{N^-\delta:L} + \Upsilon_{N^+N^+:[K]:L}\Big] - \Upsilon_{\delta\delta:L}. \quad \square$$

Implicit Representations for Hawkes Trees

Proof (Lemma 4.24): This is the second part of the proof of Lemma 4.24.

(*i*) *Ordinary product measure.* Due to the relation between the inclusive and

exclusive Hawkes tree and with Equation (4.17) one has:

$$
\begin{aligned}
N_m^{+(ij)}(A \times B) &= N_m^{+(i)}(A) N_m^{+(j)}(B) \\
&= \Big[\delta_{im}\delta_0(A) + \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,y_1}^{+(j)}(A) K_m^{(k)}(dy_1)\Big] \\
&\quad \Big[\delta_{jm}\delta_0(B) + \sum_{l=1}^{d} \int_{\mathbb{E}} N_{l,y_2}^{+(j)}(B) K_m^{(l)}(dy_2)\Big] \\
&= \delta_{ijm}\delta_0(A)\delta_0(B) + \delta_{im}\delta_0(A) \sum_{l=1}^{d} \int_{\mathbb{E}} N_{l,y_2}^{+(j)}(B) K_m^{(l)}(dy_2) \\
&\quad + \delta_{jm}\delta_0(B) \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,y_1}^{+(i)}(A) K_m^{(k)}(dy_1) \\
&\quad + \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) K_m^{(kl)}(d\boldsymbol{y}).
\end{aligned}
$$

One can now identify three of the placeholders from Definition 4.23, so that the previous expression becomes:

$$
\begin{aligned}
N_m^{+(ij)}(A \times B) &= \Upsilon_{\delta\delta} + \Upsilon_{\delta|N^+:K} + \Upsilon_{N^+:K|\delta} \\
&\quad + \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) K_m^{(kl)}(d\boldsymbol{y}).
\end{aligned}
$$

It remains to show that the last term is equal to $\Upsilon_{(N^+):K} + \Upsilon_{N^+N^+:[K]}$.

But since $1 = \delta^{(2)} + \delta^{[2]}$, and with Equation (2.16), this follows from:

$$\begin{aligned}
&\sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) K_m^{(kl)}(d\boldsymbol{y}) \\
&\quad = \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) \delta^{(2)}(kl, \boldsymbol{y}) K_m^{(kl)}(d\boldsymbol{y}) \\
&\qquad + \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) \delta^{[2]}(kl, \boldsymbol{y}) K_m^{(kl)}(d\boldsymbol{y}) \\
&\quad = \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,y}^{+(i)}(A) N_{k,y}^{+(j)}(B) K_m^{(k)}(dy) \\
&\qquad + \sum_{k,l=1}^{d} \int_{\mathbb{E}} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) K_m^{[kl]}(d\boldsymbol{y}) \\
&\quad = \Upsilon_{(N^+):K} + \Upsilon_{N^+N^+:[K]}.
\end{aligned}$$

(*ii*) *Factorial product measure.* According to the definition of the factorial product measure, the fact that $\delta^{[2]} = 1 - \delta^{(2)}$, and due to Equation (2.16), one has:

$$\begin{aligned}
N_m^{+[ij]}(A \times B) &= \int_{\mathbb{E}^2} \delta_{y_1}(A) \delta_{y_2}(B) \delta^{[2]}(ij, \boldsymbol{y}) N_m^{+(ij)}(d\boldsymbol{y}) \\
&= \int_{\mathbb{E}^2} \delta_{y_1}(A) \delta_{y_2}(B) N_m^{+(ij)}(d\boldsymbol{y}) \\
&\quad - \int_{\mathbb{E}^2} \delta_{y_1}(A) \delta_{y_2}(B) \delta^{(2)}(ij, \boldsymbol{y}) N_m^{+(ij)}(d\boldsymbol{y}) \\
&= N_m^{+(ij)}(A \times B) - \delta_{ij} \int_{\mathbb{E}} \delta_y(A) \delta_y(B) N_m^{+(i)}(dy) \\
&= N_m^{+(ij)}(A \times B) - \delta_{ij} \int_{\mathbb{E}} \delta_y(A) \delta_y(B) \Big[\delta_{im} \delta_0(dy) + N_m^{-(i)}(dy) \Big].
\end{aligned}$$

Next one can identify the placeholders $\Upsilon_{\delta\delta}$ and $\Upsilon_{\delta\delta:N^-}$, and due to first

part of Equation (4.19), the above expression becomes:

$$\begin{aligned}
N_m^{+[ij]}(A\times B) &= N_m^{+(ij)}(A\times B) - \delta_{ijm}\delta_0(A)\delta_0(B)\\
&\qquad - \delta_{ij}\int_{\mathbb{E}} \delta_y(A)\delta_y(B) N_m^{-(i)}(dy)\\
&= N_m^{+(ij)}(A\times B) - \Upsilon_{\delta\delta} - \Upsilon_{\delta\delta:N^-}\\
&= \Big[\Upsilon_{\delta\delta} + \Upsilon_{\delta|N^+:K} + \Upsilon_{N^+:K|\delta} + \Upsilon_{(N^+):K} + \Upsilon_{N^+N^+:[K]}\Big]\\
&\qquad - \Upsilon_{\delta\delta} - \Upsilon_{\delta\delta:N^-}\\
&= \Upsilon_{\delta|N^+:K} + \Upsilon_{N^+:K|\delta} + \Upsilon_{N^+N^+:[K]} + \Upsilon_{(N^+):K} - \Upsilon_{\delta\delta:N^-}.
\end{aligned} \tag{4.24}$$

Except for the last two terms, this decomposition coincides with the claimed formula. Hence, we need to decompose the last term once more: Due to (4.17) and (2.16), one obtains:

$$\begin{aligned}
\Upsilon_{\delta\delta:N^-} &= \delta_{ij}\int_{\mathbb{E}} \delta_y(A)\delta_y(B) N_m^{-(i)}(dy)\\
&= \delta_{ij}\int_{\mathbb{E}} \delta_y(A)\delta_y(B)\Big[\sum_{k=1}^{d}\int_{\mathbb{E}} N_{k,x}^{+(i)}(dy) K_m^{(k)}(dx)\Big]\\
&= \sum_{k=1}^{d}\int_{\mathbb{E}}\Big[\delta_{ij}\int_{\mathbb{E}} \delta_y(A)\delta_y(B) N_{k,x}^{+(i)}(dy)\Big] K_m^{(k)}(dx)\\
&= \sum_{k=1}^{d}\int_{\mathbb{E}}\Big[\int_{\mathbb{E}^2} \delta^{(2)}(ij,\boldsymbol{y})\delta_{y_1}(A)\delta_{y_2}(B) N_{k,x}^{+(ij)}(d\boldsymbol{y})\Big] K_m^{(k)}(dx).
\end{aligned}$$

Using that $\delta^{(2)}(ij,\boldsymbol{y}) = 1 - \delta^{[2]}(ij,\boldsymbol{y})$, and according to the definition of

the factorial product measure, the above expression becomes:

$$\begin{aligned}
\Upsilon_{\delta\delta:N^-} &= \sum_{k=1}^{d} \int_{\mathbb{E}^3} \big[1 - \delta^{[2]}(ij, \boldsymbol{y})\big] \delta_{y_1}(A)\delta_{y_2}(B) N_{k,x}^{+(ij)}(d\boldsymbol{y}) K_m^{(k)}(dx) \\
&= \sum_{k=1}^{d} \int_{\mathbb{E}} \Big[\int_{\mathbb{E}^2} \delta_{y_1}(A)\delta_{y_2}(B) N_{k,x}^{+(ij)}(d\boldsymbol{y}) \Big] K_m^{(k)}(dx) \\
&\quad - \sum_{k=1}^{d} \int_{\mathbb{E}} \Big[\int_{\mathbb{E}^2} \delta^{[2]}(ij, \boldsymbol{y}) \delta_{y_1}(A)\delta_{y_2}(B) N_{k,x}^{+(ij)}(d\boldsymbol{y}) \Big] K_m^{(k)}(dx) \\
&= \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,x}^{+(ij)}(A \times B) K_m^{(k)}(dx) - \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,x}^{+[ij]}(A \times B) K_m^{(k)}(dx) \\
&= \Upsilon_{(N^+):K} - \Upsilon_{[N^+]:K}.
\end{aligned}$$

The claimed formula follows, if one substitutes this expression for $\Upsilon_{\delta\delta:N^-}$ in (4.24). □

Representations for Hawkes Processes

Proof (Lemma 4.24): This is the third part of the proof of Lemma 4.24.

(*i*) *Ordinary product measure.* According to (4.15), and with $1 = \delta^{(2)} + \delta^{[2]}$, one finds:

$$\begin{aligned}
N^{(ij)}(A \times B) &= N^{(i)}(A) N^{(j)}(B) \\
&= \Big[\sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,x_1}^{+(i)}(A) J^{(k)}(dx_1) \Big] \Big[\sum_{l=1}^{d} \int_{\mathbb{E}} N_{l,x_2}^{+(j)}(B) J^{(l)}(dx_2) \Big] \\
&= \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N_{k,x_1}^{+(i)}(A) N_{l,x_2}^{+(j)}(B) J^{(kl)}(d\boldsymbol{x}) \\
&= \sum_{k,l=1}^{d} \int_{\mathbb{E}} N_{k,x_1}^{+(i)}(A) N_{l,x_2}^{+(j)}(B) \delta^{(2)}(kl, \boldsymbol{x}) J^{(kl)}(d\boldsymbol{x}) \\
&\quad + \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} N_{k,x_1}^{+(i)}(A) N_{l,x_2}^{+(j)}(B) \delta^{[2]}(kl, \boldsymbol{x}) J^{(kl)}(d\boldsymbol{x}).
\end{aligned}$$

The claim now follows, due to (2.16), with:

$$\begin{aligned} N^{(ij)}(A\times B) &= \sum_{k=1}^{d}\int_{\mathbb{E}} N_{k,x}^{+(ij)}(A\times B)J^{(k)}(dx) \\ &\quad + \sum_{k,l=1}^{d}\int_{\mathbb{E}^2} N_{k,x_1}^{+(i)}(A)N_{l,x_2}^{+(j)}(B)J^{[kl]}(d\boldsymbol{x}) \\ &= \Upsilon_{(N^+):J} + \Upsilon_{N^+N^+:[J]}. \end{aligned}$$

(*ii*) *Factorial product measure.* Due to $1-\delta^{(2)}=\delta^{[2]}$ and with the definition of the factorial product measure, one finds:

$$\begin{aligned} \Upsilon_{(N^+):J} - \Upsilon_{\delta\delta:(N^+):J} &= \sum_{k=1}^{d}\int_{\mathbb{E}^3} \delta_{y_1}(A)\delta_{y_2}(B)N_{k,x}^{+(ij)}(d\boldsymbol{y})J^{(k)}(dx) \\ &\quad - \sum_{k=1}^{d}\int_{\mathbb{E}^3} \delta_{y_1}(A)\delta_{y_2}(B)\delta^{(2)}(ij,\boldsymbol{y})N_{k,x}^{+(ij)}(d\boldsymbol{y})J^{(k)}(dx) \\ &= \sum_{k=1}^{d}\int_{\mathbb{E}^3} \delta_{y_1}(A)\delta_{y_2}(B)\big[1-\delta^{(2)}(ij,\boldsymbol{y})\big]N_{k,x}^{+(ij)}(d\boldsymbol{y})J^{(k)}(dx) \\ &= \sum_{k=1}^{d}\int_{\mathbb{E}^3} \delta_{y_1}(A)\delta_{y_2}(B)N_{k,x}^{+[ij]}(d\boldsymbol{y})J^{(k)}(dx) \\ &= \sum_{k=1}^{d}\int_{\mathbb{E}} N_{k,x}^{+[ij]}(A\times B)J^{(k)}(dx) = \Upsilon_{[N^+]:J}. \end{aligned} \tag{4.25}$$

Use once more $\delta^{[2]} = 1 - \delta^{(2)}$ and apply (4.15), so that:

$$\begin{aligned}
N^{[ij]}(A \times B) &= \int_{\mathbb{E}^2} \delta_{y_1}(A)\delta_{y_2}(B)\delta^{[2]}(ij, \boldsymbol{y})N^{(ij)}(d\boldsymbol{y}) \\
&= \int_{\mathbb{E}^2} \delta_{y_1}(A)\delta_{y_2}(B)N^{(ij)}(d\boldsymbol{y}) \\
&\quad - \int_{\mathbb{E}^2} \delta_{y_1}(A)\delta_{y_2}(B)\delta^{(2)}(ij, \boldsymbol{y})N^{(ij)}(d\boldsymbol{y}) \\
&= N^{(ij)}(A \times B) \\
&\quad - \sum_{k=1}^{d} \int_{\mathbb{E}^3} \delta_{y_1}(A)\delta_{y_2}(B)\delta^{(2)}(ij, \boldsymbol{y})N_{k,x}^{+(ij)}(d\boldsymbol{y})J^{(k)}(dx) \\
&= N^{(ij)}(A \times B) - \Upsilon_{\delta\delta:(N^+):J}.
\end{aligned}$$

Now with the first part of Equation (4.20) and the intermediate result from (4.25), the claim follows with:

$$\begin{aligned}
N^{[ij]}(A \times B) &= \Upsilon_{(N^+):J} + \Upsilon_{N^+N^+:[J]} - \Upsilon_{\delta\delta:(N^+):J} \\
&= \Upsilon_{[N^+]:J} + \Upsilon_{N^+N^+:[J]}.
\end{aligned}$$

□

Chapter 5

Moment Measures

In this chapter, we calculate the first and second order moment measures of a Hawkes process. From now on we always assume that the branching matrix Q satisfies $\mathrm{Spr}(Q) < 1$, see Definition 4.2. This condition implies that the first order moment measures are finite, and therefore the Hawkes process is well-defined.

5.1 Motivation and Objectives

Without a doubt, the calculation of the moment measures of a Hawkes process is a tedious undertaking, even if it is basically a straight-forward procedure. The first order moment measures can be obtained quite easily. The second order moment measures require much more work and it gets already difficult to present the calculation in a readable form. For the third order moment measures, let alone higher order measures, one would quite likely hit the limit of what can be presented in text form. This explains also why, to my knowledge, only first and second order moment measures of Hawkes processes are discussed in the literature.

As a very helpful reference for this chapter has served the book [DVJ03]. It touches in several places second order moment measures of Hawkes processes, see especially Exercise 5.5.7 and Example 8.3(c). Some of the results given there are stated without proofs, the reason most likely being space considerations. I will give in this chapter a complete derivation of the moment measures of a Hawkes process of the first two orders. In Theorem 5.4, explicit and implicit equations for the second order moment measures of a Hawkes process are

presented. The corresponding reduced versions are given in Theorem 5.7.

Markov Renewal Equations. Let me explain in a few words the purpose behind this chapter, as it may not be obvious what all these lengthy calculations are useful for. The presentation given here deviates in a few points from the one in [DVJ03]. I want to point out that there is a close connection between moment measures of Hawkes processes and Markov renewal equations. I will not make this connection explicit, since all results can be derived without using the theory of Markov renewal equations. The similarity to renewal equations however influences the presentation considerably. This can be seen from the frequent use of convolutions. This also explains why I called the measure in Definition 4.9 a *cluster kernel*. Using this notation, one can give quite compact, explicit as well as implicit equations for the moment measures of a Hawkes process.

The results in this chapter could be derived more easily using Fourier transform techniques. I think however that the approach chosen has some advantages: If one has become familiar with the notation, the results obtained usually have quite intuitive interpretations. If nothing else, the approach chosen hopefully gives some insight into the stochastic dependence of the location of the nodes of a Hawkes process. I sometimes added a few captions that should give a hint of how the corresponding expression can be interpreted.

The ground work for this chapter has already been laid in Chapter 3 where Poisson cluster processes have been discussed and corresponding formulas for the moment measures have been derived. As already explained in the introduction to the previous chapter, the trick is to find a suitable decomposition such that these moment measure formulas can be applied. I gave such a decomposition in Lemma 4.24 of the previous chapter.

Marked vs. Unmarked Hawkes Processes. All results are given for the case of an unmarked, multivariate Hawkes process. The calculations for the univariate and the multivariate case are very similar, and therefore I consider directly the more general situation of a multivariate Hawkes process. For marked Hawkes processes however, the situation is different: In general, the given results can not be extended to marked Hawkes processes, not even in the univariate case.

If one wanted to deal with marked Hawkes processes, basically the same techniques could be used. The trick would be to consider an *extended event space*, $\bar{\mathbb{E}}$ say, and incorporate the marks in this space. One could then show

that the marked Hawkes process is equivalent to an unmarked Hawkes process on this extended event space $\bar{\mathbb{E}}$. So it seems that all results could be carried over from the unmarked to the marked case. But there is one problem: On this extended event space, the transfer measures matrices $\mathbb{H}_{y\in\bar{\mathbb{E}}}$, see Definition 4.7, are not any more shifted versions of each other. This implies that the calculations could not any more be done using convolutions and therefore one would get stuck sooner or later with terms that cannot be simplified further.

This is the reason why I consider the case of unmarked Hawkes processes only, since only in this case it is possible to carry out the calculations with the convenient notation based on transfer measures matrices and cluster kernels.

Independent Marks. As an aside, if one has a marked Hawkes process with iid marks, then the moment measure formulas can be derived trivially from the formulas for the unmarked case, see Proposition 6.4.IV in [DVJ03] for the univariate case and Example 8.3(d) for the multivariate case. These results are true in general and not specific to Hawkes processes.

5.2 First Order Moment Measures

The first order moment measures of a Hawkes process are easy to calculate. We will use the following formulas later for the derivation of the second order moment measures:

5.1 Proposition (First order moment measures). Consider a multivariate Hawkes process and its associated Hawkes trees and generation processes.

(*1*) *Generation process.* The first order moment measures of the generation process of order n are given by:

$$M^{(j)}_{n|m}(dz) := \mathbb{E}\big[L^{(j)}_{n|m}(dz)\big] = H^{*n}_{jm}(dz), \qquad M^{(j)}_{n|m}(\mathbb{E}) = Q^n_{jm}. \tag{5.1}$$

(*2*) *Hawkes tree.* The first order moment measures of the Hawkes tree are given by:

$$M^{+(j)}_{m}(dz) := \mathbb{E}\big[N^{+(j)}_{m}(dz)\big] = U^{+}_{jm}(dz), \quad M^{+(j)}_{m}(\mathbb{E}) = (\mathbb{1}_d - Q)^{-1}_{jm}. \tag{5.2}$$

(*3*) *Hawkes process.* Assume that the immigration intensity $\boldsymbol{G}$ is a locally-finite measure on $\mathbb{E}^d$. Then one has for the first order moment measures

of the Hawkes process:

$$M^{(j)}(dz) = \big[\mathbb{U}^+ * \boldsymbol{G}\big]_j(dz).$$

Now assume that $\boldsymbol{G}(d\boldsymbol{x}) = \boldsymbol{\eta} d\boldsymbol{x}$, for a constant $\boldsymbol{\eta} \geq \mathbf{0}$, i.e. the immigration intensity is a multiple of the Lebesgue measure on $\mathbb{E}^d$. Then:

$$M^{(j)}(dz) = \big[(\mathbb{1}_d - Q)^{-1}\boldsymbol{\eta}\big]_j \lambda_{\mathbb{E}}(dz). \qquad \diamondsuit$$

5.2 Notation (Moment measures of subtrees). In correspondence with the notation used so far, let us introduce the following symbols for the moment measures of subtrees:

$$M^{+(j)}_{k,y}(dz) := \mathbb{E}\big[N^{+(j)}_{k,y}(dz)\big], \qquad M^{+(ij)}_{k,y}(d\boldsymbol{z}) := \mathbb{E}\big[N^{+(ij)}_{k,y}(d\boldsymbol{z})\big]. \tag{5.3}$$

We use the analog notation for the factorial instead of the ordinary moment measure. $\diamondsuit$

The distribution of a Hawkes subtree is closely related to the distribution of the complete tree. This leads to the following results concerning the first order moment measures of a Hawkes subtree:

5.3 Corollary (First order moment measures of subtrees). Consider a Hawkes tree N with associated subtrees $N_{y\in\mathbb{E}}$. The moment measures of the complete tree and its subtrees are related by:

$$M^{+(j)}_{k,y}(dz) = M^{+(j)}_{k}(dz - y), \qquad M^{+(ij)}_{k,y}(d\boldsymbol{z}) = M^{+(ij)}_{k}(d\boldsymbol{z} - y). \tag{5.4}$$

Moreover, the following explicit expressions can be found:

$$M^{+(j)}_{k,y}(dz) = U^{+}_{jk,y}(dz), \qquad M^{+(j)}_{k,y}(\mathbb{E}) = (\mathbb{1}_d - Q)^{-1}_{jk}. \tag{5.5}$$

$\diamondsuit$

5.3 Second Order Moment Measures

In Lemma 4.24, we have derived two decompositions of the product measures of a Hawkes tree, which we have called the *explicit* and *implicit* version. Correspondingly, there are two different representations for the second order moment measures of a Hawkes tree:

5.4 Theorem (Second order moment measures). Assume the second order moment measures are finite.

(1) *Explicit representation for Hawkes trees.* The second order moment measures are given by:

$$\begin{aligned} M_m^{+(ij)}(A \times B) &= \sum_{r=1}^{d} \int_{\mathbb{E}} U_{ir,y}^{+}(A) U_{jr,y}^{+}(B) U_{rm}^{+}(dy), \\ M_m^{+[ij]}(A \times B) &= \int_{\mathbb{E}} \delta_y(A) U_{ji,y}^{-}(B) U_{im}^{+}(dy) + \int_{\mathbb{E}} \delta_y(B) U_{ij,y}^{-}(A) U_{jm}^{+}(dy) \\ &\quad + \sum_{r=1}^{d} \int_{\mathbb{E}} U_{ir,y}^{-}(A) U_{jr,y}^{-}(B) U_{rm}^{+}(dy). \end{aligned} \tag{5.6}$$

Based on the self-similarity structure of a Hawkes tree, the second order moment measures of a Hawkes tree can be represented in a second way:

(2) *Implicit representations for Hawkes trees.* The second order moment measures of a Hawkes tree solve the following implicit equations:

$$\begin{aligned} M_m^{+(ij)}(A \times B) &= \sum_{k=1}^{d} \int_{\mathbb{E}} M_{k,y}^{+(ij)}(A \times B) H_{km}(dy) + U_{im}^{+}(A) U_{jm}^{+}(B), \\ M_m^{+[ij]}(A \times B) &= \sum_{k=1}^{d} \int_{\mathbb{E}} M_{k,y}^{+[ij]}(A \times B) H_{km}(dy) + \delta_{im} \delta_0(A) U_{jm}^{-}(B) \\ &\quad + \delta_{jm} \delta_0(B) U_{im}^{-}(A) + U_{im}^{-}(A) U_{jm}^{-}(B). \end{aligned} \tag{5.7}$$

For a Hawkes process, not a tree, there is only one representation for the moment measures:

(3) *Representation for Hawkes processes.* The second order moment measures

of a Hawkes process are given by:

$$
\begin{aligned}
M^{(ij)}(A \times B) &= \sum_{k=1}^{d} \int_{\mathbb{E}} M_{k,x}^{+(ij)}(A \times B) G_k(dx) \\
&\quad + \Big[\sum_{k=1}^{d} \int_{\mathbb{E}} U_{ik,x_1}^{+}(A) G_k(dx_1) \Big] \Big[\sum_{l=1}^{d} \int_{\mathbb{E}} U_{jl,x_2}^{+}(B) G_l(dx_2) \Big], \\
M^{[ij]}(A \times B) &= \sum_{k=1}^{d} \int_{\mathbb{E}} M_{k,x}^{+[ij]}(A \times B) G_k(dx) \\
&\quad + \Big[\sum_{k=1}^{d} \int_{\mathbb{E}} U_{ik,x_1}^{+}(A) G_k(dx_1) \Big] \Big[\sum_{l=1}^{d} \int_{\mathbb{E}} U_{jl,x_2}^{+}(B) G_l(dx_2) \Big]. \qquad \diamondsuit
\end{aligned} \tag{5.8}
$$

5.5 Remark (Second order moment measures). The following facts can be observed concerning the second order moment measures:

(1) *Implicit representation.* Obviously, the implicit formulas given in the second part above need first to be solved in order to obtain the second order moment measures. This is also the reason why we call this the *implicit* representation. One can easily check that the implicit and explicit representations are equivalent. Even so, we mention the implicit versions because they reflect nicely the self-similarity structure of a Hawkes tree.

(2) *Non-symmetric expressions.* Since the expressions are not symmetric in the indexes i and j, they are not symmetric with respect to the sets A and B. Also note that in the univariate case, the second order moment measures are clearly symmetric.

(3) *Non-translation invariant measures.* The second order moment measures are not translation-invariant. This is important, as we want to calculate the reduced versions of these measures. $\diamondsuit$

5.4 Reduced Moment Measures

Since the second order moment measures are not translation-invariant, we can only calculate the pseudo-reduced measures and not the regular reduced measures, see Definition 2.22 for the details.

5.6 Remark (Univariate and multivariate case). Let us for a moment go back to the theory of reduced measures from Chapter 2. We have seen that a non-symmetric measure μ has d different reduced measures $\mu^{\langle k \rangle}$, $1 \leq k \leq d$, and moreover the formulas tend to be more cumbersome.

Since the second order moment measures are only symmetric in the univariate case, we restrict the following exposure to the univariate case. This simplifies the calculation by a great deal, as we do not need to distinguish between the different versions of reduced measures. Once one has found the formulas for the moment measures in the univariate case, one could extend them easily to the multivariate case. ♢

5.7 Theorem (Reduced second order moment measures). Assume the second order moment measures are finite.

(1) *Explicit representation for Hawkes trees.* The pseudo-reduced second order moment measures of a Hawkes tree are given by:

$$\begin{aligned}
\mathring{M}^{+(2)}(A) &= (1-Q)^{-1} \int_{\mathbb{E}} U^+(A+w) U^+(dw), \\
\mathring{M}^{+[2]}(A) &= (1-Q)^{-1} U^-(-A) + (1-Q)^{-1} U^-(A) \\
&\qquad + (1-Q)^{-1} \int_{\mathbb{E}} U^-(A+w) U^-(dw).
\end{aligned}$$

(2) *Implicit representation for Hawkes trees.* The pseudo-reduced second order moment measures of a Hawkes tree solve the following implicit equations:

$$\begin{aligned}
\mathring{M}^{+(2)}(A) &= Q\mathring{M}^{+(2)}(A) + \int_{\mathbb{E}} U^+(A+w) U^+(dw), \\
\mathring{M}^{+[2]}(A) &= Q\mathring{M}^{+[2]}(A) + U^-(-A) + U^-(A) + \int_{\mathbb{E}} U^-(A+w) U^-(dw).
\end{aligned}$$

For Hawkes trees we have taken the pseudo-reduction of the moment measures, but for Hawkes processes we calculate the regular reduction:

(3) *Representation for Hawkes processes.* Assume the immigration measure G is homogeneous, i.e. assume without loss of generality that $G(dx) = dx$. In this case, the reduced second order moment measures of a Hawkes

process are:

$$\breve{M}^{(2)}(A) = \mathring{M}^{+(2)}(A) + (\mathbb{1}_d - Q)^{-2}\lambda_{\mathbb{E}}(A),$$
$$\breve{M}^{[2]}(A) = \mathring{M}^{+[2]}(A) + (\mathbb{1}_d - Q)^{-2}\lambda_{\mathbb{E}}(A). \qquad \diamondsuit$$

5.5 Some Intermediate Results

To better present the overall structure of the proofs, we formulate some of the intermediate steps as separate Lemmas. Recall that the decompositions in Lemma 4.24 are expressed in terms of the placeholders given in Definition 4.23. In order to calculate the second order moment measures, one basically needs to calculate the expectations of these placeholders:

5.8 Lemma (Expectation of placeholders). Let $1 \leq i, j \leq d$ and $A, B \subseteq \mathbb{E}$. Below we list the expectations of the placeholders from Definition 4.23. For later use, we introduce again placeholders for the resulting expressions:

(1a) $\Upsilon_{\delta\delta:U^+} := \mathbb{E}[\Upsilon_{\delta\delta:L}] = \delta_{ij} \int_{\mathbb{E}} \delta_y(A)\delta_y(B)U^+_{im}(dy).$

(1b) $\Upsilon_{\delta U^-:U^+} := \mathbb{E}[\Upsilon_{\delta N^-:L}] = \int_{\mathbb{E}} \delta_y(A)U^-_{ji,y}(B)U^+_{im}(dy).$

(1c) $\Upsilon_{U^-\delta:U^+} := \mathbb{E}[\Upsilon_{N^-\delta:L}] = \int_{\mathbb{E}} \delta_y(B)U^-_{ij,y}(A)U^+_{jm}(dy).$

(1d) $\Upsilon_{U^-U^-:U^+} := \mathbb{E}[\Upsilon_{N^+N^+:[K]:L}] = \sum_{r=1}^{d} \int_{\mathbb{E}} U^-_{ir,y}(A)U^-_{jr,y}(B)U^+_{rm}(dy).$

(1d) $\Upsilon_{U^+U^+:U^+} := \sum_{r=1}^{d} \int_{\mathbb{E}} U^+_{ir,y}(A)U^+_{jr,y}(B)U^+_{rm}(dy).$

(2a) $\Upsilon_{\delta U^-} := \mathbb{E}[\Upsilon_{\delta|N^+:K}] = \delta_{im}\delta_0(A)U^-_{jm}(B).$

(2b) $\Upsilon_{U^-\delta} := \mathbb{E}[\Upsilon_{N^+:K|\delta}] = \delta_{jm}\delta_0(B)U^-_{im}(A).$

(2c) $\Upsilon_{(M^+):H} := \mathbb{E}[\Upsilon_{(N^+):K}] = \sum_{k=1}^{d} \int_{\mathbb{E}} M_{k,y}^{+(ij)}(A \times B) H_{km}(dy).$

(2d) $\Upsilon_{[M^+]:H} := \mathbb{E}[\Upsilon_{[N^+]:K}] = \sum_{k=1}^{d} \int_{\mathbb{E}} M_{k,y}^{+[ij]}(A \times B) H_{km}(dy).$

(2e) $\Upsilon_{U^-U^-} := \mathbb{E}[\Upsilon_{N^+N^+:[K]}] = U_{im}^{-}(A) U_{jm}^{-}(B).$

(2f) $\Upsilon_{U^+U^+} := U_{im}^{+}(A) U_{jm}^{+}(B).$

(3a) $\Upsilon_{(M^+):G} := \mathbb{E}[\Upsilon_{(N^+):J}] = \sum_{k=1}^{d} \int_{\mathbb{E}} M_{k,x}^{+(ij)}(A \times B) G_k(dx).$

(3b) $\Upsilon_{[M^+]:G} := \mathbb{E}[\Upsilon_{[N^+]:J}] = \sum_{k=1}^{d} \int_{\mathbb{E}} M_{k,x}^{+[ij]}(A \times B) G_k(dx).$

(3c) $\Upsilon_{U^+:G|U^+:G} := \mathbb{E}[\Upsilon_{N^+N^+:[J]}]$

$$= \Big[\sum_{k=1}^{d} \int_{\mathbb{E}} U_{ik,x_1}^{+}(A) G_k(dx_1)\Big] \Big[\sum_{l=1}^{d} \int_{\mathbb{E}} U_{jl,x_2}^{+}(B) G_l(dx_2)\Big]. \quad \diamondsuit$$

As a next preparation for the proof of the moment formulas, we reformulate Theorem 5.4 in terms of the placeholders given above:

5.9 Lemma (Reformulation of moment measure formulas). The following three statements are equivalent to the three statements given in Theorem 5.4:

(1) *Explicit representation for Hawkes trees.* The two Equations (5.6) are equivalent to:

$$\begin{aligned} M_m^{+(ij)}(A \times B) &= \Upsilon_{U^+U^+:U^+}, \\ M_m^{+[ij]}(A \times B) &= \Upsilon_{\delta U^-:U^+} + \Upsilon_{U^-\delta:U^+} + \Upsilon_{U^-U^-:U^+}. \end{aligned} \tag{5.9}$$

(2) *Implicit representations for Hawkes trees.* The two Equations (5.7) are equivalent to:

$$\begin{aligned} M_m^{+(ij)}(A \times B) &= \Upsilon_{(M^+):H} + \Upsilon_{U^+U^+}, \\ M_m^{+[ij]}(A \times B) &= \Upsilon_{[M^+]:H} + \Upsilon_{\delta U^-} + \Upsilon_{U^-\delta} + \Upsilon_{U^-U^-}. \end{aligned} \tag{5.10}$$

(3) *Representation for Hawkes processes.* The two Equations (5.8) are equivalent to:

$$\begin{aligned} M^{(ij)}(A \times B) &= \Upsilon_{(M^+):G} + \Upsilon_{U^+:G|U^+:G}, \\ M^{[ij]}(A \times B) &= \Upsilon_{[M^+]:G} + \Upsilon_{U^+:G|U^+:G}. \end{aligned} \tag{5.11}$$

♢

In order to calculate the reduced moment measures one basically needs to calculate the reduced versions of the corresponding placeholders. One should be careful not to confuse the regular reduction $\breve{\mu}$ and the pseudo-reduction $\mathring{\mu}$ of a measure μ in the next lemma:

5.10 Lemma (Reductions of placeholders). Let $1 \leq i, j \leq d$ and $A, B \subseteq \mathbb{E}$. The pseudo-reductions of the placeholders from Lemma 5.8 are given by:

(1a) $\mathring{\Upsilon}_{\delta U^-:U^+} + \mathring{\Upsilon}_{U^-\delta:U^+} = (1-Q)^{-1}U^-(-A) + (1-Q)^{-1}U^-(A).$

(1b) $\mathring{\Upsilon}_{U^-U^-:U^+} = (1-Q)^{-1}\int_{\mathbb{E}} U^-(A+w)U^-(dw).$

(1c) $\mathring{\Upsilon}_{U^+U^+:U^+} = (1-Q)^{-1}\int_{\mathbb{E}} U^+(A+w)U^+(dw).$

(2a) $\mathring{\Upsilon}_{\delta U^-} + \mathring{\Upsilon}_{U^-\delta} = U^-(-A) + U^-(A).$

(2b) $\mathring{\Upsilon}_{(M^+):H} = Q\mathring{M}^{+(2)}(A).$

(2c) $\mathring{\Upsilon}_{[M^+]:H} = Q\mathring{M}^{+[2]}(A).$

(2d) $\mathring{\Upsilon}_{U^-U^-} = \int_{\mathbb{E}} U^-(A+w)U^-(dw).$

(2e) $\mathring{\Upsilon}_{U^+U^+} = \int_{\mathbb{E}} U^+(A+w)U^+(dw).$

For the next three placeholders, assume that the immigration intensity is the Lebesgue measure $G(dx) = dx$. This time, the regular reduction and not the pseudo-reduction is calculated:

(3a) $\breve{\Upsilon}_{(M^+):G} = \mathring{M}^{+(2)}(A)$.

(3b) $\breve{\Upsilon}_{[M^+]:G} = \mathring{M}^{+[2]}(A)$.

(3c) $\breve{\Upsilon}_{U^+:G|U^+:G} = (\mathbb{1}_d - Q)^{-2}\lambda_{\mathbb{E}}(A)$. $\diamondsuit$

Proofs for Chapter 5

The proofs of this chapter are quite long. They are therefore sometimes split up in several parts.

First Order Moment Measures

PROOF (PROPOSITION 5.1): This is the first part of the proof, concerning the first order moment measure of a generation process.

(i) *Moment measure.* We show this by induction: The base case $n = 0$ follows from:

$$M_{0|m}^{(j)}(dz) = \mathbb{E}\big[L_{0|m}^{(j)}(dz)\big] = \mathbb{E}\big[\delta_{jm}\delta_0(dz)\big] = \delta_{jm}\delta_0(dz) = H_{jm}^{*0}(dz).$$

Now assume ${M_{n-1|m}}^{(j)} = H_{jm}^{*n-1}$ has already been shown for $(n-1) \geq 0$. Then, due to (4.16), one finds:

$$M_{n|m}^{(j)}(dz) = \mathbb{E}\Big[L_{n|m}^{(j)}(dz)\Big] = \mathbb{E}\Big[\sum_{k=1}^{d}\int_{\mathbb{E}} K_{k,y}^{(j)}(dz)L_{n-1|m}^{(k)}(dy)\Big].$$

Note that the first order moment measures of K_y are given by the family $\mathbb{H}_y$. Since K and L are conditionally independent, the first part of (3.9)

applies. Together with the induction assumption, this shows that:

$$\begin{aligned} M^{(j)}_{n|m}(dz) &= \sum_{k=1}^{d} \int_{\mathbb{E}} \mathbb{E}\big[K^{(j)}_{k,y}(dz)\big]\, \mathbb{E}\big[L^{(k)}_{n-1|m}(dy)\big] \\ &= \sum_{k=1}^{d} \int_{\mathbb{E}} H_{jk,y}(dz) M^{(k)}_{n-1|m}(dy) = \sum_{k=1}^{d} \int_{\mathbb{E}} H_{jk,y}(dz) H^{*n-1}_{km}(dy) \\ &= \sum_{k=1}^{d} \big[H_{jk} * H^{*n-1}_{km}\big](dz) = H^{*n}_{jm}(dz). \end{aligned}$$

(*ii*) *Total mass.* We show this by induction: The base case $n = 0$ follows from:

$$M^{(j)}_{0|m}(\mathbb{E}) = H^{*0}_{jm}(\mathbb{E}) = \delta_{jm}\delta_0(\mathbb{E}) = \delta_{jm} = Q^0_{jm}.$$

Now assume $H^{*n-1}_{jm}(\mathbb{E}) = Q^{n-1}_{jm}$ has already been shown for $(n-1) \geq 0$. Due to the first part of (5.1), and the fact that $H_{jk,y}(\mathbb{E}) = H_{jk}(\mathbb{E} - y) = H_{jk}(\mathbb{E})$, the statement follows with:

$$\begin{aligned} M^{(j)}_{n|m}(\mathbb{E}) = H^{*n}_{jm}(\mathbb{E}) &= \sum_{k=1}^{d} \int_{\mathbb{E}} H_{jk,y}(\mathbb{E}) H^{*n-1}_{km}(dy) \\ &= \sum_{k=1}^{d} \int_{\mathbb{E}} H_{jk}(\mathbb{E}) H^{*n-1}_{km}(dy) = \sum_{k=1}^{d} H_{jk}(\mathbb{E}) \int_{\mathbb{E}} H^{*n-1}_{km}(dy) \\ &= \sum_{k=1}^{d} \Big[H_{jk}(\mathbb{E}) H^{*n-1}_{km}(\mathbb{E})\Big] = \sum_{k=1}^{d} Q_{jk} Q^{n-1}_{km} = Q^n_{jm}. \end{aligned}$$ □

Proof (Proposition 5.1): This is second part of the proof, concerning the first order moment measure of a Hawkes tree.

(*i*) *Moment measure.* First apply Equations (4.14) and (5.1). Using the definition of $\mathbb{U}^+$, see Equation (4.1), the claim follows with:

$$\begin{aligned} M^{+(j)}_{m}(dz) = \mathbb{E}\big[N^{+(j)}_{m}(dz)\big] &= \mathbb{E}\Big[\sum_{n=0}^{\infty} L^{(j)}_{n|m}(dz)\Big] = \sum_{n=0}^{\infty} M^{(j)}_{n|m}(dz) \\ &= \sum_{n=0}^{\infty} H^{*n}_{jm}(dz) = \Big[\sum_{n=0}^{\infty} \mathbb{H}^{*n}\Big]_{jm}(dz) = U^{+}_{jm}(dz). \end{aligned}$$

(*ii*) *Total mass.* Since by assumption $\mathrm{Spr}(Q) < 1$, one can apply Equation (4.5), which shows that $\mathbb{U}^+$ is finite with total mass $(\mathbb{1}_d - Q)^{-1}$. Hence:

$$M_m^{+(j)}(\mathbb{E}) = U_{jm}^+(\mathbb{E}) = \big[(\mathbb{1}_d - Q)^{-1}\big]_{jm} = (\mathbb{1}_d - Q)^{-1}_{jm}. \qquad \square$$

Proof (Proposition 5.1): This is third part of the proof, concerning the first order moment measure of a Hawkes process.

(*i*) *General immigration intensity.* Use first Equation (4.15), and then, since N^+ and J are conditionally independent, apply the first part of Equation (3.9). Finally, due to Equation (5.2), the claim follows with:

$$\begin{aligned} M^{(j)}(dz) &= \mathbb{E}\big[N^{(j)}(dz)\big] = \mathbb{E}\Big[\sum_{m=1}^{d}\int_{\mathbb{E}} N_{m,x}^{+(j)}(dz)J^{(m)}(dx)\Big] \\ &= \sum_{m=1}^{d}\int_{\mathbb{E}} \mathbb{E}\big[N_{m,x}^{+(j)}(dz)\big]\,\mathbb{E}\big[J^{(m)}(dx)\big] = \sum_{m=1}^{d}\int_{\mathbb{E}} M_{m,x}^{+(j)}(dz)G_m(dx) \\ &= \sum_{m=1}^{d}\int_{\mathbb{E}} U_{jm,x}^{+}(dz)G_m(dx) = \big[\mathbb{U}^+ * \boldsymbol{G}\big]_j(dz). \end{aligned}$$

(*ii*) *Stationary immigration intensity.* Assume now that the immigration intensity is $G_m(dx) = \eta_m dx$, for constants $\eta_m \geq 0$, $1 \leq m \leq d$. Since ${U_{jm,x}}^+(dz) = {U_{jm}}^+(dz - x)$, and with the above calculation, one has:

$$M^{(j)}(A) = \sum_{m=1}^{d}\int_{\mathbb{E}} U_{jm,x}^{+}(A)\eta_m dx = \sum_{m=1}^{d}\eta_m\int_{\mathbb{E}} U_{jm}^{+}(A-x)dx. \quad (5.12)$$

The integral can be rewritten as:

$$\begin{aligned} \int_{\mathbb{E}} U_{jm}^{+}(A-x)dx &= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}}\delta_z(A)U_{jm}^{+}(dz-x)\Big]dx \\ &= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}}\delta_{z+x}(A)U_{jm}^{+}(dz)\Big]dx = \int_{\mathbb{E}}\Big[\int_{\mathbb{E}}\delta_x(A-z)dx\Big]U_{jm}^{+}(dz) \\ &= \int_{\mathbb{E}}\Big[\int_{\mathbb{E}}\delta_x(A)dx\Big]U_{jm}^{+}(dz) = \int_{\mathbb{E}}\lambda_{\mathbb{E}}(A)U_{jm}^{+}(dz). \end{aligned}$$

Substituting this into (5.12) and then applying the first part of (4.5), one

obtains, as claimed:

$$M^{(j)}(A) = \sum_{m=1}^{d} \eta_m \int_{\mathbb{E}} \lambda_{\mathbb{E}}(A) U_{jm}^{+}(dz) = \sum_{m=1}^{d} \eta_m \Big[\int_{\mathbb{E}} U_{jm}^{+}(dz) \Big] \lambda_{\mathbb{E}}(A)$$

$$= \Big[\sum_{m=1}^{d} (\mathbb{1}_d - Q)_{jm}^{-1} \eta_m \Big] \lambda_{\mathbb{E}}(A) = \big[(\mathbb{1}_d - Q)^{-1} \boldsymbol{\eta} \big]_j \lambda_{\mathbb{E}}(A). \qquad \square$$

Proof (Corollary 5.3): By construction, the subtree $N_{k,y}$ with root node at $y \in \mathbb{E}$ has the same distribution as the complete tree N_k, except that the subtree $N_{k,y}$ is shifted by the amount y.

(*i*) The moment measures $M_{k,y}$ and M_k are related in the same way as $N_{k,y}$ and N_k are. Hence, $M_{k,y}$ coincides with M_k, except that it is shifted by y, as claimed.

(*ii*) From the first part, and due to (5.2), one knows that:

$$M_{k,y}^{+(j)}(dz) = M_k^{+(j)}(dz - y) = U_{jk}^{+}(dz - y) = U_{jk,y}^{+}(dz).$$

Since $U_{jk,y}^{+}(dz) = U_{jk}^{+}(dz - y)$, the second equality follows with:

$$M_{k,y}^{+(j)}(\mathbb{E}) = U_{jk,y}^{+}(\mathbb{E}) = U_{jk}^{+}(\mathbb{E} - y) = U_{jk}^{+}(\mathbb{E}) = (\mathbb{1}_d - Q)_{jk}^{-1}. \qquad \square$$

Second Order Moment Measures

Proof (Lemma 5.8): This is the first part of the proof, concerning the explicit representation.

(*1a*) *Relationship between an event and its duplicate.* Since $\mathbb{E}[{L_{n|m}}^{(i)}] = H_{im}^{*n}$,

see Equation (5.1), the statement follows with:

$$\begin{aligned}\mathbb{E}\big[\Upsilon_{\delta\delta:L}(A\times B)\big] &= \mathbb{E}\Big[\sum_{n=0}^{\infty}\delta_{ij}\int_{\mathbb{E}}\delta_y(A)\delta_y(B)L^{(i)}_{n|m}(dy)\Big]\\ &= \sum_{n=0}^{\infty}\delta_{ij}\int_{\mathbb{E}}\delta_y(A)\delta_y(B)\,\mathbb{E}\big[L^{(i)}_{n|m}(dy)\big]\\ &= \delta_{ij}\int_{\mathbb{E}}\delta_y(A)\delta_y(B)\Big[\sum_{n=0}^{\infty}H^{*n}_{im}(dy)\Big]\\ &= \delta_{ij}\int_{\mathbb{E}}\delta_y(A)\delta_y(B)U^{+}_{im}(dy) = \Upsilon_{\delta\delta:U^+}(A\times B).\end{aligned}$$

(1b) *Relationship between an event and its descendants.* Since N^- and L are conditionally independent, the first part of Equation (3.9) applies. Thus:

$$\begin{aligned}\mathbb{E}\big[\Upsilon_{\delta N^-:L}(A\times B)\big] &= \mathbb{E}\Big[\sum_{n=0}^{\infty}\int_{\mathbb{E}}\delta_y(A)N^{-(j)}_{i,y}(B)L^{(i)}_{n|m}(dy)\Big]\\ &= \sum_{n=0}^{\infty}\int_{\mathbb{E}}\delta_y(A)\,\mathbb{E}\big[N^{-(j)}_{i,y}(B)\big]\,\mathbb{E}\big[L^{(i)}_{n|m}(dy)\big].\end{aligned}$$

Next use the first part of Equation (5.1) and the first part of Equation (5.5), but for the exclusive instead of the inclusive version of the subtree. The statement now follows with:

$$\begin{aligned}(\ldots) &= \int_{\mathbb{E}}\delta_y(A)U^{-}_{ji,y}(B)\Big[\sum_{n=0}^{\infty}H^{*n}_{im}(dy)\Big]\\ &= \int_{\mathbb{E}}\delta_y(A)U^{-}_{ji,y}(B)U^{+}_{im}(dy) = \Upsilon_{\delta U^-:U^+}(A\times B).\end{aligned}$$

(1c) *Relationship between the descendants and their ancestor.* The calculation for $\Upsilon_{\delta N^-:L}$ is almost the same as for $\Upsilon_{N^-\delta:L}$. Therefore:

$$\mathbb{E}\big[\Upsilon_{N^-\delta:L}(A\times B)\big] = \int_{\mathbb{E}}\delta_y(B)U^{-}_{ij,y}(A)U^{+}_{jm}(dy) = \Upsilon_{U^-\delta:U^+}(A\times B).$$

(1d) *Relationship between the two families of two different children of an event.* Since the inner integral below is conditionally independent of L, one can

apply the first part of Equation (3.8), so that:

$$\begin{aligned}
&\mathbb{E}\big[\Upsilon_{N^+N^+:[K]:L}(A\times B)\big]\\
&\quad=\mathbb{E}\Big[\sum_{n=0}^{\infty}\sum_{r,k,l=1}^{d}\int_{\mathbb{E}}\Big[\int_{\mathbb{E}^2}N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{[kl]}_{r,x}(d\boldsymbol{y})\Big]L^{(r)}_{n|m}(dx)\Big]\\
&\quad=\sum_{n=0}^{\infty}\sum_{r,k,l=1}^{d}\int_{\mathbb{E}}\mathbb{E}\Big[\int_{\mathbb{E}^2}N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{[kl]}_{r,x}(d\boldsymbol{y})\Big]\,\mathbb{E}\big[L^{(r)}_{n|m}(dx)\big].
\end{aligned}$$

From (5.1) one knows that $\mathbb{E}[{L_{n|m}}^{(r)}]=H^{*n}_{rm}$, and this shows that:

$$\sum_{n=0}^{\infty}\mathbb{E}\big[L^{(r)}_{n|m}(dx)\big]=\sum_{n=0}^{\infty}H^{*n}_{rm}(dx)=U^{+}_{rm}(dx).$$

Substituting this in the expression above gives:

$$\begin{aligned}
&\mathbb{E}\big[\Upsilon_{N^+N^+:[K]:L}(A\times B)\big]\\
&\quad=\sum_{r=1}^{d}\int_{\mathbb{E}}\sum_{k,l=1}^{d}\mathbb{E}\Big[\int_{\mathbb{E}^2}N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{[kl]}_{r,x}(d\boldsymbol{y})\Big]U^{+}_{rm}(dx).
\end{aligned}\tag{5.13}$$

Note that the integral below is of the form $\int K_{y_1}K_{y_2}L^{[2]}(d\boldsymbol{y})$, so that the second part of Equation (3.9) applies. Moreover, since $K_{r,x}$ is a Poisson process, one knows from the multivariate version of Equation (3.11) that $\mathbb{E}\big[K^{[kl]}(d\boldsymbol{y})\big]=\mathbb{E}\big[K^{(k)}(dy_1)\big]\,\mathbb{E}\big[K^{(l)}(dy_2)\big]$. Together, this shows that:

$$\begin{aligned}
&\sum_{k,l=1}^{d}\mathbb{E}\Big[\int_{\mathbb{E}^2}N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{[kl]}_{r,x}(d\boldsymbol{y})\Big]\\
&\quad=\sum_{k,l}^{d}\int_{\mathbb{E}^2}\mathbb{E}\big[N^{+(i)}_{k,y_1}(A)\big]\,\mathbb{E}\big[N^{+(j)}_{l,y_2}(B)\big]\,\mathbb{E}\big[K^{[kl]}_{r,x}(d\boldsymbol{y})\big]\\
&\quad=\sum_{k,l=1}^{d}\int_{\mathbb{E}^2}\mathbb{E}\big[N^{+(i)}_{k,y_1}(A)\big]\,\mathbb{E}\big[N^{+(j)}_{l,y_2}(B)\big]\,\mathbb{E}\big[K^{(k)}_{r,x}(dy_1)\big]\,\mathbb{E}\big[K^{(l)}_{r,x}(dy_2)\big].
\end{aligned}$$

Now use that $\mathbb{E}[{K_{r,x}}^{(j)}]=H_{jr,x}$, and due to Equation (5.5), one knows

that $\mathbb{E}[N_{m,y}{}^{+(j)}] = U_{jm,y}{}^{+}$. Together with Equation (4.4), this yields:

$$\begin{aligned}
&\sum_{k,l=1}^{d} \mathbb{E}\Big[\int_{\mathbb{E}^2} N_{k,y_1}^{+(i)}(A) N_{l,y_2}^{+(j)}(B) K_{r,x}^{[kl]}(d\boldsymbol{y})\Big] \\
&\quad = \sum_{k,l=1}^{d} \int_{\mathbb{E}^2} U_{ik,y_1}^{+}(A) U_{jl,y_2}^{+}(B) H_{kr,x}(dy_1) H_{lr,x}(dy_2) \\
&\quad = \Big[\sum_{k=1}^{d} \int_{\mathbb{E}} U_{ik,y_1}^{+}(A) H_{kr,x}(dy_1)\Big]\Big[\sum_{l=1}^{d} \int_{\mathbb{E}} U_{jl,y_2}^{+}(B) H_{lr,x}(dy_2)\Big] \\
&\quad = U_{ir,x}^{-}(A) U_{jr,x}^{-}(B).
\end{aligned}$$

After substituting this in (5.13), the statement follows with:

$$\begin{aligned}
\mathbb{E}\big[\Upsilon_{N^+N^+:[K]:L}(A \times B)\big] &= \sum_{r=1}^{d} \int_{\mathbb{E}} U_{ir,x}^{-}(A) U_{jr,x}^{-}(B) U_{rm}^{+}(dx) \\
&= \Upsilon_{U^-U^-:U^+}(A \times B). \qquad \square
\end{aligned}$$

Proof (Lemma 5.8): This is the second part of the proof, concerning the implicit representation.

(2a) *Relationship between the root node and the family of its children.* Since N_y^+ and K are conditionally independent, one can apply the first part of Equation (3.9). Due to Equation (5.5), one knows that $\mathbb{E}[N_{k,y}{}^{+(j)}] = U_{jk,y}^{+}$ and since K is a Poisson process with mean measure family $\mathbb{H}$, one has $\mathbb{E}[K_m{}^{(k)}] = H_{km}$. Together with Equation (4.4), the claim follows with:

$$\begin{aligned}
\mathbb{E}\big[\Upsilon_{\delta|N^+:K}(A \times B)\big] &= \mathbb{E}\Big[\delta_{im}\delta_0(A) \sum_{k=1}^{d} \int_{\mathbb{E}} N_{k,y}^{+(j)}(B) K_m^{(k)}(dy)\Big] \\
&= \delta_{im}\delta_0(A) \sum_{k=1}^{d} \int_{\mathbb{E}} \mathbb{E}\big[N_{k,y}^{+(j)}(B)\big]\, \mathbb{E}\big[K_m^{(k)}(dy)\big] \\
&= \delta_{im}\delta_0(A) \sum_{k=1}^{d} \int_{\mathbb{E}} U_{jk,y}^{+}(B) H_{km}(dy) \\
&= \delta_{im}\delta_0(A) \big[\mathbb{U}^+ * \mathbb{H}\big]_{jm}(B) = \delta_{im}\delta_0(A) U_{jm}^{-}(B) = \Upsilon_{\delta U^-}(A \times B).
\end{aligned}$$

(2b) *Relationship between the family of the children and their generating root*

node. The calculation for $\Upsilon_{N^+:K|\delta}$ is almost the same as for $\Upsilon_{\delta|N^+:K}$, so that:

$$\begin{aligned}\mathbb{E}\big[\Upsilon_{N^+:K|\delta}(A\times B)\big] &= \mathbb{E}\Big[\delta_{jm}\delta_0(B)\sum_{k=1}^{d}\int_{\mathbb{E}} N_{k,y}^{+(i)}(A)K_m^{(k)}(dy)\Big]\\ &= \delta_{jm}\delta_0(B)U_{im}^{-}(A) = \Upsilon_{U^-\delta}(A\times B).\end{aligned}$$

(2c) *Relationship among family members including duplicates of the children of the root node.* Since the integral below is of the form $\int N_y K(dy)$, and N_y and K are conditionally independent, one can apply the first part of Equation (3.8), so that:

$$\begin{aligned}\mathbb{E}\big[\Upsilon_{(N^+):K}(A\times B)\big] &= \mathbb{E}\Big[\int_{\mathbb{E}} N_{k,y}^{+(ij)}(A\times B)K_m^{(k)}(dy)\Big]\\ &= \int_{\mathbb{E}} \mathbb{E}\big[N_{k,y}^{+(ij)}(A\times B)\big]\,\mathbb{E}\big[K_m^{(k)}(dy)\big].\end{aligned}$$

Now due to Equation (5.3), and using the fact that $\mathbb{E}[K_m{}^{(k)}] = H_{km}$, one finds for the above expression:

$$(\ldots) = \sum_{k=1}^{d}\int_{\mathbb{E}} M_{k,y}^{+(ij)}(A\times B)H_{km}(dy) = \Upsilon_{(M^+):H}(A\times B).$$

(2d) *Relationship among different family members of the children of the root node.* This calculation is almost identical to the previous one, except that there is a factorial instead of an ordinary product measure. Therefore:

$$\begin{aligned}\mathbb{E}\big[\Upsilon_{[N^+]:K}(A\times B)\big] &= \mathbb{E}\Big[\int_{\mathbb{E}} N_{k,y}^{+[ij]}(A\times B)K_m^{(k)}(dy)\Big]\\ &= \sum_{k=1}^{d}\int_{\mathbb{E}} M_{k,y}^{+[ij]}(A\times B)H_{km}(dy) = \Upsilon_{[M^+]:H}(A\times B).\end{aligned}$$

(2e) *Relationship between two families of two different children of the root node.* Since the integral below is of the form $\int N_{y_1}N_{y_2}K^{[kl]}(d\boldsymbol{y})$ and N_y and K are conditionally independent, the second part of Equation (3.9)

applies, and one has:

$$\mathbb{E}\big[\Upsilon_{N^+N^+:[K]}(A\times B)\big] = \mathbb{E}\Big[\sum_{k,l=1}^{d}\int_{\mathbb{E}^2} N^{+(i)}_{k,y_1}(A)N^{+(j)}_{l,y_2}(B)K^{[kl]}_m(d\boldsymbol{y})\Big]$$
$$= \sum_{k,l=1}^{d}\int_{\mathbb{E}^2}\mathbb{E}\big[N^{+(i)}_{k,y_1}(A)\big]\,\mathbb{E}\big[N^{+(j)}_{l,y_2}(B)\big]\,\mathbb{E}\big[K^{[kl]}_m(d\boldsymbol{y})\big].$$

The first two expectations are given in (5.5). Moreover, since K is a Poisson process, one has $\mathbb{E}\big[K^{[kl]}(d\boldsymbol{y})\big] = \mathbb{E}\big[H_{km}(dy_1)\big]\,\mathbb{E}\big[H_{lm}(dy_2)\big]$. Finally apply two times Equation (4.4), and the claim follows with:

$$\mathbb{E}\big[\Upsilon_{N^+N^+:[K]}(A\times B)\big]$$
$$= \sum_{k,l=1}^{d}\int_{\mathbb{E}^2} U^+_{ik,y_1}(A)U^+_{jl,y_2}(B)H_{km}(dy_1)H_{lm}(dy_2)$$
$$= \Big[\sum_{k=1}^{d}\int_{\mathbb{E}} U^+_{ik,y_1}(A)H_{km}(dy_1)\Big]\Big[\sum_{l=1}^{d}\int_{\mathbb{E}} U^+_{jl,y_2}(B)H_{lm}(dy_2)\Big]$$
$$= U^-_{im}(A)U^-_{jm}(B) = \Upsilon_{U^-U^-}(A\times B). \qquad \square$$

Proof (Lemma 5.8): This is the third part of the proof, concerning the representation for a Hawkes process.

(3a) *Relationship between any two events inside the same tree.* First apply the first part of Equation (3.8) and then use that J is a Poisson process with mean measure family $\boldsymbol{G}$. The statement follows with:

$$\mathbb{E}\big[\Upsilon_{(N^+):J}\big] = \mathbb{E}\Big[\sum_{k=1}^{d}\int_{\mathbb{E}} N^{+(ij)}_{k,x}(A\times B)J^{(k)}(dx)\Big]$$
$$= \sum_{k=1}^{d}\int_{\mathbb{E}}\mathbb{E}\big[N^{+(ij)}_{k,x}(A\times B)\big]\,\mathbb{E}\big[J^{(k)}(dx)\big]$$
$$= \sum_{k=1}^{d}\int_{\mathbb{E}} M^{+(ij)}_{k,x}(A\times B)G_k(dx) = \Upsilon_{(M^+):G}.$$

(3b) *Relationship between two different events inside the same tree.* The calculation for $\Upsilon_{[N^+]:J}$ is almost the same as for $\Upsilon_{(N^+):J}$, except that there is a factorial instead of an ordinary product measure.

(3c) *Relationship between two events from two different trees.* Due to the second part of Equation (3.9), one gets in a first step:

$$\begin{aligned}\mathbb{E}\big[\Upsilon_{N^+N^+:[J]}\big] &= \mathbb{E}\Big[\sum_{k,l=1}^{d}\int_{\mathbb{E}^2} N^{+(i)}_{k,x_1}(A)N^{+(j)}_{l,x_2}(B)J^{[kl]}(d\boldsymbol{x})\Big]\\ &= \sum_{k,l=1}^{d}\int_{\mathbb{E}^2}\mathbb{E}\big[N^{+(i)}_{k,x_1}(A)\big]\,\mathbb{E}\big[N^{+(j)}_{l,x_2}(B)\big]\,\mathbb{E}\big[J^{[kl]}(d\boldsymbol{x})\big].\end{aligned}$$

Since $\mathbb{E}\big[J^{[kl]}(d\boldsymbol{x})\big] = M_J{}^{[kl]}(d\boldsymbol{x}) = G_k(dx_1)G_l(dx_2)$, and due to Equation (5.5), the claim now follows with:

$$\begin{aligned}\mathbb{E}\big[\Upsilon_{N^+N^+:[J]}\big] &= \sum_{k,l=1}^{d}\int_{\mathbb{E}^2} M^{+(i)}_{k,x_1}(A)M^{+(j)}_{l,x_2}(B)G_k(dx_1)G_l(dx_2)\\ &= \Big[\sum_{k=1}^{d}\int_{\mathbb{E}} U^{+}_{ik,x_1}(A)G_k(dx_1)\Big]\Big[\sum_{l=1}^{d}\int_{\mathbb{E}} U^{+}_{jl,x_2}(B)G_l(dx_2)\Big]\\ &= \Upsilon_{U^+:G|U^+:G}.\end{aligned}$$

□

Proof (Theorem 5.4): This is the first part of the proof, concerning the explicit representation.

(i) *Ordinary moment measure.* Start with the decomposition of N^+_m given in (4.18) and take expectations on both sides. Then, due to Lemma 5.8, one finds:

$$\begin{aligned}M^{+(ij)}_m(A\times B) &= \mathbb{E}\Big[\Upsilon_{\delta\delta:L} + \Upsilon_{\delta N^-:L} + \Upsilon_{N^-\delta:L} + \Upsilon_{N^+N^+:[K]:L}\Big]\\ &= \Upsilon_{\delta\delta:U^+} + \Upsilon_{\delta U^-:U^+} + \Upsilon_{U^-\delta:U^+} + \Upsilon_{U^-U^-:U^+}.\end{aligned}$$

Replace the placeholders with their definitions, and apply (4.3) twice, so

that:

$$
\begin{aligned}
M_m^{+(ij)}(A \times B) &= \delta_{ij} \int_{\mathbb{E}} \delta_y(A)\delta_y(B) U_{im}^+(dy) + \int_{\mathbb{E}} \delta_y(A) U_{ji,y}^-(B) U_{im}^+(dy) \\
&+ \int_{\mathbb{E}} \delta_y(B) U_{ij,y}^-(A) U_{jm}^+(dy) + \sum_{r=1}^{d} \int_{\mathbb{E}} U_{ir,y}^-(A) U_{jr,y}^-(B) U_{rm}^+(dy) \\
&= \sum_{r=1}^{d} \int_{\mathbb{E}} \Big[\delta_{ir}\delta_y(A) + U_{ir,y}^-(A)\Big] \Big[\delta_{jr}\delta_y(B) + U_{jr,y}^-(B)\Big] U_{rm}^+(dy) \\
&= \sum_{r=1}^{d} \int_{\mathbb{E}} U_{ir,y}^+(A) U_{jr,y}^+(B) U_{rm}^+(dy) = \Upsilon_{U^+U^+:U^+}.
\end{aligned}
$$

(*ii*) *Factorial moment measure.* This time start with the second decomposition given in (4.18). Take expectations on both sides and use Lemma 5.8, so that:

$$
\begin{aligned}
M_m^{+[ij]}(A \times B) &= \mathbb{E}\Big[\Upsilon_{\delta N^-:L} + \Upsilon_{N^-\delta:L} + \Upsilon_{N^+N^+:[K]:L}\Big] \\
&= \Upsilon_{\delta U^-:U^+} + \Upsilon_{U^-\delta:U^+} + \Upsilon_{U^-U^-:U^+}.
\end{aligned}
$$

After substitution of the placeholders, the claimed formula follows. □

Proof (Theorem 5.4): This is the second part of the proof, concerning the implicit representation.

(*i*) *Ordinary moment measure.* Start with the decomposition of $N^{+(ij)}$ from the first part of (4.19) and take expectations on both sides. Due to Lemma 5.8, one has in a first step:

$$
\begin{aligned}
M_m^{+(ij)}(A \times B) &= \mathbb{E}\Big[\Upsilon_{\delta\delta} + \Upsilon_{\delta|N^+:K} + \Upsilon_{N^+:K|\delta} + \Upsilon_{(N^+):K} + \Upsilon_{N^+N^+:[K]}\Big] \\
&= \Upsilon_{\delta\delta} + \Upsilon_{\delta U^-} + \Upsilon_{U^-\delta} + \Upsilon_{(M^+):H} + \Upsilon_{U^-U^-} \\
&= \Upsilon_{(M^+):H} + \Big[\Upsilon_{\delta\delta} + \Upsilon_{\delta U^-} + \Upsilon_{U^-\delta} + \Upsilon_{U^-U^-}\Big].
\end{aligned}
$$

It remains to show that the last four terms are equal to $\Upsilon_{U^+U^+}$. But

with the help of (4.3), this follows from:

$$\begin{aligned}
&\Upsilon_{\delta\delta} + \Upsilon_{\delta U^-} + \Upsilon_{U^-\delta} + \Upsilon_{U^-U^-} \\
&\quad = \delta_{ijm}\delta_0(A)\delta_0(B) + \delta_{im}\delta_0(A)U^-_{jm}(B) \\
&\qquad + \delta_{jm}\delta_0(B)U^-_{im}(A) + U^-_{im}(A)U^-_{jm}(B) \\
&\quad = \Big[\delta_{im}\delta_0(A) + U^-_{jm}(A)\Big]\Big[\delta_{jm}\delta_0(B) + U^-_{im}(B)\Big] \\
&\quad = U^+_{im}(A)U^+_{jm}(B) = \Upsilon_{U^+U^+}.
\end{aligned}$$

In summary one has:

$$M^{+(ij)}_m(A \times B) = \Upsilon_{(M^+):H} + \Upsilon_{U^+U^+}.$$

The claim follows if one substitutes the definitions of the placeholders.

(*ii*) *Factorial moment measure.* This time start with the second part of (4.19) and again take expectations on both sides. Due to Lemma 5.8, one finds:

$$\begin{aligned}
M^{+[ij]}_m(A \times B) &= \mathbb{E}\Big[\Upsilon_{\delta|N^+:K} + \Upsilon_{N^+:K|\delta} + \Upsilon_{[N^+]:K} + \Upsilon_{N^+N^+:[K]}\Big] \\
&= \Upsilon_{\delta U^-} + \Upsilon_{U^-\delta} + \Upsilon_{[M^+]:H} + \Upsilon_{U^-U^-}.
\end{aligned}$$

The claim follows if one substitutes the definitions of the placeholders. □

Proof (Theorem 5.4): This is the third part of the proof, concerning a Hawkes process.

(*i*) *Ordinary moment measure.* Start with the first part of (4.20) and take expectations on both sides. The claimed formula follows due to Lemma 5.8 with:

$$M^{(ij)}(A \times B) = \mathbb{E}\Big[\Upsilon_{(N^+):J} + \Upsilon_{N^+N^+:[J]}\Big] = \Upsilon_{(M^+):G} + \Upsilon_{U^+:G|U^+:G}.$$

(*ii*) *Factorial moment measure.* This time take the second part of (4.20). Again due to Lemma 5.8, the claim follows with:

$$M^{[ij]}(A \times B) = \mathbb{E}\Big[\Upsilon_{[N^+]:J} + \Upsilon_{N^+N^+:[J]}\Big] = \Upsilon_{[M^+]:G} + \Upsilon_{U^+:G|U^+:G}.$$ □

Proof (Lemma 5.9): There is nothing to prove. We simply need to substitute the placeholders from Lemma 5.8 and check whether we indeed get the claimed equations. □

Reduced Moment Measures

From now on we will restrict ourselves to univariate Hawkes processes, see also Remark 5.6. Recall the Definition 2.22 of the pseudo-reduction $\mathring{\mu}$ of a finite, symmetric measure μ on $\mathbb{E}^n$.

Proof (Lemma 5.10): This is the first part of the proof, concerning the explicit representation.

(1a) *Relationship between an event and its descendants.* Note that neither $\Upsilon_{\delta U^-:U^+}$ nor $\Upsilon_{U^-\delta:U^+}$ is symmetric, but the sum of the two is symmetric. If we treat the pair as a unit, we do not have to distinguish between different versions of reduced measures. Now consider first the expression:

$$\begin{aligned}\int_{\mathbb{E}} \Upsilon_{\delta U^-:U^+}(A + w \times dw) &= \int_{\mathbb{E}^2} \delta_y(A + w) U_y^-(dw) U^+(dy) \\ &= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}} \delta_y(A + w) U_y^-(dw) \Big] U^+(dy) = \int_{\mathbb{E}} \Big[\int_{y-A} U_y^-(dw) \Big] U^+(dy).\end{aligned}$$

Due to (4.5), one obtains the intermediate result:

$$\begin{aligned}\int_{\mathbb{E}} \Upsilon_{\delta U^-:U^+}(A + w \times dw) &= \int_{\mathbb{E}} U_y^-(y - A) U^+(dy) \\ &= \int_{\mathbb{E}} U^-(-A) U^+(dy) = U^-(-A) U^+(\mathbb{E}) = (1 - Q)^{-1} U^-(-A).\end{aligned}$$

In the same way, one obtains a second intermediate result:

$$\begin{aligned}\int_{\mathbb{E}} \Upsilon_{U^-\delta:U^+}(A + w \times dw) &= \int_{\mathbb{E}^2} \delta_y(dw) U_y^-(A + w) U^+(dy) \\ &= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}} \delta_y(dw) U_y^-(A + w) \Big] U^+(dy) = \int_{\mathbb{E}} U_y^-(A + y) U^+(dy) \\ &= \int_{\mathbb{E}} U^-(A) U^+(dy) = U^-(A) U^+(\mathbb{E}) = (1 - Q)^{-1} U^-(A).\end{aligned}$$

Due to these two equations, one now finds:

$$\begin{aligned}\big[\mathring{\Upsilon}_{\delta U^-:U^+} + \mathring{\Upsilon}_{U^-\delta:U^+}\big](A) &= \int_{\mathbb{E}} \big[\Upsilon_{\delta U^-:U^+} + \Upsilon_{U^-\delta:U^+}\big](A + w \times dw) \\ &= \int_{\mathbb{E}} \Upsilon_{\delta U^-:U^+}(A + w \times dw) + \int_{\mathbb{E}} \Upsilon_{U^-\delta:U^+}(A + w \times dw) \\ &= (1 - Q)^{-1} U^-(-A) + (1 - Q)^{-1} U^-(A).\end{aligned}$$

(1b) *Relationship between events not related to each other in direct line.* Recall that by definition $U_y^+(dw) = U^+(dw - y)$. Due to (4.5) one knows that $U^+(\mathbb{E}) = (1-Q)^{-1}$, so that:

$$\begin{aligned}
\mathring{\Upsilon}_{U^-U^-:U^+}(A) &= \int_{\mathbb{E}} \Upsilon_{U^-U^-:U^+}(A + w \times dw) \\
&= \int_{\mathbb{E}^2} U_y^-(A+w) U_y^-(dw) U^+(dy) \\
&= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}} U^-(A + w - y) U^-(dw - y) \Big] U^+(dy) \\
&= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}} U^-(A + w) U^-(dw) \Big] U^+(dy) \\
&= \Big[\int_{\mathbb{E}} U^+(dy) \Big] \Big[\int_{\mathbb{E}} U^-(A+w) U^-(dw) \Big] \\
&= (1-Q)^{-1} \int_{\mathbb{E}} U^-(A+w) U^-(dw).
\end{aligned}$$

(1c) *Relationship between any event and all others, including its duplicate.* The measure $\Upsilon_{U^+U^+:U^+}$ can be treated in the same way as $\Upsilon_{U^-U^-:U^+}$ and a similar calculation shows that:

$$\begin{aligned}
\mathring{\Upsilon}_{U^+U^+:U^+}(A) &= \int_{\mathbb{E}^2} U_y^+(A+w) U_y^+(dw) U^+(dy) \\
&= (1-Q)^{-1} \int_{\mathbb{E}} U^+(A+w) U^+(dw).
\end{aligned}$$

□

Proof (Lemma 5.10): This is the second part of the proof, concerning the implicit representation.

(2a) *Relationship between the root node and the family of its children, in both directions.* We prefer again to treat $\Upsilon_{\delta U^-:U^+}$ and $\Upsilon_{U^-\delta:U^+}$ as a unit, since the sum of the two measures is symmetric. Hence:

$$\begin{aligned}
\big[\mathring{\Upsilon}_{\delta U^-} + \mathring{\Upsilon}_{U^-\delta}\big](A) &= \int_{\mathbb{E}} \Upsilon_{\delta U^-}(A + w \times dw) + \int_{\mathbb{E}} \Upsilon_{U^-\delta}(A + w \times dw) \\
&= \int_{\mathbb{E}} \delta_0(A+w) U^-(dw) + \int_{\mathbb{E}} \delta_0(dw) U^-(A+w) \\
&= U^-(-A) + U^-(A).
\end{aligned}$$

(2b) *Relationship among family members including duplicates of the children*

of the root node. From (5.4) one knows that ${M_y}^{+(2)}(d\boldsymbol{z}) = M^{+(2)}(d\boldsymbol{z}-y)$, and due to (4.2) one has $H(\mathbb{E}) = Q$, so that:

$$\begin{aligned}
\mathring{\Upsilon}_{(M^+):H}(A) &= \int_{\mathbb{E}} \Upsilon_{(M^+):H}(A + w \times dw) \\
&= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}} M_y^{+(2)}(A + w \times dw) \Big] H(dy) \\
&= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}} M^{+(2)}(A + w - y \times dw - y) \Big] H(dy) \\
&= \int_{\mathbb{E}} \Big[\int_{\mathbb{E}} M^{+(2)}(A + w \times dw) \Big] H(dy) \\
&= \int_{\mathbb{E}} \mathring{M}^{+(2)}(A) H(dy) = Q\mathring{M}^{+(2)}(A).
\end{aligned}$$

(2c) *Relationship among different family members of the children of the root node.* This is the same calculation as above, except for a factorial instead of an ordinary product moment measure:

$$\mathring{\Upsilon}_{[M^+]:H}(A) = \int_{\mathbb{E}^2} M_y^{+[2]}(A + w \times dw) H(dy) = Q\mathring{M}^{+[2]}(A).$$

(2d) *Relationship between two families of two different children of the root node, the root node exclusive.* This expression cannot be simplified much and one obtains:

$$\mathring{\Upsilon}_{U^-U^-}(A) = \int_{\mathbb{E}} \Upsilon_{U^-U^-}(A + w \times dw) = \int_{\mathbb{E}} U^-(A + w) U^-(dw).$$

(2e) *Relationship between two families of two different children of the root node, the root node inclusive.* This follows in the same way as the previous statement. □

Proof (Lemma 5.10): This is the third part of the proof, concerning a Hawkes process. Recall that now $G(dx) = dx$.

(3a) *Relationship between any two events inside the same tree.* First note that:

$$\Upsilon_{(M^+):G}(d\boldsymbol{z}) = \int_{\mathbb{E}} M_x^{+(2)}(d\boldsymbol{z}) dx = \int_{\mathbb{E}} M^{+(2)}(d\boldsymbol{z} - x) dx.$$

This integral is of the same form as in (2.8), so that:

$$\breve{\Upsilon}_{(M^+):G} = \mathring{M}^{+(2)}(A).$$

(3b) *Relationship between two different events inside the same tree.* This is almost the same calculation, except for a factorial instead of an ordinary moment measure.

(3c) *Relationship between two events from two different trees.* Note in a first step that:

$$\begin{aligned}\Upsilon_{U^+:G|U^+:G}(d\boldsymbol{z}) &= \Big[\int_{\mathbb{E}} U^+_{x_1}(dz_1)dx_1\Big]\Big[\int_{\mathbb{E}} U^+_{x_2}(dz_2)dx_2\Big] \\ &= \prod_{r=1}^{2}\Big[\int_{\mathbb{E}} U^+_{x_r}(dz_r)dx_r\Big] = \prod_{r=1}^{2}\Big[\int_{\mathbb{E}} U^+(dz_r - x_r)dx_r\Big].\end{aligned}$$

This expression is of the same form as in (2.9), which shows that:

$$\breve{\Upsilon}_{U^+:G|U^+:G} = \big[U^+(\mathbb{E})\big]^2\lambda_{\mathbb{E}}(A) = (\mathbb{1}_d - Q)^{-2}\lambda_{\mathbb{E}}(A). \qquad \square$$

Proof (Theorem 5.7): Most of the required calculation has already been done in Lemma 5.10.

(1) *Explicit representation for Hawkes trees.* Take the decompositions from Equation (5.9) and apply the pseudo-reduction operation on both sides:

$$\begin{aligned}\mathring{M}^{+(2)}(A) &= \mathring{\Upsilon}_{U^+U^+:U^+}, \\ \mathring{M}^{+[2]}(A) &= \mathring{\Upsilon}_{\delta U^-:U^+} + \mathring{\Upsilon}_{U^-\delta:U^+} + \mathring{\Upsilon}_{U^-U^-:U^+}.\end{aligned}$$

Now replace the placeholders with the expressions given in Lemma 5.10.

(2) *Implicit representation for Hawkes trees.* This time take the decompositions from Equation (5.10). Then take again the pseudo-reduction on both sides, so that:

$$\begin{aligned}\mathring{M}^{+(2)}(A) &= \mathring{\Upsilon}_{(M^+):H} + \mathring{\Upsilon}_{U^+U^+}, \\ \mathring{M}^{+[2]}(A) &= \mathring{\Upsilon}_{[M^+]:H} + \mathring{\Upsilon}_{\delta U^-} + \mathring{\Upsilon}_{U^-\delta} + \mathring{\Upsilon}_{U^-U^-}.\end{aligned}$$

Again substitute the placeholders with the explicit expressions.

(3) *Representation for Hawkes processes.* Take the univariate versions of

Equation (5.11) and then take the regular reduction on both sides:

$$\breve{M}^{(2)}(A) = \breve{\Upsilon}_{(M^+):G} + \breve{\Upsilon}_{U^+:G|U^+:G},$$
$$\breve{M}^{[2]}(A) = \breve{\Upsilon}_{[M^+]:G} + \breve{\Upsilon}_{U^+:G|U^+:G}.$$

Now replace the placeholders with their definitions. □

Chapter 6

Intensities

This chapter is more or less independent of the previous ones and the statements and proofs are self-contained. It should therefore be possible to read this chapter independently of the others. As the title suggests, the main topic concerns intensity processes and how they can be used to define Hawkes processes. Although we have already defined Hawkes processes, it is well worth giving an alternative definition using intensity processes. Because one can only reasonably define intensity processes if there is some sort of time dimension, we will assume from now on that the event space is $\mathbb{E} := \mathbb{R}$.

6.1 Motivation and Objectives

The results given in this chapter are based strongly on a series of papers written by Brémaud, Massoulié and Torrisi. Notably, these are the papers [BM96], [Mas98], [BNT02] and [Tor02], check also the bibliography. Additionally, some standard results are taken from the book [DVJ03].

The knowledgable reader might not find too much that is surprising in the following lines. In some sense, the material presented here is even less general than the results given in the aforementioned papers. But I would still like to point out a few differences and additions to the existing literature and also explain the reason for my approach.

In writing this chapter, it was my intent to work out a few of the concepts more clearly. By breaking up some of the original proofs, the main ideas can hopefully be better appreciated.

Point Configuration Spaces. In dealing with point processes, one is quickly confronted with the abstract definitions of point configuration spaces. I wanted to use an approach as simple as possible, but still powerful enough so that all the given results can be stated rigourously and in full generality.

Usually, the predictable σ-algebra is defined based on continuous or left-continuous processes. But there is the well-known fact that if a filtration is generated by a point process, an alternative, direct and intuitive characterization of predictable processes exists. The price one has to pay is that this convenient characterization of predictability brakes down one if one would take a general filtration.

I decided to define from the beginning on a canonical probability space. It is used all over, with a few exceptions, as the underlying probability space. This has a few advantages: Because the driving process is a point process, the natural filtration on the canonical probability space inherits all the nice properties mentioned above. This allows to give more compact, and more intuitively clear definitions, see e.g. the definition of an intensity process in Definition 6.9.

The canonical probability space is actually a point configurations space, which implies that elementary events ω are point configurations. To emphasize the special nature of the elementary events from the canonical probability space, I call them ϖ, instead of ω. But more importantly, ϖ can and will be interpreted as the underlying *driving process*. It is then a matter of choosing an appropriate canonical probability space, which is rich enough, so that all required stochastic objects can be constructed.

The mentioned characterization of predictable processes can be found in the book [Bré81]. The papers [Jac75] and [Las93] have also served as a rich, informative source.

Intensities Processes. I make a distinction between the standard intensity function, see Definition 6.9, and what I call the *intrinsic* intensity function, see Definition 6.13. Basically, an intrinsic intensity function γ is a function such that $\lambda(t|\varpi) = \gamma(t|N)$, where N is a point process and λ is the intensity process of N.

I will usually explicitly mention the driving process and write $\lambda(t|\varpi)$ for the intensity process, instead of only $\lambda(t)$. For this reason, I often speak about a function and not a process, although this is a matter of taste. It has also the advantage that the notation becomes more regular: The intensity functions

$\lambda(t|\varpi)$ are then of the same form as the intrinsic intensity functions, which are of the form $\gamma(t|\nu)$.

Note that in a general setting, the intensity process λ and the intrinsic intensity function γ would be quite different objects. But the chosen approach hopefully helps to shed some light onto the similarities. One should also point out that intrinsic intensity functions are not specific to Hawkes processes.

The definition of Hawkes processes is split up into two parts: Firstly, the definition of the *intrinsic Hawkes intensity function* is given. For simplicity, I call it the Hawkes intensity function, see Definition 6.19. Secondly, a Hawkes process is defined as the solution of a thinning problem. Basically, a Hawkes process is a point process N which satisfies the implicit Equation (6.9), see Definition 6.33.

The definition in two steps has a couple of advantages: It is possible to look at the Hawkes intensity function separately, without even having to use a probability space. Indeed, the probability measure is only introduced in Definition 6.28, when the thinning part of the definition comes into play.

This approach differentiates between a first deterministic part and a second probabilistic part of the definition. It should also be noted that the thinning construction, and especially the implicit Equation (6.9), are not specific to Hawkes processes. Indeed, a quite general class of point processes could be defined in the same way.

Initial States and Continuations. In [BNT02] and [Tor02], the rate of convergence of a transient, i.e. non-stationary, Hawkes process with some initial condition is analyzed. In these papers, the idea of a what I call a continuation appears indirectly. A formal definition is given in Definition 6.35. Note that an *initial state*, given in Definition 6.34, is a deterministic point configuration, and this deviates slightly from what is called a *initial condition* in the references above.

I give a precise definition of an initial condition and its continuation for three reasons: Firstly, the definition of a continuation parallels the one of a strong solution very nicely and a couple of almost trivial, but interesting relations between the two concepts can be found. Secondly, the proof of uniqueness of a strong solution, Theorem 6.55, actually relies on uniqueness of continuations, Corollary 6.51. By introducing the notion of a continuation, this step of the proof can be examined separately. Thirdly, if one wants to analyze the speed of convergence to equilibrium, continuations are what one essentially is looking at, see the coupling result in Proposition 6.52

As can be seen in the paper [Tor02], if one deals with Hawkes processes, a lot of calculations involve convolutions of matrix-valued functions. By using convolution notation consequently, quite compact results can be obtained, as e.g. in Equation (6.12). I also wanted to avoid complications in another way: Whenever possible, I prefer an explicit equation to an implicit equation. E.g. the first moment measure densities in Propositions 6.47 and 6.48 are derived without the need to solve a Markov renewal equation.

Hazard Rates. For the proof of Proposition 6.43, a result concerning conditional extinction probabilities is required. The starting point was Lemma 1 in [BM96]. But my aim was to avoid the use of the cutoff operator that appears in there. Instead I introduced what I call a *hazard rate*, see Definition 6.38, to clarify the idea behind this result, which goes back to [Jac75]. Clearly, the notion of a hazard rate is not new, but the definitions I am aware of are not as general as the one used in Definition 6.38 and Corollary 6.40.

6.2 Canonical Spaces

To explain the concepts more clearly, we restrict ourselves to the situation where we have a univariate point process without marks, i.e. a point process with values in $\mathscr{N}(\mathbb{R})$. The general case is similar.

6.1 Definition (Restrictions and projections). Let $D \subseteq \mathbb{R}$ and fix a point configuration $\nu \in \mathscr{N}(\mathbb{R})$. The *restriction* ν_D, which is a point configuration in $\mathscr{N}(D)$, is

$$\nu_D(dt) := \nu(dt \cap D).$$

The *projection* π_D is the function

$$\pi_D : \mathscr{N}(\mathbb{R}) \to \mathbb{N}_0, \qquad \pi_D(\nu) := \nu(D). \qquad \diamondsuit$$

The following definition is equivalent to Definition 3.1, but it should help clarify the role of the projection operator π_D. Recall that we stick to the following convention: Whenever we say we take a subset of $\mathbb{R}$, we actually mean a *measurable* subset.

6.2 Definition (Canonical filtration). On the space $\mathscr{N}(\mathbb{R})$, define the σ-

algebras:

$$\mathfrak{F} := \sigma\big\{\pi_D, D \subseteq \mathbb{R}\big\} \qquad \text{and} \qquad \mathfrak{F}_t := \sigma\big\{\pi_D, D \subseteq (-\infty, t]\big\},$$

for $t \in \mathbb{R}$. The associated *filtration* is $\mathfrak{F}_\bullet := (\mathfrak{F}_t)_{t\in\mathbb{R}}$.

Consider now the space $\mathfrak{N} \times \mathbb{R}$. The predictable σ-algebra $\mathfrak{P}$ is defined on this space and given by

$$\mathfrak{P} := \sigma\Big\{\mathfrak{F}_t \otimes \mathscr{B}_{(t,\infty)}, t \in \mathbb{R}\Big\}.$$

Clearly, $\mathfrak{P}$ is a sub-σ-algebra of $\mathfrak{F} \otimes \mathscr{B}(\mathbb{R})$. $\diamondsuit$

More about the Borel-σ-algebra on a measure space and the associated Prohorov metric on the space of finite measures can be found in Appendix 2 of [DVJ03]. Compare also the above definition with Proposition A2.5.IV in this reference.

For the following definition, take a space $\mathbb{Y}$ of the form $\mathbb{Y} := \mathbb{X} \times \mathbb{R}_+$, where $\mathbb{X} := \mathbb{R}^e$ is a mark space. The space $\mathbb{Y}$ has the interpretation of an *extended* mark space, where $\mathbb{R}_+$ is used in for a thinning procedure given later.

Whenever we speak about a *canonical probability space* we refer to a probability space of the following form:

6.3 Definition (Canonical measurable space). The *canonical measurable space* consists of the triple $(\mathfrak{N}, \mathfrak{F}, \mathfrak{F}_\bullet)$, where one has the sample space $\mathfrak{N} := \mathscr{N}_{\mathbb{Y}}(\mathbb{R})$, the associated σ-algebra $\mathfrak{F}$, and the filtration $\mathfrak{F}_\bullet$.

The elements of the sample space $\mathfrak{N}$ are denoted by ϖ, that is

$$\varpi : \mathfrak{N} \to \mathfrak{N}, \qquad \varpi := \mathrm{id}_{\mathfrak{N}}, \qquad \varpi \mapsto \varpi. \qquad \diamondsuit$$

One can think of ϖ as a *driving process*, from which other processes are constructed: The symbol ϖ is used to remind one that the sample events are not arbitrary sample events but point configurations.

The idea of a driving process can be found e.g. in section 3 of [BM96]. Because the driving process is all that is needed in the following exposition, we decided to introduce at this point the canonical probability space. This simplifies some of the definitions given later.

In the definition above, no associated probability measure $\mathbb{P}$ is specified. $\mathbb{P}$ will be defined later, when the thinning procedure becomes relevant.

6.4 Definition (Processes on canonical space). Assume $\mathfrak{N}$ is a canonical measurable space with driving process ϖ.

(1) *Stochastic process on canonical space.* Assume H is a function of the form

$$H : \mathfrak{N} \times \mathbb{R} \to \mathbb{R}.$$

Then H is called a *stochastic process* if it is measurable with respect to the σ-algebra $\mathfrak{F} \otimes \mathscr{B}(\mathbb{R})$. It is called *predictable* if it is measurable with respect to the σ-algebra $\mathfrak{P}$.

The following is equivalent to the above definition: A stochastic process H on $\mathfrak{N}$ is *predictable* if and only if for all $t \in \mathbb{R}$:

H_t is $\mathfrak{F}_{t-}$-measurable.

(2) *Point process on canonical space.* A function N of the form

$$N : \mathfrak{N} \to \mathscr{N}(\mathbb{R})$$

is called a *point process* if it is measurable with respect to the canonical σ-algebras on the two spaces. $\diamondsuit$

Note that we do not define predictable point processes, as such processes would rarely make sense.

The above definition of predictability is not the usual one found in the literature. For the standard definition see e.g. Definition D4 in Section I.3 of [Bré81]. But here we are working with the canonical filtration, which is the internal history of a point process. The predictable σ-algebra generated by the internal history has especially nice properties. As one of the convenient consequences, predictable processes can be directly characterized, see e.g. Exercise E4 in Section III.2 or Theorem T34 in Section A2.3 of [Bré81].

We use this specific definition of predictability, since it is less abstract and more intuitive. The price one pays is that it cannot be generalized to arbitrary filtrations. See also Section 1 of [Las93], where the same characterization is used.

Next, we consider a filtration generated by some point process, i.e. not necessarily the canonical filtration. Note that this filtration is therefore still an internal history. We extend the notions of measurability and predictability to this case, but make use of the previous definition for the case of a canonical filtration. This leads to the following compact definition:

6.5 Definition (Induced filtration). Let N be a point process defined on the canonical probability space $\mathfrak{N}$ with values in $\mathscr{N}(\mathbb{R})$.

(1) *Induced σ-algebra.* The σ-algebra $\mathscr{F}^N$ and the element $\mathscr{F}_t^N$ of the filtration $\mathscr{F}_\bullet^N$ are given by

$$\mathscr{F}^N := N^{-1}(\mathfrak{F}) \qquad \text{and} \qquad \mathscr{F}_t^N := N^{-1}(\mathfrak{F}_t).$$

We define the *induced filtration* by $\mathscr{F}_\bullet^N := (\mathscr{F}_t^N)_{t\in\mathbb{R}}$.

(2) *Induced predictable σ-algebra.* The predictable σ-algebra $\mathscr{P}^N$ is defined on the space $\mathfrak{N}\times\mathbb{R}$ and is a sub-σ-algebra of the form

$$\mathscr{P}^N := \sigma\Big\{\mathscr{F}_t^N \otimes \mathscr{B}_{(t,\infty)}, t\in\mathbb{R}\Big\}, \qquad \mathscr{P}^N \subseteq \mathscr{F}^N \otimes \mathscr{B}(\mathbb{R}). \qquad \diamondsuit$$

Since the two notions *canonical* and *induced* will appear often, let us clarify their exact meaning: We use the term *canonical* whenever we want to emphasize that the measurability is considered with respect to the *canonical* measurable space $\mathfrak{N}$ and its driving process ϖ.

Similarly, we use the term *induced* to emphasize the measurability is considered with respect to the filtration *induced* by some point process N.

6.6 Definition (Induced measurability). Let $\mathfrak{N}$ be a canonical probability space equipped with a point process N with values in some point configuration space $\mathscr{N}(\mathbb{R})$. As explained, N generates the induced filtration $\mathscr{F}_\bullet^N$.

Let X be an arbitrary stochastic process on $\mathfrak{N}$. Instead of canonical measurable and canonical predictable, one defines the following extension of these two notions. We say that X is:

$$\begin{array}{lcl} N\text{-}measurable & \Leftrightarrow & \mathscr{F}^N\otimes\mathscr{B}(\mathbb{R})\text{-measurable}, \\ N\text{-}predictable & \Leftrightarrow & \mathscr{P}^N\text{-measurable}. \end{array} \qquad \diamondsuit$$

Based on this definition, one obtains the following characterization: Assume L is a point process on $\mathfrak{N}$. Then L is N-*measurable*, or even N-*predictable*, if and only if the associated counting process

$$H(t) := N\big((-\infty,t]\big), \text{ where } t\in\mathbb{R},$$

is N-measurable, or even N-predictable.

So far, only univariate point processes without marks were treated. But the definitions above extend easily to the case of multivariate and marked point

processes. We demonstrate this for the predictable σ-algebra $\mathscr{P}^N$; the idea should then become clear:

6.7 Remark (Predictable σ-algebra, revisited). Consider some canonical probability space $\mathfrak{N}$.

(1) *Marked point process.* Let N be a marked point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$, for some univariate mark space $\mathbb{X}$.

The idea is to consider the projections $N_{\mathbb{R}\times X}(dt) := N(dt \times X)$, for arbitrary sets $X \subseteq \mathbb{X}$. The induced predictable σ-algebra $\mathscr{P}^N$ is then defined as

$$\mathscr{P}^N := \sigma\Big\{ \mathscr{P}^K, \text{ for all } K := N_{\mathbb{R}\times X} \text{ with } X \subseteq \mathbb{X} \Big\}.$$

(2) *Multivariate point process.* Let N be a multivariate marked point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$, where $\mathbb{X}$ is some multivariate mark space. In the same way as above, the induced predicable σ-algebra $\mathscr{P}^N$ is given by

$$\mathscr{P}^N := \sigma\Big\{ \mathscr{P}^K, \text{ for all } K := N^{(j)} \text{ with } 1 \leq j \leq d \Big\}.$$

$N^{(j)}$ denotes the j-th component process of N. ♢

6.8 Definition (Intensity kernel). Let $\mathscr{N}_{\mathbb{Y}}(\mathbb{R})$ be a canonical probability space with driving process ϖ. Assume $\lambda(dt|\varpi)$ is a family of measures on $\mathbb{R}$, indexed with a parameter ϖ from the space $\mathscr{N}_{\mathbb{Y}}(\mathbb{R})$.

We say λ is a *locally-finite, predictable sub-probability kernel*, or in short an *intensity kernel*, if λ satisfies the following three conditions:

(1) It is a *kernel* in the sense that

(1) $\lambda(\cdot|\varpi)$ is a locally-finite measure on $\mathbb{R}$, for all $\varpi \in \mathscr{N}_{\mathbb{Y}}(\mathbb{R})$.

(2) $\varpi \mapsto \lambda(E|\varpi)$ is an $\mathfrak{F}$-measurable function, for all $E \subseteq \mathbb{R}$.

(2) It is a *sub-probability*, i.e. it satisfies

$$\lambda(\{t\}|\varpi) \leq 1, \text{ for all } t \in \mathbb{R} \text{ and } \varpi \in \mathscr{N}_{\mathbb{Y}}(\mathbb{R}).$$

(3) It is *predictable* with respect to the filtration $\mathfrak{F}_\bullet$ in the following sense: For all $t \in \mathbb{R}$ it holds that

$$\lambda(E|\varpi) = \lambda(E|\varpi_{(-\infty,t)}), \text{ for all } E \subseteq (-\infty, t].$$ ♢

The above definition is essentially taken from Section 1 of [Las93], but some additional conditions have been left out, as we are exclusively dealing with locally-finite, simple point processes.

We have given the definition of an intensity kernel only in the case of a univariate point process. In the same way as before, the extension to the general case is trivial.

6.9 Definition (Intensity measure). Let $\mathfrak{N} := \mathscr{N}_{\mathbb{Y}}(\mathbb{R})$ be a univariate canonical measurable space with driving process ϖ and let $\mathbb{X}$ be some mark space.

(1) *Base measure.* Let $\mu_{\mathbb{X}}$ be a measure on $\mathbb{X}$. If $\mu_{\mathbb{X}}$ is used in the sense given below, then it is called a *base measure*.

(2) *Density of intensity kernel.* Let $\lambda(dt \times dx|\varpi)$ be an intensity kernel and assume we can decompose it according to

$$\lambda(dt \times dx|\varpi) = \lambda(t, x|\varpi)\mu_{\mathbb{X}}(dx)dt = f(x|t, \varpi)\lambda(t|\varpi)\mu_{\mathbb{X}}(dx)dt. \quad (6.1)$$

Then $\lambda(t, x|\varpi)$ is called the *time-space-intensity function*, $\lambda(t|\varpi)$ the *time-intensity function* and $f(x|t, \varpi)$ the *conditional mark density*.

(3) *Normalizing condition.* If $\lambda(dt \times dx|\varpi)$ has a decomposition as above, it is obviously not unique. Hence, we always assume that $f(x|t, \varpi)$ is a probability density in the parameter x, i.e. we assume for all t and ϖ:

$$\int_{\mathbb{X}} f(x|t, \varpi)\mu_{\mathbb{X}}(dx) = 1. \qquad \diamondsuit$$

6.10 Definition (Intensity process). Let $\mathfrak{N} := \mathscr{N}_{\mathbb{Y}}(\mathbb{R})$ be a univariate canonical measurable space with driving process ϖ and let N be a point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$, for some mark space $\mathbb{X}$.

Assume λ is an intensity kernel with decomposition as in Equation (6.1). Then the following statements are equivalent:

(1) For all $t \in \mathbb{R}$ and $E \subseteq (t, \infty)$, $X \subseteq \mathbb{X}$ it holds that

$$\mathbb{E}\Big[\int_{E\times X} N(dt \times dx) \;\Big|\; \mathfrak{F}_t\Big] = \mathbb{E}\Big[\int_{E\times X} f(x|t, \varpi)\lambda(t|\varpi)\mu_{\mathbb{X}}(dx)dt \;\Big|\; \mathfrak{F}_t\Big].$$

(2) For all non-negative, predictable processes H on $\mathfrak{N}$ it holds that

$$\mathbb{E}\Big[\int_{\mathbb{R}\times\mathbb{X}} H(t,x|\varpi)N(dt\times dx)\Big] = \mathbb{E}\Big[\int_{\mathbb{R}\times\mathbb{X}} H(t,x|\varpi)f(x|t,\varpi)\lambda(t|\varpi)\mu_{\mathbb{X}}(dx)dt\Big]. \qquad \diamondsuit$$

The above definition of the intensity process can be found e.g. in Section 1 of [BM96]; see also Definition D7 in Section II.3 of [Bré81]. Actually, what we have defined above is the so-called *predictable* intensity process. We will only deal with predictable intensity processes, but this is not a restriction at all. It follows from Theorem T13 in Section II.4 of [Bré81] that whenever one has a general intensity process one can find a predictable intensity process. Moreover, predictable intensity processes are unique in some specific sense, see Theorem T12 in Section II.4 of [Bré81].

So far, we have only considered the univariate case. But again, the multivariate case is a trivial extension. Let us clarify the relation between univariate and multivariate intensity functions: Recall that a multivariate point process can be identified with a univariate one in the following sense: If one has a pair $y \equiv (j,x)$, for some $x \in \mathbb{X}_j$, one can consider it as an element of the combined mark space $y \in \mathbb{X}$.

6.11 Remark (Multivariate intensity function). Let $\mathbb{X} \equiv \{\mathbb{X}_1, \dots \mathbb{X}_d\}$ be a multivariate mark space.

Assume $\lambda_j(t,x|\varpi)$, for $1 \le j \le d$, is a family of univariate intensity functions as in Equation (6.1) with decomposition

$$\lambda_j(t,x|\varpi) = f_j(x|t,\varpi)\lambda_j(t|\varpi), \tag{6.2}$$

so that $f_j(x|t,\varpi)$ are probability densities on $\mathbb{X}_j$. We say $\lambda_{j=1\dots,d}$ is a multivariate *family of intensity functions* if the characterizing property in Definition 6.10 is satisfied for each component $j = 1, \dots, d$.

It is possible to interpret this family of intensity functions in two ways: Either as vector-valued intensity function on the multivariate mark space, or as a scalar-valued intensity function on the combined mark space $\mathbb{X}$.

To perform this transition, we need some notation:

(1) *Change from multivariate to univariate representation.* Assume we start

with a multivariate family of intensity functions $\lambda_j(t, x|\varpi)$. Then define:

$$\lambda_*(t|\varpi) := \sum_{j=1}^{d} \lambda_j(t|\varpi), \qquad f_*(j, x|t, \varpi) := p_*(j|t, \varpi) f_j(x|t, \varpi),$$
$$p_*(j|t, \varpi) := \frac{\lambda_j(t|\varpi)}{\lambda_*(t|\varpi)}, \qquad \lambda_*(t, j, x|\varpi) := f_*(j, x|t, \varpi)\lambda_*(t|\varpi).$$

As a consequence, we have the properties

$$\sum_{j=1}^{d} p_*(j|t, \varpi) = 1, \qquad \sum_{j=1}^{d} \int_{\mathbb{X}_j} f_*(j, x|t, \varpi)\mu_{\mathbb{X}_j}(dx) = 1.$$

It is then clear that $\lambda_*(t, j, x|\varpi) = \lambda_j(t, x|\varpi)$ and $\lambda_*(t, j, x|\varpi)$ is an intensity function on $\mathbb{R} \times \mathbb{X}$ with decomposition

$$\lambda_*(t, j, x|\varpi) = f_*(j, x|t, \varpi)\lambda_*(t|\varpi).$$

For the other direction:

(2) *Change from univariate to multivariate representation.* Let $\lambda_*(t, j, x|\varpi)$ be an intensity function on the space $\mathbb{R} \times \mathbb{X}$ with decomposition

$$\lambda_*(t, j, x|\varpi) = f_*(j, x|t, \varpi)\lambda_*(t|\varpi).$$

Then define for all $1 \le j \le d$ the functions

$$p_*(j|t, \varpi) := \int_{\mathbb{X}_j} f_*(j, x|t, \varpi)\mu_{\mathbb{X}_j}(dx), \qquad \lambda_j(t|\varpi) := p_*(j|t, \varpi)\lambda_*(t|\varpi),$$
$$f_j(x|t, \varpi) := \frac{f_*(j, x|t, \varpi)}{p_*(j|t, \varpi)}, \qquad \lambda_j(t, x|\varpi) := f_j(x|t, \varpi)\lambda_j(t|\varpi).$$

One can now easily check that $\lambda_j(t, x|\varpi) = \lambda_*(t, j, x|\varpi)$ and λ_j is again a multivariate family of intensity functions on the spaces $\mathbb{R} \times \mathbb{X}_j$. ♢

The above expressions have some nice interpretations: $\lambda_*(t|\varpi)$ is the infinitesimal probability that an event occurs in the time interval $(t, t + dt]$, $p_*(j|t, \varpi)$ is the probability that the event lies in component j, given that there is an event at time t, and $f_*(j, x|t, \varpi)$ is the probability density that the event is in component j at location x, given that there is an event at time t.

As a summary, we have shown that a multivariate family of intensity functions can be considered also as a univariate intensity function, but one has to take a higher dimensional mark space.

6.12 Definition (Classification of multivariate intensity functions). Fix a multivariate mark space $\mathbb{X}$ and let $\lambda_j(t, x|\varpi)$ be a multivariate family of intensity functions which can be decomposed according to Equation (6.2).

(1) λ is *stationary* or *translation-covariant* if it satisfies the two equivalent conditions:

$$\lambda_j(t, x|\varpi) = \lambda_j(t + h, x|\varpi + h) \quad \text{or} \quad \lambda_j(t + h, x|\varpi) = \lambda_j(t, x|\varpi - h),$$

for all $t, h \in \mathbb{R}$, $x \in \mathbb{X}_j$, $\varpi \in \mathfrak{N}$ and all $1 \leq j \leq d$. In this case, define:

$$f_j(x|\varpi) := f_j(x|h, \varpi + h) \qquad \text{and} \qquad \lambda_j(\varpi) := \lambda_j(h|\varpi + h),$$

for all $t \in \mathbb{R}$, where $h \in \mathbb{R}$ can be chosen arbitrarily. Then the general form of a stationary intensity function is:

$$\lambda_j(t, x|\varpi) = f_j(x|\varpi - t)\lambda_j(\varpi - t).$$

The next property describes whether the mark distribution is independent of the past of the process. We distinguish between a weaker and a stronger version:

(2) λ has *componentwise* unpredictable marks if the densities $f_j(x|t, \varpi)$ are independent of ϖ. In this case, the intensity function is of the general form:

$$\lambda_j(t, x|\varpi) = f_j(x|t)\lambda_j(t|\varpi).$$

For the next definition, recall the definition of p_* given in Remark 6.11:

(3) λ has *completely* unpredictable marks, if both, $f_j(x|t, \varpi)$ and $p_*(j|t, \varpi)$, are independent of ϖ. In this case, the general form of the intensity function is:

$$\lambda_j(t, x|\varpi) = f_j(x|t)p_*(j|t)\lambda_*(t|\varpi) = f_*(j, x|t)\lambda_*(t|\varpi).$$

Note that in a univariate setting the above two definitions collapse and one simply speaks about *unpredictable marks*. ♢

The terminology for unpredictable mark distributions has been taken from Definition 6.4.III in [DVJ03]. The definition given above is slightly more general, since it distinguishes between componentwise and completely unpredictable marks.

Often, the intensity function depends only on the past of the associated point process. In this case, it is more convenient to consider the so-called intrinsic intensity function. We treat only the univariate case, since the extension to the multivariate case is trivial.

6.13 Definition (Intrinsic intensity function). Let $\mathfrak{N} := \mathscr{N}_{\mathbb{X}\times\mathbb{R}_+}(\mathbb{R})$ be a univariate canonical measurable space and assume N a point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$. Further, let λ be the predictable intensity function of N and assume there exists an $\mathfrak{P}_{\mathbb{X}}(\mathbb{R})$-predictable function $\gamma(t|\nu)$, for $\nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{R})$, such that

$$\lambda(t|\varpi) = \gamma(t|N),$$

for all $t \in \mathbb{R}$. Then γ is called the *intrinsic intensity function* of N. $\diamondsuit$

As far as we know, the notion of an *intrinsic intensity function* is not used in the literature. But we think the idea behind it justifies this name.

For the next definition, the time character of the event space $\mathbb{E}$ is irrelevant. This means that $\mathbb{E} = \mathbb{R}^e$ can again be a general event space and does not need to be $\mathbb{R}$.

We define the Janossy measure only in the case of a point process with values in $\mathscr{N}(\mathbb{E})$. The general case is then very similar.

6.14 Definition (Local Janossy measure). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and N a point process with values in $\mathscr{N}(\mathbb{E})$.

(1) *Local Janossy measure.* For an integer $n \geq 1$ and a bounded set $D \subseteq \mathbb{E}$, the *local Janossy measure* on D of order n is defined as:

$$J_D^n(A) := \mathbb{E}\Big[N^{[n]}(A), N(D) = n\Big], \text{ for sets } A \subseteq D^n.$$

Here $N^{[n]}$ denotes the factorial product measure of order n, see also Definition 2.36. By convention, the Janossy measure of order 0 is defined as $J_D^0 := \mathbb{P}\big[N(D) = 0\big]$.

(2) *Local Janossy density.* Assume that all local Janossy measures J_D^n are absolutely continuous with respect to the Lebesgue measure on $\mathbb{E}^n$. The

densities are then denoted by $j_D^n(x_1, \ldots, x_n)$ and are called the *local Janossy densities*. That is

$$J_D^n(dx_1 \times \ldots \times dx_n) = j_D^n(x_1, \ldots, x_n)\, dx_1 \cdot \ldots \cdot dx_n. \qquad \diamondsuit$$

Note that there is also a *global* Janossy measure. But for our purposes, we will only need the definition of the local version.

The following notion of exclusion probabilities and the relation to Janossy measures can be found in Section 5.3, specifically Equation (5.3.11), in [DVJ03]: Recall the definition of the falling factorial $r^{[n]}$, given in Definition 3.8.

6.15 Definition (Exclusion probabilities). Let $D \subseteq \mathbb{E}$ be a bounded set and $\{A_1, \ldots, A_m\}$ a partition of D. Moreover, let n be an integer with decomposition $n = \sum_{k=1}^m n_k$, for $n_k \geq 0$. Then:

$$\begin{aligned} J_D^n\big(A_1^{n_1} \times \ldots \times A_m^{n_m}\big) &= \mathbb{E}\Big[N(A_1)^{[n_1]} \cdot \ldots \cdot N(A_m)^{[n_m]}, N(D) = n\Big] \\ &= n_1! \cdot \ldots \cdot n_m! \cdot \mathbb{P}\Big[N(A_1) = n_1, \ldots, N(A_m) = n_m\Big]. \end{aligned} \qquad \diamondsuit$$

The following expression for the likelihood function can be found in Definition 7.1.II and Proposition 7.3.III in [DVJ03]:

6.16 Definition (Likelihood function). Fix a multivariate mark space $\mathbb{X}$ and consider the associated point configuration space $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$. Moreover, let $D \subseteq \mathbb{R}$ be a bounded set. Let $\{\gamma_j; 1 \leq j \leq d\}$ be a multivariate family of intrinsic intensity functions. The *likelihood function* L_D relative to D is

$$\begin{aligned} \log L_D(\nu) = &\sum_{j=1}^d \int_{D \times \mathbb{X}_j} \log \gamma_j(t, x|\nu) \nu^{(j)}(dt \times dx) \\ &- \sum_{j=1}^d \int_{D \times \mathbb{X}_j} \gamma_j(t, x|\nu) \mu_{\mathbb{X}_j}(dx) dt, \end{aligned} \tag{6.3}$$

where $\nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{R})$. $\diamondsuit$

Note that the likelihood function is a local property and depends on the set D, in the same way as the local Janossy density does.

It turns out that the likelihood function and the Janossy densities are very closely related. Again for simplicity, we only consider the case where $\nu \in \mathscr{N}(D)$ is a univariate point configuration without marks:

6.17 Remark (Connection with Janossy densities). Assume $\nu(D) =: n$ and enumerate these n events with the vector $\{t_1, \ldots, t_n\}$. Then

$$L_D(\nu) = j_D^n(t_i, \ldots, t_n). \qquad \diamondsuit$$

Note that there is an inherent problem if one wants to calculate the likelihood in an actual application. Assume we can observe the point process only during the finite time interval D. Then the likelihood $L_D(\nu)$ is in general *not* the same as the likelihood $L_D(\nu_D)$ of the restricted point configuration ν_D. Let us explain this in more detail:

6.18 Remark (Observed & realized). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and N a point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$. Obviously, the following distinction has to be made:

(1) *Realized configuration.* If $\omega \in \Omega$ is the realized state of the world, the *realized* point configuration is $\nu := N(\omega)$.

(2) *Observed configuration.* Because we are only able to observe the time interval D, the *observed* configuration is the truncation $\nu_D := N_{D\times\mathbb{X}}(\omega)$. Note that ν_D is an element of $\mathscr{N}_{\mathbb{X}}(D)$.

Clearly, one has to expect that $\nu \neq \nu_D$, so in general

$$L_D(\nu) \neq L_D(\nu_D).$$

Hence, the fact that there is an index D in L_D does not mean events outside of the observation period D are not relevant. $\diamondsuit$

Next we introduce the intrinsic Hawkes intensity function. It consists of two components, the *immigration intensity* and the *transfer-functions*, which specify the self- and mutual-excitation behavior of the Hawkes process:

6.19 Definition (Hawkes intrinsic intensity function). Let $\mathbb{X}$ be a multivariate mark space. A Hawkes process, or better its associated intensity function, is specified by the following three elements, where the indexes run over $j, k \in \{1, \ldots, d\}$:

(1) *Immigration intensity.* A family η_j of locally-finite, non-negative functions, defined on $\mathbb{R}$. They are called the *immigration intensities* and specify the intensity at which new events arrive over time.

(2) *Transfer function.* A family $h_{jk}(t, x)$ of continuous in t, non-negative functions, defined for $t > 0$ and $x \in \mathbb{X}_k$. This family is called the family of *transfer functions*.

(3) *Mark distribution.* A family f_j of probability densities on the mark spaces $\mathbb{X}_j$. They are called the *mark distributions.*

Now let $\nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{R})$ be a point configuration:

(1) The *Hawkes excitation functions* β_j and *Hawkes intensity functions* α_j are defined as

$$\begin{aligned} \beta_j(t|\nu) &:= \sum_{k=1}^{d} \int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)\nu^{(k)}(ds \times dx), \\ \alpha_j(t|\nu) &:= \eta_j(t) + \beta_j(t|\nu), \end{aligned} \tag{6.4}$$

for all $1 \leq j \leq d$ and $t \in \mathbb{R}$. These two parts represent the immigration part and a the self-excitation part of the Hawkes process.

Note that $\alpha_j(t|\nu)$ are only the time-intensity functions of a Hawkes process, see also Definition 6.9. But in order to specify the full dynamics of a Hawkes process, one needs to know the time-space-intensity functions $\alpha_j(t,x|\nu)$:

(2) The *time-space* Hawkes intensity functions and intensity measures are defined as

$$\alpha_j(t,x|\nu) = f_j(x)\alpha_j(t|\nu), \quad \alpha_j(dt \times dx|\nu) = f_j(x)\alpha_j(t|\nu)\mu_{\mathbb{X}_j}(dx)dt. \ \diamondsuit$$

Compare the definition of the time-space Hawkes intensity functions with the general formulation of an intensity measure given in Equation (6.1).

In Definition 6.33, we give the exact definition of a Hawkes process, what we then call a strong solution. But if we assume for a moment that N is a univariate Hawkes process, then the intensity process is given by $\lambda(t) := \alpha(t|N)$. This means that the intensity process does not depend directly on the underlying driving process ϖ, but only on the history of the Hawkes process N itself. According to Definition 6.13, the Hawkes intensity function is therefore an intrinsic intensity.

Let us now check on some other properties of the Hawkes intensity function:

6.20 Remark (Properties of the Hawkes Intensity). According to the classification scheme in Definition 6.12, one observes that:

(1) *Stationary intensity function.* The Hawkes excitation functions β_j are stationary, i.e. for all $t, h \in \mathbb{R}$ and $\nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{R})$ the following equality is

satisfied:

$$\beta_j(t+h|\nu) = \beta_j(t|\nu - h).$$

One can easily check this:

$$\begin{aligned}\beta_j(t+h|\nu) &= \sum_{k=1}^{d} \int_{(-\infty,t+h)\times\mathbb{X}_k} h_{jk}(t+h-s,x)\nu^{(k)}(ds\times dx)\\ &= \sum_{k=1}^{d} \int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)\nu^{(k)}(ds+h\times dx)\\ &= \sum_{k=1}^{d} \int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)[\nu^{(k)}-h](ds\times dx) = \beta_j(t|\nu-h).\end{aligned}$$

As a consequence, the self-excitation part given by the family β_j is homogeneous in time. This shows that the functions $\beta_j(\nu) := \beta_j(h|\nu + h)$, for an arbitrary $h \in \mathbb{R}$, contain enough information to reconstruct the functions $\beta_j(t|\nu)$.

Note that in general only the Hawkes excitation functions β_j are stationary. If additionally, the immigration intensities η_j are constants, then also the intensity functions α_j are stationary. In this case, the Hawkes process is a stationary process, if one can show its existence.

(2) *Component-wise unpredictable marks.* In the case of a Hawkes process, the conditional mark distributions are of the form $f_j(x)$, i.e. they do not dependent on the parameters t and ν. If one considers Definition 6.12, this shows that a Hawkes process has time-homogeneous, component-wise unpredictable conditional mark distributions. ♢

In the above definition, we give an explicit expression for the Hawkes excitation functions β_j. But one could specify β_j also in a more abstract way, as follows: For simplicity, only the univariate case is considered.

6.21 Remark (Alternative characterization). Assume $\beta(t|\nu)$, for $t \in \mathbb{R}$ and $\nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{R})$, is a non-negative function with the following three properties:

(1) It is *stationary*, i.e. $\beta(t|\nu) = \beta(0|\nu - t) =: \beta(\nu - t)$.

(2) It is *linear* in ν, i.e. $\beta(a\nu + b\kappa) = a\beta(\nu) + b\beta(\kappa)$,

(3) It is *continuous* in ν, i.e. if $\nu_n \to \nu$ converges weakly in $\mathscr{N}(\mathbb{R})$, then $\lim_{n\to\infty} \beta(\nu_n) = \beta(\nu)$.

One can now easily check that all functions β which satisfy these three properties must be of the form of a Hawkes excitation function. $\diamondsuit$

For many calculations related to theoretical properties of Hawkes processes, one does not need to know the transfer functions $h_{jk}(t,x)$, but it is enough if one knows what we call the *averaged transfer functions* $H_{jk}(t)$. The reason for this is that Hawkes processes have component-wise unpredictable marks, and this allows us to integrate out the space parameter $x \in \mathbb{X}_j$ in many calculations.

6.22 Definition (Averaged transfer functions). Let $\mathbb{X}$ be a multivariate mark space with base measures $\mu_{\mathbb{X}_j}$ and $\boldsymbol{\beta}$ a multivariate family of Hawkes excitation functions. For all $t > 0$ and $1 \leq j, k \leq d$ define the functions

$$H_{jk}(t) := \int_{\mathbb{X}_k} h_{jk}(t,x) f_k(x) \mu_{\mathbb{X}_k}(dx). \tag{6.5}$$

$\diamondsuit$

Note that $H_{jk}(t)$ is an expected value in the following sense: If X is an $\mathbb{X}_k$-valued random variable with distribution $f_k(x)\mu_{\mathbb{X}_k}(dx)$ then

$$H_{jk}(t) = \mathbb{E}\big[h_{jk}(t,X)\big].$$

The averaged transfer functions H_{jk} often occur in vector- and matrix-valued convolutions. It is therefore convenient to introduce a matrix $\mathbb{H}$, that comprises all averaged transfer functions.

6.23 Definition (Transfer function matrix). Let $\boldsymbol{\beta}$ be a vector-valued Hawkes excitation function. Define for $t > 0$ the matrix-valued function $\mathbb{H}$ with values in $\mathbb{R}_+^{d \times d}$ by:

$$\mathbb{H}(t) := \Big\{ H_{jk}(t), \text{ for } 1 \leq j, k \leq d \Big\}. \qquad \diamondsuit$$

The global properties of a Hawkes process depend crucially on the branching behavior, which is related to the total mass of the transfer functions h_{jk}, and as a consequence, to the total mass of the averaged transfer functions H_{jk}. To capture the branching behavior of a Hawkes excitement function $\boldsymbol{\beta}$, one has to look at the so-called branching matrix Q:

6.24 Definition (Branching matrix). Let $\boldsymbol{\beta}$ be a multivariate family of Hawkes excitation functions. Define the $d \times d$-dimensional matrix Q with non-

negative components

$$Q_{jk} := \int_{\mathbb{R}_+} H_{jk}(t)dt = \int_{\mathbb{R}_+ \times \mathbb{X}_k} h_{jk}(t,x) f_k(x) \mu_{\mathbb{X}_k}(dx)dt \tag{6.6}$$

for $1 \leq j, k \leq d$. ♢

Separation of Transfer Function. Often, the transfer functions h_{jk} are a product of two functions that depend only on t and x, respectively. In this case, there exists a very convenient representation of h_{jk} that allows one to read off the branching coefficients Q_{jk} directly.

The assumption, that the Hawkes intensity function is of this form is not uncommon. The same decomposition is used e.g. in Example 7.6(c) and Example 7.3(b) in [DVJ03].

6.25 Definition (Separation of transfer function). Let h_{jk} be a family of transfer functions and f_j be a family of densities with respect to the base measures $\mu_{\mathbb{X}_j}$ on the mark spaces $\mathbb{X}_j$.

Assume there are coefficients ϑ_{jk} and functions w_{jk} and g_{jk} such that the transfer functions h_{jk} are of the form

$$h_{jk}(t,x) = \vartheta_{jk} w_{jk}(t) g_{jk}(x).$$

In order to determine the three functions above uniquely, we need to impose two additional conditions:

$$\int_{\mathbb{R}_+} w_{jk}(t)dt = 1, \qquad \int_{\mathbb{X}_k} g_{jk}(x) f_k(x) \mu_{\mathbb{X}_k}(dx) = 1.$$

Then the components of the branching matrix are given by $Q_{jk} = \vartheta_{jk}$. Moreover, the averaged transfer functions are $H_{jk} = w_{jk}$. ♢

The three components of this decomposition have the interpretation:

(1) The *decay functions* w_{jk} control how fast the effect of an event decays in time.

(2) The *impact functions* g_{jk} control how strong the effect is, given that the event has the mark $x \in \mathbb{X}_k$.

Note that both, the decay and the impact functions, only specify the relative impact of an event. This is due to the normalizing conditions given above. The absolute impact of an event is governed by the branching matrix.

(3) The *branching coefficients* ϑ_{jk} control in absolute terms, how much a given event is going to increase the intensity process.

Note that the normalizing conditions are not chosen arbitrarily. Only under these conditions one can interpret the decay functions, impact functions and the branching coefficients as explained above.

Likelihood Function. In order to express the likelihood function in a more compact form, we first define what we call the cumulative transfer functions $\bar{h}_{jk}$:

6.26 Definition (Cumulative transfer functions). Let $\boldsymbol{\alpha}$ be a multivariate family of Hawkes intensity functions as given in Equation (6.4). For $1 \leq j, k \leq d$ and $x \in \mathbb{X}_k$ define the *cumulative transfer functions* by

$$\bar{h}_{jk}(t,x) := \int_0^t h_{jk}(s,x)ds, \text{ if } t > 0, \text{ and } \bar{h}_{jk}(t,x) := 0, \text{ if } t \leq 0. \tag{6.7}$$

♢

Note that the cumulative transfer functions $\bar{h}_{jk}$ should not be confused with the averaged transfer functions H_{jk} defined earlier.

6.27 Proposition (Hawkes likelihood function). Let $\boldsymbol{\alpha}$ be a multivariate family of Hawkes intensity functions and let the cumulative transfer functions $\bar{h}_{jk}$ be as in Equation (6.7).

Assume $T_* < T^*$ and let the observation domain be $D := [T_*, T^*]$. For all realization $\nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{R})$ one has

$$\begin{aligned}\log L_{[T_*,T^*]}(\nu) &= \sum_{j=1}^d \int_{[T_*,T^*]} \log \alpha_j(s|\nu)\nu^{(j)}(ds \times \mathbb{X}_j) \\ &+ \sum_{j=1}^d \int_{\mathbb{X}_j} \log f_j(x)\nu^{(j)}([T_*,T^*] \times dx) - \sum_{j=1}^d \int_{T_*}^{T^*} \eta_j(s)ds \\ &- \sum_{j,k=1}^d \int_{(-\infty,T^*)\times\mathbb{X}_k} \Big[\bar{h}_{jk}(T^*-s,x) - \bar{h}_{jk}(T_*-s,x)\Big]\nu^{(k)}(ds \times dx). \quad \diamondsuit\end{aligned}$$

Recall from Remark 6.18 the distinction between the realized and the observed point configuration. If one can only observe the truncated point configuration ν_D, one could e.g. calculate the value $\log L_D(\nu_D)$ and use this as an approximation to the actual logarithmic likelihood $\log L_D(\nu)$. But this is more a problem

of parameter estimation than a problem concerning the likelihood function itself.

6.3 Strong Solutions

Next, we construct a Hawkes process as a thinning of the underlying driving process ϖ. But first we need to deal with the canonical probability measure $\mathbb{P}$, which has not yet been defined in Definition 6.3. We will now add this missing piece. We first define the canonical probability space in the univariate case and then we extend the definition to the multivariate case:

6.28 Definition (Univariate canonical space). Assume $\mathbb{X}$ is a univariate mark space and $\mathscr{T} \subseteq \mathbb{R}$ a time interval, possibly unbounded. Furthermore, let $\mu_{\mathbb{X}}$ be the base measure on the mark space $\mathbb{X}$ and f be a probability density on $\mathbb{X}$ with respect to this measure.

(1) *Canonical space and driving process.* Consider the extended mark space $\mathbb{Y} := \mathbb{X} \times \mathbb{R}_+$. The associated canonical measurable space and its σ-algebra are

$$\mathfrak{N} := \mathscr{N}_{\mathbb{X}\times\mathbb{R}_+}(\mathscr{T}) \qquad \text{and} \qquad \mathfrak{F} := \mathfrak{F}_{\mathbb{X}\times\mathbb{R}_+}(\mathscr{T}).$$

Recall that the *driving process* ϖ is the identity function on the space $\mathfrak{N}$.

(2) *Distribution of the driving process.* Let $\mathbb{P}$ be the probability measure on $\mathfrak{N}$, so that ϖ is a compound Poisson process with the following characteristics:

(a) The process $\varpi(dt \times \mathbb{X} \times dz)$ is a Poisson process on $\mathbb{R}_+ \times \mathscr{T}$ with mean measure $dt \times dz$.

(b) The marks $\varpi(t \times dx \times z)$ have distribution $f(x)\mu_{\mathbb{X}}(dx)$ and are independent of the ground process $\varpi(dt \times \mathbb{X} \times dz)$ and independent of each other.

Note that $\mathbb{P}$ is uniquely determined by these two properties. ♢

We give an equivalent definition of the canonical probability measure $\mathbb{P}$. Instead of defining the driving process ϖ as a compound Poisson process we could also proceed as follows:

6.29 Remark (Equivalent Definition). The distribution $\mathbb{P}$ of the driving process is so that ϖ is a Poisson process on the space $\mathscr{T} \times \mathbb{X} \times \mathbb{R}_+$ with mean

measure

$$\mu_{\mathscr{T}\times\mathbb{R}_+\times\mathbb{X}}(dt \times dx \times dz) := f(x)\mu_{\mathbb{X}}(dx)dt\,dz. \qquad \diamondsuit$$

The extension to the multivariate case is now straightforward:

6.30 Definition (Multivariate canonical space). Let $\mathscr{T} \subseteq \mathbb{R}$ be a time interval, $\mathbb{X}$ be a multivariate mark space, $\mu_{\mathbb{X}_j}$ the base measures on $\mathbb{X}_j$ and f_j probability densities on $\mathbb{X}_j$.

(1) *Canonical space and driving process.* The multivariate canonical probability space is $\mathfrak{N} := \mathscr{N}_{\mathbb{X}\times\mathbb{R}_+}(\mathscr{T})$. Consequently, the driving process ϖ has components $\varpi^{(j)}$ with values in $\mathfrak{N}_j := \mathscr{N}_{\mathbb{X}_j\times\mathbb{R}_+}(\mathscr{T})$.

(2) *Distribution of the driving process.* The canonical probability measure $\mathbb{P}$ is so that the components $\varpi^{(j)}$ are independent and have the same distribution as in the univariate case. $\diamondsuit$

In the multivariate case, there is a weaker and a stronger version of the notion of a *simple* process. We call the weaker version *component-wise* simpleness and the stronger version *overall* simpleness. It is important to note that our driving process satisfies the stronger version:

6.31 Remark (Simpleness of Driving Process). Take a realization $\varpi^{(j)}$ of the j-th component of the driving process.

(1) The driving process is *component-wise simple* because for each time point $t \in \mathbb{R}$, the projection $\varpi^{(j)}(t \times dx \times dz)$ restricted to the set $\{t\} \times \mathbb{X}_j \times \mathbb{R}_+$ has at most one event in the space $\mathbb{X}_j \times \mathbb{R}_+$. This is clear by the definition of a compound Poisson process.

(2) The driving process is also *overall simple*: If one fixes some $t \in \mathscr{T}$ then even the projection $\sum_{j=1}^{d} \varpi^{(j)}(t \times dx \times dz)$ consists of at most one event in the space $\{t\} \times \mathbb{X} \times \mathbb{R}_+$. This means that for a fixed time point t there is at most one component that contains an event. $\diamondsuit$

These two properties of the driving process ϖ are a consequence of the fact that a Poisson process in the plane $\mathbb{R}^2$ never has multiple points with the same first or second coordinate. Hence, if (x_1, y_1) and (x_2, y_2) are two points of a Poisson process in $\mathbb{R}^2$, then $x_1 \neq x_2$ and $y_1 \neq y_2$, with probability one.

For the following definition, we restrict ourselves to the univariate case, as the multivariate case is a trivial extension:

6.32 Definition (Thinning). Let $\mathbb{X}$ be a univariate mark space with mark density f and $\mathscr{T} \subseteq \mathbb{R}$ the time domain. Moreover, let $\mathfrak{N} := \mathscr{N}_{\mathbb{X}\times\mathbb{R}_+}(\mathscr{T})$ be the associated univariate canonical probability space.

(1) *Thinning with respect to intensity.* Let λ be a non-negative, predictable stochastic process. The *thinning* Θ of ϖ with respect to λ is

$$\Theta_\varpi[\lambda](E \times X) := \int_{E\times\mathbb{R}_+} \mathbb{1}_{\{z\leq\lambda(t|\varpi)\}}\varpi(dt \times X \times dz), \tag{6.8}$$

where $E \subseteq \mathscr{T}$, $X \subseteq \mathbb{X}$ are two sets and $\varpi \in \mathfrak{N}$ is the driving process.

Hence, the thinning $\Theta_\varpi[\lambda]$ is a point process with values in $\mathscr{N}_{\mathbb{X}}(\mathscr{T})$. Now recall the definition of an intrinsic intensity function, given in Definition 6.13.

(2) *Thinning with respect to intrinsic intensity.* Assume the intensity process λ is defined in terms of an intrinsic intensity function γ, associated with some point process N. In this case, the thinning is of the form

$$\Theta_\varpi\big[\gamma(\cdot|N)\big](E \times X) := \int_{E\times\mathbb{R}_+} \mathbb{1}_{\{z\leq\gamma(t|N)\}}\varpi(dt \times X \times dz). \qquad \diamondsuit$$

The notion of a strong solution in the context of Hawkes processes appears first in [Mas98], see Equation (1). Instead of the Hawkes intrinsic intensity function, one could obviously take any other intrinsic intensity function as well.

We give the definition of a strong solution only in the univariate case. The definition in the multivariate case is a trivial extension:

6.33 Definition (Strong solution). Let $\mathfrak{N}$ be a univariate canonical space and α a Hawkes intensity function.

(1) *Formulation with thinning operator.* Let N be a predictable point process on the canonical space with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$. Then N is called a *strong solution* if it satisfies, for all $E \subseteq \mathbb{R}$ and $X \subseteq \mathbb{X}$:

$$\Theta_\varpi\big[\alpha(\cdot|N)\big](E \times X) = N(E \times X).$$

Instead of using the abstract thinning operator Θ, one can write the condition also in explicit form:

(2) *Formulation without thinning operator.* A point process N is a strong

solution if and only if, for all sets $E \subseteq \mathbb{R}$ and $X \subseteq \mathbb{X}$:

$$N(E \times X) = \int_{E \times \mathbb{R}_+} \mathbb{1}_{\{z \leq \alpha(t|N)\}} \varpi(dt \times X \times dz). \tag{6.9}$$

♢

For the next two definitions of initial states and continuations we restrict the time domain: Let $\chi \in \mathbb{R}$ be a fixed time and consider the time domain $\mathscr{T} := (\chi, \infty)$. The canonical probability space is then $\mathfrak{N}_+ := \mathscr{N}_{\mathbb{X} \times \mathbb{R}_+}\big((\chi, \infty)\big)$. To emphasize that the time domain is only a half-line, we use the symbol $\mathfrak{N}_+$ instead of $\mathfrak{N}$.

Before we can define a continuation, we need first something that can be continued, which is the so-called initial state. Again, we only consider the univariate case:

6.34 Definition (Initial state). Let $\chi \in \mathbb{R}$ be a fixed time and $\mathfrak{N}_+$ the associated canonical space on the (χ, ∞). Furthermore, let α be a Hawkes intensity function.

(1) An *initial state* is a point configuration $\nu_- \in \mathscr{N}_{\mathbb{X}}\big((-\infty, \chi]\big)$. One should interpret ν_- as the past of a point process N, that has already been observed.

A special case of an initial state is the *empty* or *void* initial state:

(2) We use the following two symbols for *empty point configurations*, i.e. point configurations without any events:

$$\varnothing_- \in \mathscr{N}_{\mathbb{X}}\big((-\infty, \chi]\big) \qquad \text{and} \qquad \varnothing \in \mathscr{N}_{\mathbb{X}}(\mathbb{R}). \qquad \diamondsuit$$

The notion of an initial state is also used in the papers [BM96], [Mas98], [BNT02] and [Tor02]. Our definition deviates slightly, as an initial state is assumed to be deterministic, although we will see in the proof of Theorem 6.55, that this is not really a restriction.

A continuation is essentially a point process whose past up to time χ coincides with ν_- and its future after time χ follows the dynamics determined by the Hawkes intensity function α. The precise definition is:

6.35 Definition (Continuation). Let N_+ be a point process on $\mathfrak{N}_+$ with values in $\mathscr{N}_{\mathbb{X}}\big((\chi, \infty)\big)$ and $\hat{N} := \nu_- + N_+$ be the superposition of ν_- and N_+.

(1) *Formulation with thinning operator.* If the following thinning relation is satisfied for all sets $E_+ \subseteq (\chi, \infty)$ and $X \subseteq \mathbb{X}$, then N_+ is called a

continuation of ν_-:

$$\hat{N}(E_+ \times X) = \Theta_{\varpi_+}\big[\alpha(\cdot|\hat{N})\big](E_+ \times X).$$

One can formulate this condition also in a less abstract form:

(2) *Formulation without thinning operator.* A point process N_+ on $\mathfrak{N}_+$ is a continuation of ν_- if and only if

$$\hat{N}(E_+ \times X) = \int_{E_+ \times \mathbb{R}_+} \mathbb{1}_{\{z \le \alpha(t|\hat{N})\}} \varpi_+(dt \times X \times dz),$$

for all sets $E_+ \subseteq (\chi, \infty)$ and $X \subseteq \mathbb{X}$. ♢

Below we will discuss the relationship between continuations with different initial states and different immigration intensities. In order to avoid confusion, we introduce a corresponding notation:

6.36 Notation (Continuation triple). Let ν_- be an initial state and η an immigration intensity. If N_+ is a continuation of ν_- with respect to the intensity process $\alpha(t|\hat{N}) := \eta(t) + \beta(t|\hat{N})$, we say that

$$(\nu_-, \eta, N_+)$$

is a *continuation triple*, where $\hat{N} := \nu_- + N_+$. Note that we did not include β in this triple, since we assume it is fixed. ♢

If we compare the definition of a continuation with the one for a strong solution, we see many similarities: Basically, a continuation satisfies the same condition as a strong solution, but only on the time interval (χ, ∞). We will work out this connection in more detail later.

The following important theorem states that the thinning procedure creates a point process with the expected intensity:

6.37 Theorem (Thinning). Let $\mathbb{X}$ be a univariate mark space with mark space density f and $\mathscr{T} \subseteq \mathbb{R}$ be a time interval. We consider the univariate canonical probability space $\mathfrak{N} := \mathscr{N}_{\mathbb{X} \times \mathbb{R}_+}(\mathscr{T})$.

Assume λ is a non-negative, locally-integrable, predictable process. Then the thinning $N = \Theta_\varpi[\lambda]$ of ϖ with respect to λ is a locally-finite point process, adapted to the canonical filtration. Moreover, the intensity process and intensity measure of N are given by

$$\lambda(t, x|\varpi) = f(x)\lambda(t|\varpi), \qquad \Lambda(dt \times dx|\varpi) = f(x)\lambda(t|\varpi)\mu_{\mathbb{X}}(dx)dt. \qquad ♢$$

PROOF: See Lemma 3 in [BM96]. The theorem goes back to [Ker64] and [LS79]. □

6.4 Hazard Rate

6.38 Definition (Hazard rate). Let $\mathbb{X}$ be a multivariate mark space and $\mathfrak{N} := \mathscr{N}_{\mathbb{X}\times\mathbb{R}_+}(\mathbb{R})$ the associated multivariate canonical probability space with driving process ϖ.

Assume the following is given: Let N be an adapted point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$ and let $\lambda_j(t, x|\varpi)$ be the predictable time-space-intensity functions of N. Additionally assume that the marks are component-wise unpredictable, i.e. the intensity functions are of the form

$$\lambda_j(t, x|\varpi) = f_j(x)\lambda_j(t|\varpi).$$

Then we define:

(1) *Exact hazard rate.* Let $\boldsymbol{\lambda}^+(s, t|\varpi)$ be a vector-valued function defined for all $t > s$ and $\varpi \in \mathfrak{N}$ and assume that for all $s \in \mathbb{R}$ the function

$$(\varpi, t) \mapsto \boldsymbol{\lambda}^+(s, t|\varpi) \text{ is } \mathfrak{F}_s \otimes \mathscr{B}(s, \infty)\text{-measurable.}$$

We call $\boldsymbol{\lambda}^+$ the *hazard rate* of N, if for all $t > s$:

$$\boldsymbol{\lambda}(t|\varpi) = \boldsymbol{\lambda}^+(s, t|\varpi) \text{ on the set } \big\{\varpi : N(s, t)[\varpi] = 0\big\}. \tag{6.10}$$

Note that the hazard rate, if it exists, does not need to be unique.

(2) *Upper bound for hazard rate.* Assume $\tilde{\boldsymbol{\lambda}}^+(s, t|\varpi)$ is a function of the same type and satisfies the same measurability condition as the hazard rate $\boldsymbol{\lambda}^+$. We call $\tilde{\boldsymbol{\lambda}}^+$ an *upper bound* for the hazard rate if for all $t > s$:

$$\boldsymbol{\lambda}(t|\varpi) \le \tilde{\boldsymbol{\lambda}}^+(s, t|\varpi) \text{ on the set } \big\{\varpi : N(s, t)[\varpi] = 0\big\}. \qquad \diamondsuit$$

Note that we need the upper bound $\tilde{\boldsymbol{\lambda}}^+$ only in cases where we cannot find the exact hazard rate $\boldsymbol{\lambda}^+$. The upper bound then still allows us to apply the proposition below:

The next statement is based on Lemma 1 in [BM96], but has been reformulated and extended: Firstly, a separate concept of a hazard rate function is introduced, whereas in the original statement, the intrinsic intensity function is

used, although it is not called like this. Secondly, the statement is generalized to the case where only an upper bound is available. This will be helpful in the proof of Proposition 6.43, because there the hazard rate of the difference process L is not known.

6.39 Proposition (Extinction probability). Consider the multivariate canonical space $\mathfrak{N} := \mathscr{N}_{\mathbb{X}\times\mathbb{R}_+}(\mathbb{R})$ and let N be a point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$.

(1) *Exact hazard rate.* If $\boldsymbol{\lambda}^+$ is the hazard rate of N, then for all $t > s$:

$$\mathbb{P}\Big[N\big((s,t)\times\mathbb{X}\big) = 0 \;\Big|\; \mathfrak{F}_s\Big] = \exp\Big\{-\sum_{j=1}^{d}\int_s^t \lambda_j^+(s,u|\varpi)du\Big\}.$$

(2) *Upper bound for hazard rate.* If $\tilde{\boldsymbol{\lambda}}^+$ is an upper bound for the hazard rate, then for all $t > s$:

$$\mathbb{P}\Big[N\big((s,t)\times\mathbb{X}\big) = 0 \;\Big|\; \mathfrak{F}_s\Big] \geq \exp\Big\{-\sum_{j=1}^{d}\int_s^t \tilde{\lambda}_j^+(s,u|\varpi)du\Big\}.$$

Note that in both cases, the terms on the right-hand side are $\mathfrak{F}_s$-measurable.$\diamondsuit$

There is a common situation where we can easily find a hazard rate for some given point process. The idea is based on the so-called intrinsic hazard rate. The difference between the general hazard rate and the intrinsic hazard rate is the same as between the intensity function and the intrinsic intensity function, see Definition 6.13. We only consider the univariate case:

6.40 Corollary (Intrinsic hazard rate). Consider the univariate canonical space $\mathfrak{N} := \mathscr{N}_{\mathbb{X}\times\mathbb{R}_+}(\mathbb{R})$ and assume N is a point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$.

(1) *Exact hazard rate.* Assume N has the intrinsic intensity function γ, i.e. the intensity of N satisfies $\lambda(t|\varpi) = \gamma(t|N)$. Then the hazard rate of N is

$$\lambda^+(s,t|\varpi) := \gamma\big(t\big|N|_{(-\infty,s]}\big).$$

Assume one cannot find the intrinsic intensity function γ of N, but one has at least an upper bound $\tilde{\gamma}$ for the intrinsic intensity function of N:

(2) *Upper bound for hazard rate.* Assume there is an $\mathfrak{P}_{\mathbb{X}}(\mathbb{R})$-predictable function $\tilde{\gamma}$ such that the intensity of N is bounded from above by $\lambda(t|\varpi) \leq$

$\tilde{\gamma}(t|N)$. Then an upper bound for the hazard rate is given by

$$\tilde{\lambda}^+(s,t|\varpi) := \tilde{\gamma}\big(t\big|N|_{(-\infty,s]}\big). \qquad \diamondsuit$$

Next we give the obvious, but important notion of a difference process:

6.41 Definition (Difference process). Let K and N be two point processes with values in $\mathscr{N}(\mathbb{R})$, i.e. two univariate, point processes without marks.

The *difference process* between K and N is denoted by $L := |K - N|$. A more formal definition of L is

$$L(E) = \int_E \big[1 - N(t)\big] K(dt) + \int_E \big[1 - K(t)\big] N(dt),$$

for sets $E \subseteq \mathbb{R}$. $\diamondsuit$

Thus, the difference process contains all events of the difference set between the events of two point processes. Note that L is again a simple point process.

Since we will construct couplings on the canonical space $\mathfrak{N}_+$, which is defined on the half-line (χ, ∞), it will be convenient to consider not only the coupling time but also the truncated coupling time:

6.42 Definition (Coupling time). Let K and N be two point processes with values in $\mathscr{N}(\mathbb{R})$.

(1) *Truncated coupling time.* Let $\chi \in \mathbb{R}$ be a fixed time. The *truncated coupling time* T_χ between K and N on the time interval $[\chi, \infty)$ is

$$T_\chi := \inf_{t \geq \chi} \Big\{ K|_{(t,\infty)} = N|_{(t,\infty)} \Big\} = \inf_{t \geq \chi} \Big\{ L|_{(t,\infty)} = 0 \Big\}. \tag{6.11}$$

By convention, one sets $T_\chi := \chi$ in case the sets on the right-hand side are empty.

(2) *Coupling time.* To define the *coupling time* T between K and N, one takes the infimum over the whole real line instead, i.e. one defines

$$T := T_{-\infty}. \qquad \diamondsuit$$

Note that T_χ and T are random times, but in general not stopping times.

The following statement is motivated by a similar result of Proposition 2.1 in [Tor02]:

6.43 Proposition (Coupling). Consider a multivariate mark space $\mathbb{X}$ with associated configuration space $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$. Let K and N be two strong solutions for the two Hawkes intensity functions

$$\boldsymbol{\alpha}_K(t|\nu) := \boldsymbol{\eta}_K(t) + \boldsymbol{\beta}(t|\nu), \qquad \boldsymbol{\alpha}_N(t|\nu) := \boldsymbol{\eta}_N(t) + \boldsymbol{\beta}(t|\nu).$$

Note that the immigration intensities $\boldsymbol{\eta}_K$ and $\boldsymbol{\eta}_N$ are different but the Hawkes excitation functions $\boldsymbol{\beta}_K = \boldsymbol{\beta}_N =: \boldsymbol{\beta}$ coincide. Then:

(1) *Bound on coupling time, finer version.* For all $t \in \mathbb{R}$ one has

$$\begin{aligned}\mathbb{P}[T \le t] &= \mathbb{P}\Big[K^{(j)}|_{(t,\infty)\times\mathbb{X}_j} = N^{(j)}|_{(t,\infty)\times\mathbb{X}_j}; j = 1,\dots,d\Big] \\ &\ge \exp\Big\{-\sum_{j=1}^{d}\int_t^\infty g_j(u) + \big[\mathbb{H} * \mathbb{1}_{(-\infty,t]}\boldsymbol{m}_L\big]_j(u)du\Big\}, \qquad (6.12)\end{aligned}$$

where $\boldsymbol{g}(t) := |\boldsymbol{\eta}_K(t) - \boldsymbol{\eta}_N(t)|$, for $t \in \mathbb{R}$, and $\boldsymbol{m}_L$ is the density of the first order moment measure of the difference process $L := |K - N|$.

Now assume the first order moment density $\boldsymbol{m}_L$ is not known. In this case, one can still find a lower bound for $\mathbb{P}[T \le t]$:

(2) *Bound on coupling time, cruder version.* The above inequality remains correct, if one substitutes

$$\boldsymbol{m}_L(t) \le \boldsymbol{m}_K(t) + \boldsymbol{m}_N(t),$$

where $\boldsymbol{m}_K$ and $\boldsymbol{m}_N$ are the first order moment measures densities of K and N. ♢

6.5 Minimal Solutions

In this section, a specific strong solution is construct, which is called the minimal solution. This notion is not used in the literature, at least not in the context of Hawkes processes, but it is justified by the construction presented below.

At the moment, we cannot yet prove the uniqueness of strong solutions in full generality. But it will turn out later that under suitable conditions strong solutions are unique, see Theorem 6.55. As a consequence, the minimal solution is then actually the unique strong solution.

The construction of the minimal solution involves several intermediate point processes and before we start with the construction we give a short overview:

6.44 Remark (Notation concerning minimal solution). The construction is based on *incremental* and *cumulative* processes and to distinguish between them we use the following notation:

(1) *Incremental processes.* For each generation $n \geq 0$, the process $\hat{N}_n$ corresponds to the events of the n-th generation. The associated intensity process is $\hat{\boldsymbol{\lambda}}_n$. By convention, the first generation is generation 0, which represents the immigrants.

(2) *Cumulative processes.* Once one has constructed all generations up to order n, one can define the cumulative process containing all events up to and including generation n. This cumulative process is denoted by N_n and the corresponding intensity process by $\boldsymbol{\lambda}_n$. ♢

The following construction is taken from the proof of Theorem 4 in [BM96]:

6.45 Definition (Construction of minimal solution). Assume one has the Hawkes intensity function $\boldsymbol{\alpha}(t|\nu) = \boldsymbol{\eta}(t) + \boldsymbol{\beta}(t|\nu)$. The minimal solution is constructed inductively by:

(0) Define first the intensity process $\boldsymbol{\lambda}_{-1}(t) := \mathbf{0}$, for all $t \in \mathbb{R}$.

Next define recursively for $n \geq 0$ the following three objects:

(1) The *cumulative intensity process* $\boldsymbol{\lambda}_n$ is

$$\boldsymbol{\lambda}_n(t) := \boldsymbol{\eta}(t) + \sum_{m=1}^{n} \hat{\boldsymbol{\lambda}}_m(t), \tag{6.13}$$

for all $t \in \mathbb{R}$. By convention, one sets $\boldsymbol{\lambda}_0(t) := \boldsymbol{\eta}(t)$.

(2) The *incremental process* $\hat{N}_n$ is the thinning of the driving process ϖ in the layer between $\boldsymbol{\lambda}_{n-1}$ and $\boldsymbol{\lambda}_n$, i.e.

$$\hat{N}_n^{(j)}(E \times X) := \int_{E \times \mathbb{R}_+} \mathbb{1}_{\{\lambda_{n-1,j}(t) < z \leq \lambda_{n,j}(t)\}} \varpi^{(j)}(dt \times X \times dz), \tag{6.14}$$

for all $E \subseteq \mathbb{R}$, $X \subseteq \mathbb{X}_j$ and $1 \leq j \leq d$.

(3) The *incremental intensity process* $\hat{\boldsymbol{\lambda}}_{n+1}$ of the next generation is

$$\hat{\boldsymbol{\lambda}}_{n+1}(t) := \boldsymbol{\beta}(t|\hat{N}_n)$$

for all $t \in \mathbb{R}$. The recursion now continues with $n+1$ in the first step. $\diamondsuit$

6.46 Definition (Minimal solution). Let $\boldsymbol{\alpha}$ be a Hawkes intensity function and define the incremental processes $\hat{N}_n$ as in Definition 6.45.

(1) The *minimal solution* N_∞ is

$$N_\infty^{(j)}(E \times X) := \sum_{n=0}^{\infty} \hat{N}_n^{(j)}(E \times X),$$

for sets $E \subseteq \mathbb{R}$ and $X \subseteq \mathbb{X}$.

(2) The intensity process $\boldsymbol{\lambda}_\infty$ of the minimal solution is

$$\boldsymbol{\lambda}_\infty(t) := \lim_{n\to\infty} \boldsymbol{\lambda}_n(t) = \boldsymbol{\eta}(t) + \sum_{n=1}^{\infty} \hat{\boldsymbol{\lambda}}_n(t),$$

for all $t \in \mathbb{R}$. $\diamondsuit$

Note that without further assumptions, it is not clear, whether N_∞ is a point process, i.e. for the time being it is not clear whether N_∞ is a locally-finite point measure.

Later we will give sufficient conditions so that N_∞ and $\boldsymbol{\lambda}_\infty$ become locally-finite processes. The process N_∞ will then become our candidate for a strong solution. But even without any further assumptions, we can already now state some useful facts:

6.47 Proposition (Properties of minimal solution). Let $\boldsymbol{\alpha}$ be a Hawkes intensity function and N_∞ and $\boldsymbol{\lambda}_\infty$ be the associated minimal solution and its intensity process, as given in Definition 6.46.

(1) *First moment measure.* The density $\boldsymbol{m}_\infty$ of the first order moment measure of N_∞ is

$$\boldsymbol{m}_\infty(t) = \sum_{n=0}^{\infty} [\mathbb{H}^{*n} * \boldsymbol{\eta}](t) = [\mathbb{U}^+ * \boldsymbol{\eta}](t), \tag{6.15}$$

for all $t \in \mathbb{R}$.

(2) *Thinning property.* The limit N_∞ is a adapted to the canonical filtration. Moreover, N_∞ is a thinning of ϖ with respect to the intensity process

$\boldsymbol{\alpha}(t|N_\infty)$. In other words, N_∞ satisfies the thinning condition of a strong solution from Equation (6.9), which states that

$$N_\infty^{(j)}(E \times X) = \int_{E\times\mathbb{R}_+} \mathbb{1}_{\{z\le\alpha_j(t|N_\infty)\}}\varpi^{(j)}(dt \times X \times dz). \tag{6.16}$$

Moreover, the intensity process satisfies $\boldsymbol{\lambda}_\infty(t) = \boldsymbol{\alpha}(t|N_\infty)$, all $t \in \mathbb{R}$. $\diamondsuit$

Next we consider two simultaneous minimal solutions:

6.48 Proposition (Pair of minimal solutions). Let $\boldsymbol{\beta}$ be a Hawkes excitation function and $\boldsymbol{\eta}_K$ and $\boldsymbol{\eta}_N$ be two immigration intensities. Define the two Hawkes intensity functions

$$\boldsymbol{\alpha}_K := \boldsymbol{\eta}_K + \boldsymbol{\beta} \qquad \text{and} \qquad \boldsymbol{\alpha}_N := \boldsymbol{\eta}_N + \boldsymbol{\beta}.$$

Assume K_∞ and N_∞ are the associated minimal solutions. Then the first order moment measure of the difference process $L_\infty := |K_\infty - N_\infty|$ satisfies

$$\boldsymbol{m}_{L_\infty}(t) \le \Big[\mathbb{U}^+ * |\boldsymbol{\eta}_K - \boldsymbol{\eta}_N|\Big](t), \tag{6.17}$$

for all $t \in \mathbb{R}$. $\diamondsuit$

6.49 Theorem (Existence of minimal solution). Let $\boldsymbol{\alpha} = \boldsymbol{\eta} + \boldsymbol{\beta}$ be a Hawkes intensity function and Q the associated branching matrix, see Equation (6.6). Assume the immigration intensity $\boldsymbol{\eta}$ is bounded and the branching matrix satisfies $\mathrm{Spr}(Q) < 1$. Then the minimal solution N_∞ as given in Definition 6.46 is a strong solution.

Take the same Hawkes excitation function $\boldsymbol{\beta}$ as above and assume now that $\boldsymbol{\eta}_K$ and $\boldsymbol{\eta}_N$ are be two bounded immigration intensities. Denote the corresponding minimal solutions with K_∞ and N_∞. Then the coupling time between the two minimal solutions satisfies:

$$\mathbb{P}[T \le t] \ge \exp\Big\{-\sum_{j=1}^{d}\int_t^\infty \Big[\mathbb{U}^+ * |\boldsymbol{\eta}_K - \boldsymbol{\eta}_N|\Big]_j(u)du\Big\}, \tag{6.18}$$

for all $t \in \mathbb{R}$. $\diamondsuit$

6.6 Construction of Continuations

The aim of this section is to show that continuations are unique, given some conditions are satisfied. However, before we can tackle this problem, we need to have a closer look at Definition 6.35. Because continuations and strong solutions are formally closely related, it is not surprising that one can be constructed from the other. This is the content of the next proposition.

For simplicity, we only consider point processes without marks. The general case is a trivial extension.

6.50 Proposition (Relation to strong solutions). Let $\chi \in \mathbb{R}$ be a fixed time and consider the canonical probability space $\mathfrak{N}_+ := \mathscr{N}_{\mathbb{X}\times\mathbb{R}_+}\big((\chi,\infty)\big)$. Assume $\boldsymbol{\beta}$ is a Hawkes excitation function.

(1) *Equivalence relation between continuations.* Let ν_- be an initial state on $(-\infty,\chi]$ and $\boldsymbol{\eta}_+$ and $\hat{\boldsymbol{\eta}}_+$ be two immigration intensities on (χ,∞). Assume $\boldsymbol{\eta}_+$ and $\hat{\boldsymbol{\eta}}_+$ are related by

$$\hat{\boldsymbol{\eta}}_+(t) = \boldsymbol{\eta}_+(t) + \boldsymbol{\beta}(t|\nu_-), \qquad\qquad \boldsymbol{\eta}_+(t) = \hat{\boldsymbol{\eta}}_+(t) - \boldsymbol{\beta}(t|\nu_-).$$

for all $t \in \mathbb{R}$. Then, if one of the following triples is a continuation triple, then so is the other:

$$(\nu_-, \boldsymbol{\eta}_+, N_+) \text{ continuation triple} \qquad \Longleftrightarrow \qquad (\varnothing_-, \hat{\boldsymbol{\eta}}_+, N_+) \text{ continuation triple} \tag{6.19}$$

(2) *Relation to strong solutions.* Let $\boldsymbol{\eta}_+$ be an immigration intensity on (χ,∞). Define the extension of $\boldsymbol{\eta}_+$ to the real line by

$$\boldsymbol{\eta}(t) = \mathbb{1}_{(\chi,\infty)}(t)\boldsymbol{\eta}_+(t),$$

for $t \in \mathbb{R}$. Then, for a point process N_+ on the time interval (χ,∞):

$$(\varnothing_-, \boldsymbol{\eta}_+, N_+) \text{ continuation triple} \iff (\boldsymbol{\eta}, N_+) \text{ strong solution.} \tag{6.20}$$

♢

We call the process $\hat{\boldsymbol{\eta}}_+$ given above the *augmented immigration intensity.*

6.51 Corollary (Uniqueness of continuations). Let $\boldsymbol{\beta}$ be a Hawkes excitation function, ν_- an initial state on $(-\infty,\chi]$ and $\boldsymbol{\eta}_+$ an immigration intensity

on (χ, ∞).

(1) *Relation to strong solution.* Define the augmented immigration intensity extended to the real line by

$$\hat{\boldsymbol{\eta}}(t) := \mathbb{1}_{(\chi,\infty)}(t)\Big[\boldsymbol{\eta}_+(t) + \boldsymbol{\beta}(t|\nu_-)\Big] \tag{6.21}$$

and let N_+ be a point process on the time interval (χ, ∞). Then:

$$(\nu_-, \boldsymbol{\eta}_+, N_+) \text{ continuation triple} \quad \Leftrightarrow \quad (\hat{\boldsymbol{\eta}}, N_+) \text{ strong solution.} \tag{6.22}$$

(2) *Uniqueness.* There is at most one continuation N_+ of the pair $(\nu_-, \boldsymbol{\eta}_+)$, that is if

$$(\nu_-, \boldsymbol{\eta}_+, N_+) \qquad \text{and} \qquad (\nu_-, \boldsymbol{\eta}_+, \tilde{N}_+)$$

are two continuation triples, then $N_+ = \tilde{N}_+$. ♢

6.52 Proposition (Construction of continuation). Let $\chi \in \mathbb{R}$ be a fixed time and $\boldsymbol{\beta}$ be a Hawkes excitation function.

(1) *Construction of continuation.* Assume ν_- is an initial state on $(-\infty, \chi]$ and $\boldsymbol{\eta}_+$ an immigration intensity on (χ, ∞). Let $\hat{\boldsymbol{\eta}}$ be the augmented immigration intensity, extended to the real line, as given in Equation (6.21). Now assume the minimal solution N_∞ generated by $\hat{\boldsymbol{\eta}}$ exists. Then the continuation triple

$$\big(\nu_-, \boldsymbol{\eta}_+, N_{\infty,+}\big), \qquad N_{\infty,+} := N_\infty|_{(\chi,\infty)}, \tag{6.23}$$

is the unique continuation of the pair $(\nu_-, \boldsymbol{\eta}_+)$.

(2) *Coupling with two initial conditions.* Let κ_- and ν_- be two initial states on $(-\infty, \chi]$ and $\boldsymbol{\eta}_+$ an immigration intensity on (χ, ∞). Denote by $\hat{\boldsymbol{\eta}}_\kappa$ and $\hat{\boldsymbol{\eta}}_\nu$ the associated augmented immigration intensities, extended to the real line, as given in Equation (6.21). Assume the corresponding minimal solutions K_∞ and N_∞ exist, so that one has the two continuation triples

$$(\kappa_-, \boldsymbol{\eta}_+, K_{\infty,+}) \qquad \text{and} \qquad (\nu_-, \boldsymbol{\eta}_+, N_{\infty,+}).$$

Then the coupling time between $K_{\infty,+}$ and $N_{\infty,+}$ satisfies

$$\mathbb{P}[T \leq t] \geq \exp\Big\{-\sum_{j=1}^{d} \int_t^{\infty} \big[\mathbb{U}^+ * \boldsymbol{g}_{\kappa_-,\nu_-}\big]_j(u)du\Big\}, \tag{6.24}$$

all $t \in \mathbb{R}$, where $\boldsymbol{g}_{\kappa_-,\nu_-}(t) := \mathbb{1}_{(\chi,\infty)}(t)\big|\boldsymbol{\beta}(t|\kappa_-) - \boldsymbol{\beta}(t|\nu_-)\big|$, for $t \in \mathbb{R}$. $\diamondsuit$

6.7 Existence and Uniqueness

Before we give the main result concerning the uniqueness of strong solutions, let us introduce some further notation concerning the canonical probability space. We intend to split the canonical probability space into two parts, where the first part comprises all events up to time χ and the second part all events after time χ.

Although the following definition is not needed for the formulation of Theorem 6.55 below, it is needed in the proof. We only consider the case of a univariate point processes without marks. The extension to the general case is trivial.

6.53 Definition (Splitting of canonical probability space). Let $\chi \in \mathbb{R}$ be a fixed time. Let ϖ be the driving process on the univariate canonical probability space $\mathfrak{N} := \mathscr{N}_{\mathbb{R}_+}(\mathbb{R})$. Consider the two time intervals

$$\mathscr{T}_- := (-\infty, \chi] \qquad \text{and} \qquad \mathscr{T}_+ := (\chi, \infty).$$

The point configuration space can be split up according to:

$$\mathscr{N}_{\mathbb{R}_+}(\mathbb{R}) = \mathscr{N}_{\mathbb{R}_+}(\mathscr{T}_-) \times \mathscr{N}_{\mathbb{R}_+}(\mathscr{T}_+), \qquad \mathbb{R} = \mathscr{T}_- \cup \mathscr{T}_+.$$

As a consequence, one can decompose the canonical probability space $(\mathfrak{N}, \mathfrak{F}, \mathbb{P})$ according to:

$$\mathfrak{N} = \mathfrak{N}_- \times \mathfrak{N}_+, \qquad \mathfrak{F} = \mathfrak{F}_- \otimes \mathfrak{F}_+, \qquad \mathbb{P} = \mathbb{P}_- \otimes \mathbb{P}_+. \qquad \diamondsuit$$

The above definition shows that a configuration $\varpi \in \mathfrak{N}$ of the driving process can be written as the pair (ϖ_-, ϖ_+), where $\varpi_- \in \mathfrak{N}_-$ describes the events up to time χ and $\varpi_+ \in \mathfrak{N}_+$ describes the events after time χ.

6.54 Remark (Independence of the two parts). Since the driving process ϖ is a Poisson process on the space $\mathbb{R} \times \mathbb{R}_+$, the two components ϖ_- and ϖ_+ are two independent Poisson processes with realizations in the two point configuration spaces

$$\varpi_- \in \mathscr{N}_{\mathbb{R}_+}(\mathscr{T}_-), \qquad \varpi_+ \in \mathscr{N}_{\mathbb{R}_+}(\mathscr{T}_+).$$

This shows that $\mathbb{P}$ can indeed be written as a product probability measure. $\diamondsuit$

The next theorem combines results from Section 1 and Appendix A in [Tor02]. Recall also the definition of the averaged transfer functions H_{jk} given in Equation (6.5) and the definition of the branching matrix Q given in Equation (6.6).

6.55 Theorem (Uniqueness of strong solutions). Let $\boldsymbol{\alpha} = \boldsymbol{\eta} + \boldsymbol{\beta}$ be a Hawkes intensity function with immigration intensity $\boldsymbol{\eta}$ and excitation function $\boldsymbol{\beta}$. Assume the following two conditions are satisfied:

(1) The *branching matrix* Q satisfies $\mathrm{Spr}(Q) < 1$.

(2) The *averaged transfer functions* satisfy for all $1 \leq j, k \leq d$:

$$\int_0^\infty t H_{jk}(t) dt < \infty.$$

If K and N are two strong solutions with bounded first order moment measure densities, then K and N coincide with probability one. $\diamondsuit$

Proofs for Chapter 6

PROOF (PROPOSITION 6.27): The likelihood function, see Equation (6.3), is the difference of two integrals. We calculate these two integrals in turn:

(i) *First integral.* Since $\alpha_j(t, x|\nu) = f_j(x)\alpha_j(t|\nu)$,

$$\begin{aligned}
&\int_{D\times\mathbb{X}_j} \log \alpha_j(t, x|\nu)\nu^{(j)}(dt \times dx) \\
&\quad = \int_{D\times\mathbb{X}_j} \log\big[f_j(x)\alpha_j(t|\nu)\big]\nu^{(j)}(dt \times dx) \\
&\quad = \int_D \log \alpha_j(t|\nu)\nu^{(j)}(dt \times \mathbb{X}_j) + \int_{\mathbb{X}_j} \log f_j(x)\nu^{(j)}(D \times dx).
\end{aligned}$$

$\nu^{(j)}(dt \times \mathbb{X}_j)$ is the projection of $\nu^{(j)}$ onto the first component. Similarly, $\nu^{(j)}(D \times dx)$ is the projection onto the second component. Now use that $D = [T_*, T^*]$, and the first part follows.

(ii) *Second integral.* It remains to deal with the second integral. Note that $\int_{\mathbb{X}_j} f_j(x)\mu_{\mathbb{X}_j}(dx) = 1$, since f_j is a probability density on the space $\mathbb{X}_j$. Hence

$$\begin{aligned}
\int_{D\times\mathbb{X}_j} \alpha_j(t, x|\nu)\mu_{\mathbb{X}_j}(dx)dt &= \int_{D\times\mathbb{X}_j} f_j(x)\alpha_j(t|\nu)\mu_{\mathbb{X}_j}(dx)dt \\
&= \int_D \alpha_j(t|\nu)\Big[\int_{\mathbb{X}_j} f_j(x)\mu_{\mathbb{X}_j}(dx)\Big]dt = \int_D \alpha_j(t|\nu)dt \\
&= \int_D \Big[\eta_j(t) + \beta_j(t|\nu)\Big]dt \\
&= \int_D \eta_j(t)dt + \int_D \Big[\sum_{k=1}^d \int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s, x)\nu^{(k)}(ds \times dx)\Big]dt.
\end{aligned}$$

The first term is one of the terms in the claimed formula. It remains to decompose the second integral. If one omits the sum, one has:

$$\int_D \Big[\int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)\nu^{(k)}(ds\times dx)\Big]dt$$
$$= \int_{\mathbb{R}^2\times\mathbb{X}_k} \mathbb{1}_{(-\infty,t)\times D}(s,t)h_{jk}(t-s,x)\nu^{(k)}(ds\times dx)dt.$$

Using that $D = [T_*, T^*]$ and defining $r := (t-s)$, one has:

$$\mathbb{1}_{(-\infty,t)\times D}(s,t) = \mathbb{1}_{(0,+\infty)}(r)\mathbb{1}_{[T_*-s,T^*-s]}(r).$$

Also note that

$$\int_0^\infty \mathbb{1}_{(a,b)}(r)h_{jk}(r,x)dr = \bar{h}_{jk}(b,x) - \bar{h}_{jk}(a,x), \text{ for any } a < b.$$

This is even true for negative a and b, because by definition $\bar{h}_{jk}(r,x) = 0$, whenever $r < 0$. Hence:

$$\int_{\mathbb{R}^2\times\mathbb{X}_k} \mathbb{1}_{(-\infty,t)\times D}(s,t)h_{jk}(t-s,x)\nu^{(k)}(ds\times dx)dt$$
$$= \int_{\mathbb{R}\times\mathbb{X}_k} \Big[\int_0^\infty \mathbb{1}_{[T_*-s,T^*-s]}(r)h_{jk}(r,x)dr\Big]\nu^{(k)}(ds\times dx)$$
$$= \int_{\mathbb{R}\times\mathbb{X}_k} \Big[\bar{h}_{jk}(T^*-s,x) - \bar{h}_{jk}(T_*-s,x)\Big]\nu^{(k)}(ds\times dx).$$

Finally, one can replace the integration domain $\mathbb{R}$ with $(-\infty, T^*)$, since outside this set the integrand vanishes. □

Recall that a process has stationary and component-wise unpredictable marks if its time-space intensity function is of the form $\lambda_j(t,x|\varpi) = f_j(x)\lambda_j(t|\varpi)$.

6.56 Lemma (Properties of Hawkes excitation function). Let $\boldsymbol{\beta}$ be a multivariate Hawkes excitation function. For two point configurations κ and ν in $\mathcal{N}_{\mathbb{X}}(\mathbb{R})$, one has:

$$\big|\boldsymbol{\beta}(t|\kappa) - \boldsymbol{\beta}(t|\nu)\big| \leq \boldsymbol{\beta}\big(t\big||\kappa-\nu|\big). \tag{6.25}$$

Now let f_j be densities on $\mathbb{X}_j$. Assume N is a predictable point process with

multivariate family of intensity functions $\{\lambda_j, j = 1, \ldots, d\}$ such that

$$\lambda_j(t, x|\varpi) = f_j(x)\lambda_j(t|\varpi),$$

and let $\boldsymbol{m}_N$ be the density of the first order moment measure. Then

$$\mathbb{E}\big[\boldsymbol{\beta}(t|N)\big] = [\mathbb{H} * \boldsymbol{m}_N](t). \tag{6.26}$$

♢

Proof (Lemma 6.56): Recall the definition of the Hawkes excitation function $\boldsymbol{\beta}$, given in Equation (6.4).

(*i*) The first part follows from

$$\begin{aligned}
\beta_j(t|\kappa) - \beta_j(t|\nu) &= \sum_{k=1}^{d} \int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)\kappa^{(k)}(ds\times dx) \\
&\quad - \sum_{k=1}^{d} \int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)\nu^{(k)}(ds\times dx) \\
&= \sum_{k=1}^{d} \int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)\Big[\kappa^{(k)}(ds\times dx) - \nu^{(k)}(ds\times dx)\Big] \\
&\leq \sum_{k=1}^{d} \int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)\big|\kappa^{(k)} - \nu^{(k)}\big|(ds\times dx) \\
&= \beta_j\big(t\big||\kappa-\nu|\big).
\end{aligned}$$

(*ii*) By assumption, the intensity measure of $N^{(k)}$ is

$$\Lambda_k(ds\times dx) = \lambda_k(t, x|\varpi)\mu_{\mathbb{X}_k}(dx)ds = f_k(x)\lambda_k(t|\varpi)\mu_{\mathbb{X}_k}(dx)ds.$$

Due to the fundamental property of the intensity measure, and the defi-

nition of the functions H_{jk}, see Equation (6.5), we have

$$\begin{aligned}
\mathbb{E}\big[\beta_j(t|N)\big] &= \mathbb{E}\Big[\sum_{k=1}^{d}\int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)N^{(k)}(ds\times dx)\Big] \\
&= \sum_{k=1}^{d}\mathbb{E}\Big[\int_{(-\infty,t)\times\mathbb{X}_k} h_{jk}(t-s,x)f_k(x)\lambda_k(s|\varpi)\mu_{\mathbb{X}_k}(dx)ds\Big] \\
&= \mathbb{E}\Big[\int_{(-\infty,t)}\Big\{\int_{\mathbb{X}_k} h_{jk}(t-s,x)f_k(x)\mu_{\mathbb{X}_k}(dx)\Big\}\lambda_k(s|\varpi)ds\Big] \\
&= \mathbb{E}\Big[\int_{-\infty}^{t} H_{jk}(t-s)\lambda_k(s|\varpi)ds\Big] = \int_{-\infty}^{t} H_{jk}(t-s)\,\mathbb{E}\big[\lambda_k(s|\varpi)\big]ds.
\end{aligned}$$

Recall that $\mathbb{E}[\lambda_k(s|\varpi)] = m_{N,k}(s)$. By rewriting the integral as a multivariate convolution, the statement follows from

$$\mathbb{E}\big[\beta_j(t|N)\big] = \sum_{k=1}^{d}\int_{-\infty}^{t} H_{jk}(t-s)m_{N,k}(s)ds = [\mathbb{H} * \boldsymbol{m}_N]_j(t). \qquad \square$$

Proof (Proposition 6.39): Let N be a point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$, with vector-valued intensity function $\boldsymbol{\lambda}$.

(*i*) *Exact hazard rate.* In a first step we consider only the case where N takes values in $\mathscr{N}(\mathbb{R})$, i.e. it is a univariate point process without marks.

(*ii*) Fix $s \in \mathbb{R}$. Note that for $\nu \in \mathscr{N}(\mathbb{R})$ one has

$$\int_{(s,t)} \mathbb{1}_{\{\nu(s,u)=0\}}\nu(du) = \mathbb{1}_{\{\nu(s,t)\geq 1\}}.$$

We are going to use the short notation $\nu(s,t) \equiv \nu\big((s,t)\big)$. Replacing now ν with N, one has as a first intermediate result

$$\mathbb{1}_{\{N(s,t)=0\}} = 1 - \int_{(s,t)} \mathbb{1}_{\{N(s,u)=0\}}N(du).$$

Since $N(s,u)$ is $\mathfrak{F}_{u-}$-measurable, the processes $H(u) := N(s,u)$ and $G(u) := \mathbb{1}_{\{N(s,u)=0\}}$ are predictable, see also Definition 6.4. Hence, one has by the fundamental property of the intensity function, see Defini-

tion 6.10, as a second intermediate result

$$\mathbb{E}\Big[\int_{(s,t)} \mathbb{1}_{\{N(s,u)=0\}} N(du)\Big|\mathfrak{F}_s\Big] = \mathbb{E}\Big[\int_s^t \mathbb{1}_{\{N(s,u)=0\}} \lambda(u|\varpi)du\Big|\mathfrak{F}_s\Big].$$

Combining now the two above results gives:

$$\begin{aligned}\mathbb{P}\big[N(s,t)=0\big|\mathfrak{F}_s\big] &= \mathbb{E}\Big[1-\int_{(s,t)} \mathbb{1}_{\{N(s,u)=0\}} N(du)\Big|\mathfrak{F}_s\Big] \\ &= 1-\mathbb{E}\Big[\int_s^t \mathbb{1}_{\{N(s,u)=0\}} \lambda(u|\varpi)du\Big|\mathfrak{F}_s\Big]. \qquad (6.27)\end{aligned}$$

Since λ^+ is a hazard rate, we get, due to Equation (6.10), that

$$\begin{aligned}\mathbb{P}\big[N(s,t)=0\big|\mathfrak{F}_s\big] &= 1-\mathbb{E}\Big[\int_s^t \mathbb{1}_{\{N(s,u)=0\}} \lambda^+(s,u|\varpi)du\Big|\mathfrak{F}_s\Big] \\ &= 1-\int_s^t \mathbb{E}\Big[\mathbb{1}_{\{N(s,u)=0\}} \lambda^+(s,u|\varpi)\Big|\mathfrak{F}_s\Big]du \\ &= 1-\int_s^t \mathbb{E}\Big[\mathbb{1}_{\{N(s,u)=0\}}\Big|\mathfrak{F}_s\Big]\lambda^+(s,u|\varpi)du.\end{aligned}$$

We have used that $\lambda^+(s,u|\varpi)$ is $\mathfrak{F}_s$-measurable, for all $u > s$. In summary, we have the integral equation:

$$\mathbb{P}\big[N(s,t)=0\big|\mathfrak{F}_s\big] = 1-\int_s^t \mathbb{P}\big[N(s,u)=0\big|\mathfrak{F}_s\big]\lambda^+(s,u|\varpi)du. \qquad (6.28)$$

This equation for the function $t \mapsto \mathbb{P}\big[N(s,t)=0\big|\mathfrak{F}_s\big]$ can be rewritten as an ordinary differential equation. Its unique solution is

$$\mathbb{P}\big[N(s,t)=0\big|\mathfrak{F}_s\big] = \exp\Big\{-\int_s^t \lambda^+(s,u|\varpi)du\Big\}. \qquad (6.29)$$

(*ii*) In the general case, let N be a multivariate point process with values in $\mathscr{N}_{\mathbb{X}}(\mathbb{R})$. For more details about the two processes N_* and λ_*, see Remark 6.11:

$$N_*(dt) := \sum_{j=1}^d N^{(j)}(dt \times \mathbb{X}_j), \qquad \lambda_*(t|\varpi) := \sum_{j=1}^d \lambda_j(t|\varpi).$$

Moreover, define $\lambda_*^+(s,t|\varpi) := \sum_{j=1}^d \lambda_j^+(s,t|\varpi)$. Then

$$\lambda_*(t|\varpi) = \lambda_*^+(s,t|\varpi) \text{ on the set } \big\{\varpi : N(s,t)[\varpi] = 0\big\}.$$

Now apply Equation (6.29) for the process N_*, so that

$$\begin{aligned}\mathbb{P}\Big[\bigcap_{j=1}^d N^{(j)}\big((s,t)\times\mathbb{X}_j\big) = 0\Big|\mathfrak{F}_s\Big] &= \mathbb{P}\Big[N_*\big((s,t)\big) = 0\Big|\mathfrak{F}_s\Big] \\ &= \exp\Big\{-\int_s^t \lambda_*^+(s,u|\varpi)du\Big\} = \exp\Big\{-\sum_{j=1}^d\int_s^t \lambda_j^+(s,u|\varpi)du\Big\}.\end{aligned}$$

(2) *Upper bound for hazard rate.* We state a special case *Gronwall's lemma*: The proof can e.g. be found in the original papers [Gro19], [Bel43], or any real analysis book. Let c be a constant and x, α two continuous functions on (s,∞). Then any solution x of the inequality

$$x(t) \le c + \int_s^t \alpha(u)x(u)du$$

satisfies

$$x(t) \le c\exp\Big\{\int_s^t \alpha(u)du\Big\}. \tag{6.30}$$

(i) We first treat the special case where N is a univariate point process without marks. Hence let $\lambda^+(s,t|\varpi)$ be an upper bound for the hazard rate. The proof is identical up to Equation (6.27). From there continue with

$$\begin{aligned}\mathbb{P}\big[N(s,t) = 0\big|\mathfrak{F}_s\big] &= 1 - \mathbb{E}\Big[\int_s^t \mathbb{1}_{\{N(s,u)=0\}}\lambda(u|\varpi)du\Big|\mathfrak{F}_s\Big] \\ &\ge 1 - \mathbb{E}\Big[\int_s^t \mathbb{1}_{\{N(s,u)=0\}}\tilde\lambda^+(s,u|\varpi)du\Big|\mathfrak{F}_s\Big].\end{aligned}$$

Again continue in the same way as in the first part, but use this time the upper bound for the hazard rate. Instead of Equation (6.28), one now finds:

$$\mathbb{P}\big[N(s,t) = 0\big|\mathfrak{F}_s\big] \ge 1 - \int_s^t \mathbb{P}\big[N(s,u) = 0\big|\mathfrak{F}_s\big]\tilde\lambda^+(s,u|\varpi)du.$$

Now set $y(s,t) := \mathbb{P}\big[N(s,t) = 0 \big| \mathfrak{F}_s\big]$, such that

$$y(s,t) \geq 1 - \int_s^t y(s,u)\tilde{\lambda}^+(s,u|\varpi)du.$$

Here the similarity with the first part of the proof ends and one has to continue in a different way. First fix some $s \in \mathbb{R}$ and define $x(s,t) := -y(s,t)$, so that

$$x(s,t) \leq (-1) + \int_s^t \big[-\tilde{\lambda}^+(s,u|\varpi)\big]x(s,u)du.$$

Due to Gronwall's lemma, especially Equation (6.30), one finds

$$x(s,t) \leq (-1)\exp\Big\{\int_s^t \big[-\tilde{\lambda}^+(s,u|\varpi)\big]du\Big\}.$$

In terms of $y(s,t)$, this implies

$$\mathbb{P}\big[N(s,t) = 0 \big| \mathfrak{F}_s\big] = y(s,t) = -x(s,t) \geq \exp\Big\{-\int_s^t \tilde{\lambda}^+(s,u|\varpi)du\Big\}.$$

(*ii*) In the same way as in the first part, the statement can now be extended to the general case where N is a multivariate, marked point process. □

Proof (Corollary 6.40): Define $\lambda^+(s,t|\varpi) := \gamma(t|N|_{(-\infty,s]})$, as suggested, and similarly $\tilde{\lambda}^+$.

(*1*) *Exact hazard rate.* Because γ is an intrinsic intensity function, and every intrinsic intensity function is by definition $\mathfrak{P}_{\mathbb{X}}(\mathbb{R})$-predictable, we have $\gamma(t|\nu) = \gamma(t|\nu|_{(-\infty,t)})$. We need to check Equation (6.10). But given that $\{N(s,t) = 0\}$, this follows from

$$\begin{aligned}\lambda(t|\varpi) = \gamma(t|N) = \gamma(t|N|_{(-\infty,t)}) &= \gamma\big(t\big|N|_{(-\infty,s]} + N|_{(s,t)}\big) \\ &= \gamma(t|N|_{(-\infty,s]}) = \lambda^+(s,t|\varpi).\end{aligned}$$

(*2*) *Upper bound for hazard rate.* The proof is almost identical to the first part. □

Proof (Proposition 6.43): Let K and N be two strong solutions associated with the Hawkes intensity functions $\boldsymbol{\alpha}_K = \boldsymbol{\eta}_K + \boldsymbol{\beta}$ and $\boldsymbol{\alpha}_N = \boldsymbol{\eta}_N + \boldsymbol{\beta}$. Furthermore, let $L := |K - N|$ denote the difference process.

(1) *Bound on coupling time, finer version.* Consider the two intensity processes $\boldsymbol{\lambda}_K(t) := \boldsymbol{\alpha}_K(t|K)$ and $\boldsymbol{\lambda}_N(t) := \boldsymbol{\alpha}_N(t|N)$. Because K and N are strong solutions, they are thinnings of ϖ with respect to $\boldsymbol{\lambda}_K$ and $\boldsymbol{\lambda}_N$, respectively. The difference process L is therefore again a thinning given by

$$L^{(j)}(E \times X) = \int_{E \times \mathbb{R}_+} \mathbb{1}_{\{\zeta_j^-(t) < z < \zeta_j^-(t)\}} \varpi^{(j)}(dt \times X \times dz),$$

where the thinning is with respect to the layer between the two bounds

$$\boldsymbol{\zeta}^-(t) := \min\big\{\boldsymbol{\lambda}_K(t), \boldsymbol{\lambda}_N(t)\big\}, \qquad \boldsymbol{\zeta}^+(t) := \max\big\{\boldsymbol{\lambda}_K(t), \boldsymbol{\lambda}_N(t)\big\}.$$

Due to Theorem 6.37, the difference process L has intensity process

$$\begin{aligned}\boldsymbol{\lambda}_L(t) &= \boldsymbol{\zeta}^+(t) - \boldsymbol{\zeta}^-(t) = \big|\boldsymbol{\lambda}_K(t) - \boldsymbol{\lambda}_N(t)\big| = \big|\boldsymbol{\alpha}_K(t|K) - \boldsymbol{\alpha}_N(t|N)\big| \\ &= \big|\boldsymbol{\eta}_K(t) + \boldsymbol{\beta}(t|K) - \boldsymbol{\eta}_N(t) - \boldsymbol{\beta}(t|N)\big|.\end{aligned}$$

Define $\boldsymbol{g}(t) := \big|\boldsymbol{\eta}_K(t) - \boldsymbol{\eta}_N(t)\big|$ and use Equation (6.25), so that

$$\begin{aligned}\boldsymbol{\lambda}_L(t) &\le \big|\boldsymbol{\eta}_K(t) - \boldsymbol{\eta}_N(t)\big| + \big|\boldsymbol{\beta}(t|K) - \boldsymbol{\beta}(t|N)\big| \\ &\le \boldsymbol{g}(t) + \boldsymbol{\beta}\big(t\big||K - N|\big) = \boldsymbol{g}(t) + \boldsymbol{\beta}(t|L).\end{aligned}$$

This inequality suggests the following definition for the intrinsic intensity function $\tilde{\boldsymbol{\gamma}}$, which will serve as an upper bound for the intensity process $\boldsymbol{\lambda}_L$, but only depends on the past of L:

$$\tilde{\boldsymbol{\gamma}}(t|\nu) := \boldsymbol{g}(t) + \boldsymbol{\beta}(t|\nu), \text{ for } \nu \in \mathscr{N}_{\mathbb{X}}(\mathbb{R}).$$

Then $\tilde{\boldsymbol{\gamma}}$ is an upper bound for $\boldsymbol{\lambda}_L$ in the sense that $\boldsymbol{\lambda}_L(t) \le \tilde{\boldsymbol{\gamma}}(t|L)$. Since $\tilde{\boldsymbol{\gamma}}$ satisfies the required conditions of Corollary 6.40, an upper bound for the hazard rate is given by

$$\tilde{\boldsymbol{\lambda}}^+(s,t) := \tilde{\boldsymbol{\gamma}}\big(t\big|L|_{(-\infty,s]}\big) = \boldsymbol{g}(t) + \boldsymbol{\beta}\big(t\big|L|_{(-\infty,s]}\big).$$

Hence, due to Proposition 6.39, this shows

$$\mathbb{P}\big[L(s,t) = 0\big|\mathfrak{F}_s\big] \ge \exp\Big\{-\sum_{j=1}^{d} \int_s^t \tilde{\lambda}_j^+(s,u)du\Big\}.$$

For simplicity, we write $L(s,t)$ instead of $L\big((s,t)\times\mathbb{X}\big)$. This shows, because of $\big\{L(s,\infty)=0\big\}=\big\{T\le s\big\}$ and using bounded convergence, that

$$\begin{aligned}\mathbb{P}\big[T\le s\big|\mathfrak{F}_s\big]&=\mathbb{P}\big[L(s,\infty)=0\big|\mathfrak{F}_s\big]=\lim_{t\to\infty}\mathbb{P}\big[L(s,t)=0\big|\mathfrak{F}_s\big]\\&\ge\lim_{t\to\infty}\exp\Big\{-\sum_{j=1}^{d}\int_s^t\tilde\lambda_j^+(s,u)du\Big\}=\exp\Big\{-\sum_{j=1}^{d}\int_s^\infty\tilde\lambda_j^+(s,u)du\Big\}.\end{aligned}$$

Now take the expectation on both sides. Then apply Jensen's inequality, Fubini's theorem, and use the variable t instead of s. According to the definition of $\tilde\lambda_j$, one obtains:

$$\begin{aligned}\mathbb{P}[T\le t]&\ge\exp\Big\{-\sum_{j=1}^{d}\int_t^\infty\mathbb{E}\big[\tilde\lambda_j^+(t,u)\big]du\Big\}\\&=\exp\Big\{-\sum_{j=1}^{d}\int_t^\infty g_j(u)+\mathbb{E}\Big[\beta_j\big(u\big|L|_{(-\infty,t]}\big)\Big]du\Big\}.\end{aligned}$$

Due to Equation (6.26), one knows

$$\mathbb{E}\Big[\boldsymbol{\beta}\big(u\big|L|_{(-\infty,t]}\big)\Big]=\Big[\mathbb{H}*\boldsymbol{m}_{L|_{(-\infty,t]}}\Big](u)=\Big[\mathbb{H}*\mathbb{1}_{(-\infty,t]}\boldsymbol{m}_L\Big](u).$$

Substitute this in the above inequality, and the claimed result follows.

(2) *Bound on coupling time, cruder version.* Since $L=|K-N|\le K+N$, the analog inequality holds for the densities of the first order moment measures. □

Proof (Proposition 6.47): Let $\hat N_n$, $\hat{\boldsymbol{\lambda}}_n$ be the incremental processes and N_n, $\boldsymbol{\lambda}_n$ the associated cumulative processes.

(1) *First moment measure.* According to Equation (6.14), $\hat N_n$ is the thinning of the driving process ϖ in the layer between $\boldsymbol{\lambda}_{n-1}$ and $\boldsymbol{\lambda}_n$. With Equation (6.13) one obtains $\boldsymbol{\lambda}_n-\boldsymbol{\lambda}_{n-1}=\hat{\boldsymbol{\lambda}}_n$. Therefore, due to Theorem 6.37, the space-time intensity processes of $\hat N_n$ are

$$\hat\lambda_{n,j}(t,x)=f_j(x)\hat\lambda_{n,j}(t),$$

for $1\le j\le d$. For the first order moment density $\hat{\boldsymbol{m}}_n$ of $\hat N_n$ one has,

using that f_j is a probability density:

$$\begin{aligned}\hat{m}_{n,j}(t) &= \mathbb{E}\Big[\int_{\mathbb{X}_j} \hat{\lambda}_{n,j}(t,x)\mu_{\mathbb{X}_j}(dx)\Big] = \mathbb{E}\Big[\int_{\mathbb{X}_j} \hat{\lambda}_{n,j}(t) f_j(x)\mu_{\mathbb{X}_j}(dx)\Big] \\ &= \mathbb{E}\Big[\hat{\lambda}_{n,j}(t)\int_{\mathbb{X}_j} f_j(x)\mu_{\mathbb{X}_j}(dx)\Big] = \mathbb{E}\big[\hat{\lambda}_{n,j}(t)\big].\end{aligned}$$

Now, due to Equation (6.26), one obtains a recursive formula for $\hat{\boldsymbol{m}}_n$:

$$\hat{\boldsymbol{m}}_n(t) = \mathbb{E}\big[\hat{\boldsymbol{\lambda}}_n(t)\big] = \mathbb{E}\big[\boldsymbol{\beta}(t|\hat{N}_{n-1})\big] = \big[\mathbb{H} * \hat{\boldsymbol{m}}_{n-1}\big](t).$$

The initial value $\hat{\boldsymbol{m}}_0$ of this recursion is

$$\hat{\boldsymbol{m}}_0(t) = \mathbb{E}\big[\boldsymbol{\lambda}_0(t)\big] = \mathbb{E}\big[\boldsymbol{\eta}(t)\big] = \boldsymbol{\eta}(t).$$

This shows $\hat{\boldsymbol{m}}_n(t) = [\mathbb{H}^{*n} * \boldsymbol{\eta}](t)$. The claim now follows with

$$\begin{aligned}\boldsymbol{m}_\infty(t) &= \mathbb{E}\big[\boldsymbol{\lambda}_\infty(t)\big] = \mathbb{E}\Big[\sum_{n=0}^{\infty}\hat{\boldsymbol{\lambda}}_n(t)\Big] = \sum_{n=0}^{\infty}\mathbb{E}\big[\hat{\boldsymbol{\lambda}}_n(t)\big] = \sum_{n=0}^{\infty}\hat{\boldsymbol{m}}_n(t) \\ &= \sum_{n=0}^{\infty}[\mathbb{H}^{*n} * \boldsymbol{\eta}](t) = \Big[\Big(\sum_{n=0}^{\infty}\mathbb{H}^{*n}\Big) * \boldsymbol{\eta}\Big](t) = [\mathbb{U}^{+} * \boldsymbol{\eta}](t).\end{aligned}$$

(2) *Thinning property.* Note that $\boldsymbol{\lambda}_\infty$ satisfies $\boldsymbol{\lambda}_\infty(t) = \lim_{n\to\infty}\boldsymbol{\lambda}_n(t)$. Therefore, with monotone convergence:

$$\begin{aligned}N_\infty^{(j)}(E\times X) &= \sum_{n=0}^{\infty}\hat{N}_n^{(j)}(E\times X) \\ &= \sum_{n=0}^{\infty}\int_{E\times\mathbb{R}_+}\mathbb{1}_{\{\lambda_{n-1,j}(t)<z\le\lambda_{n,j}(t)\}}\varpi^{(j)}(dt\times X\times dz) \\ &= \lim_{n\to\infty}\int_{E\times\mathbb{R}_+}\mathbb{1}_{\{z\le\lambda_{n,j}(t)\}}\varpi^{(j)}(dt\times X\times dz) \\ &= \int_{E\times\mathbb{R}_+}\mathbb{1}_{\{z\le\lambda_{\infty,j}(t)\}}\varpi^{(j)}(dt\times X\times dz).\end{aligned}$$

It remains to show $\boldsymbol{\lambda}_\infty(t) = \boldsymbol{\alpha}(t|N_\infty)$. But this follows, using monotone

convergence, from

$$\boldsymbol{\lambda}_\infty(t) = \boldsymbol{\eta}(t) + \sum_{n=1}^{\infty} \hat{\boldsymbol{\lambda}}_n(t) = \boldsymbol{\eta}(t) + \sum_{n=1}^{\infty} \boldsymbol{\beta}(t|\hat{N}_{n-1}) = \boldsymbol{\eta}(t) + \boldsymbol{\beta}\Big(t\Big|\sum_{n=1}^{\infty} \hat{N}_{n-1}\Big)$$
$$= \boldsymbol{\eta}(t) + \boldsymbol{\beta}\Big(t\Big|\sum_{n=0}^{\infty} \hat{N}_n\Big) = \boldsymbol{\eta}(t) + \boldsymbol{\beta}(t|N_\infty) = \boldsymbol{\alpha}(t|N_\infty). \qquad \square$$

PROOF (PROPOSITION 6.48): Let $\boldsymbol{\eta}_K$, $\boldsymbol{\eta}_N$ be two immigration intensities and K_∞, N_∞ the associated minimal solutions.

(*i*) *Difference process.* Let K_n and N_n be the cumulative point processes corresponding to the immigration intensities $\boldsymbol{\eta}_K$ and $\boldsymbol{\eta}_N$, see the construction in Definition 6.45. Now consider the difference processes $L_n := |K_n - N_n|$.

(*ii*) *Thinning representation of difference process.* Define the vector-valued function $\boldsymbol{g}(t) := \big|\boldsymbol{\eta}_K(t) - \boldsymbol{\eta}_N(t)\big|$. According to the construction, K_n, N_n have the associated intensity processes $\boldsymbol{\lambda}_{K_n}$, $\boldsymbol{\lambda}_{N_n}$, which are

$$\boldsymbol{\lambda}_{K_n}(t) = \boldsymbol{\eta}_K(t) + \sum_{m=1}^{n} \hat{\boldsymbol{\lambda}}_{K_m}(t) = \boldsymbol{\eta}_K(t) + \sum_{m=1}^{n} \boldsymbol{\beta}(t|\hat{K}_{m-1})$$
$$= \boldsymbol{\eta}_K(t) + \boldsymbol{\beta}\Big(t\Big|\sum_{m=0}^{n-1} \hat{K}_m\Big) = \boldsymbol{\eta}_K(t) + \boldsymbol{\beta}(t|K_{n-1}),$$

and similarly for $\boldsymbol{\lambda}_{N_n}$. From the construction of K_n, N_n follows

$$L_n^{(j)}(E \times X) = \int_{E\times\mathbb{R}} \mathbb{1}_{\{\zeta_{n,j}^-(t) < z < \zeta_{n,j}^+(t)\}} \varpi^{(j)}(dt \times X \times dz),$$

where the thinning is taken in the layer between

$$\boldsymbol{\zeta}_n^-(t) := \min\big\{\boldsymbol{\lambda}_{K_n}(t), \boldsymbol{\lambda}_{N_n}(t)\big\}, \qquad \boldsymbol{\zeta}_n^+(t) := \max\big\{\boldsymbol{\lambda}_{K_n}(t), \boldsymbol{\lambda}_{N_n}(t)\big\}.$$

Hence, by Theorem 6.37, the intensity process of L_n is

$$\boldsymbol{\lambda}_{L_n}(t) = \boldsymbol{\zeta}_n^+(t) - \boldsymbol{\zeta}_n^-(t) = \big|\boldsymbol{\lambda}_{K_n}(t) - \boldsymbol{\lambda}_{N_n}(t)\big|$$
$$= \big|\boldsymbol{\eta}_K(t) + \boldsymbol{\beta}(t|K_{n-1}) - \boldsymbol{\eta}_N(t) - \boldsymbol{\beta}(t|N_{n-1})\big|.$$

(*iii*) *Moment measure of difference process.* For the first order moment mea-

sure density $\boldsymbol{m}_{L_n}$ of L_n one has

$$\begin{aligned}\boldsymbol{m}_{L_n}(t) &= \mathbb{E}\Big[\big|\boldsymbol{\eta}_K(t) + \boldsymbol{\beta}(t|K_{n-1}) - \boldsymbol{\eta}_N(t) - \boldsymbol{\beta}(t|N_{n-1})\big|\Big] \\ &\leq \mathbb{E}\Big[\big|\boldsymbol{\eta}_K(t) - \boldsymbol{\eta}_N(t)\big|\Big] + \mathbb{E}\Big[\big|\boldsymbol{\beta}(t|K_{n-1}) - \boldsymbol{\beta}(t|N_{n-1})\big|\Big] \\ &\leq \boldsymbol{g}(t) + \mathbb{E}\Big[\boldsymbol{\beta}\big(t\big||K_{n-1} - N_{n-1}|\big)\Big] = \boldsymbol{g}(t) + \mathbb{E}\big[\boldsymbol{\beta}(t|L_{n-1})\big].\end{aligned}$$

Due to Equation (6.26), one has $\mathbb{E}\big[\boldsymbol{\beta}(t|L_{n-1})\big] = [\mathbb{H} * \boldsymbol{m}_{L_{n-1}}](t)$, and therefore

$$\boldsymbol{m}_{L_n}(t) \leq \boldsymbol{g}(t) + \big[\mathbb{H} * \boldsymbol{m}_{L_{n-1}}\big](t) \leq \sum_{m=0}^{n-1} \big[\mathbb{H}^{*m} * \boldsymbol{g}\big](t) + \big[\mathbb{H}^{*n} * \boldsymbol{m}_{L_0}\big](t),$$

and for $\boldsymbol{m}_{L_0}$ one obtains

$$\boldsymbol{m}_{L_0}(t) = \mathbb{E}\Big[\big|\boldsymbol{\lambda}_{K_0}(t) - \boldsymbol{\lambda}_{N_0}(t)\big|\Big] \leq \big|\boldsymbol{\eta}_K(t) - \boldsymbol{\eta}_N(t)\big| = \boldsymbol{g}(t).$$

Combining these inequalities, one gets

$$\boldsymbol{m}_{L_n}(t) \leq \sum_{m=0}^{n} \big[\mathbb{H}^{*m} * \boldsymbol{g}\big](t) \leq \Big[\sum_{m=0}^{\infty} \mathbb{H}^{*m} * \boldsymbol{g}\Big](t) = \big[\mathbb{U}^{+} * \boldsymbol{g}\big](t).$$

Note that $\boldsymbol{\lambda}_{K_n} \leq \boldsymbol{\lambda}_{K_\infty}$ and $\boldsymbol{\lambda}_{N_n} \leq \boldsymbol{\lambda}_{N_\infty}$. Hence, using bounded convergence, the statement follows with

$$\begin{aligned}\boldsymbol{m}_{L_\infty}(t) &= \mathbb{E}\Big[\big|\boldsymbol{\lambda}_{K_\infty}(t) - \boldsymbol{\lambda}_{N_\infty}(t)\big|\Big] = \mathbb{E}\Big[\Big|\lim_{n\to\infty} \boldsymbol{\lambda}_{K_n}(t) - \lim_{n\to\infty} \boldsymbol{\lambda}_{N_n}(t)\Big|\Big] \\ &= \mathbb{E}\Big[\lim_{n\to\infty} \big|\boldsymbol{\lambda}_{K_n}(t) - \boldsymbol{\lambda}_{N_n}(t)\big|\Big] = \lim_{n\to\infty} \mathbb{E}\Big[\big|\boldsymbol{\lambda}_{K_n}(t) - \boldsymbol{\lambda}_{N_n}(t)\big|\Big] \\ &= \lim_{n\to\infty} \boldsymbol{m}_{L_n}(t) \leq \big[\mathbb{U}^{+} * \boldsymbol{g}\big](t). \qquad \square\end{aligned}$$

Proof (Theorem 6.49): Let $\boldsymbol{\alpha}(t|\nu) = \boldsymbol{\eta}(t) + \boldsymbol{\beta}(t|\nu)$ be a Hawkes intensity function and N_∞ be the associated minimal solution as given in Definition 6.46.

(1) *Existence of minimal solution.* According to Equation (6.16), N_∞ satisfies the thinning condition. It remains to show that N_∞ is a locally-finite point process. Now, due to Equation (6.15), the density of the first order

moment measure $\boldsymbol{m}_\infty$ of N_∞ satisfies:

$$\begin{aligned} m_{\infty,j}(t) = [\mathbb{U}^+ * \boldsymbol{\eta}]_j(t) &= \sum_{k=1}^{d} \int_{\mathbb{R}} \eta_k(t-s) U^+_{jk}(ds) \\ &\leq \sum_{k=1}^{d} \big[\sup_{u\in\mathbb{R}} \eta_k(u)\big] \int_{\mathbb{R}} U^+_{jk}(ds) = \sum_{k=1}^{d} \big[\sup_{u\in\mathbb{R}} \eta_k(u)\big] U^+_{jk}(\mathbb{R}) \\ &= \sum_{k=1}^{d} (\mathbb{1}_d - Q)^{-1}_{jk} \big[\sup_{u\in\mathbb{R}} \eta_k(u)\big]. \end{aligned}$$

By assumption, the immigration intensity $\boldsymbol{\eta}$ is bounded. Hence, one obtains in vector notation:

$$\boldsymbol{m}_\infty(t) \leq (\mathbb{1}_d - Q)^{-1} \Big[\sup_{u\in\mathbb{R}} \boldsymbol{\eta}(u)\Big].$$

But this is more than sufficient for N to be locally-finite. Indeed, we need to show that N has only finitely many events in any bounded set $E \subseteq \mathbb{R}$ with probability one. Without loss of generality, we can take an interval of the form $E = [a, b]$. Then, for all $X \subseteq \mathbb{X}_j$:

$$\begin{aligned} \mathbb{E}\big[N^{(j)}_\infty(E \times X)\big] &\leq \mathbb{E}\big[N^{(j)}_\infty(E \times \mathbb{X}_j)\big] = \int_E m_{\infty,j}(t)dt = \int_a^b m_{\infty,j}(t)dt \\ &\leq (b-a)\Big[(\mathbb{1}_d - Q)^{-1} \sup_{u\in\mathbb{R}} \boldsymbol{\eta}(u)\Big]_j < \infty. \end{aligned}$$

Hence, ${N_\infty}^{(j)}(E \times X)$ is finite, and therefore N is a locally-finite point process, as claimed.

(2) *Coupling of two minimal solutions.* Consider the difference process $L_\infty := |K_\infty - N_\infty|$. Because K_∞ and N_∞ are both strong solutions, one can apply Proposition 6.43. Hence, due to Equation (6.12):

$$\mathbb{P}[T \leq t] \geq \exp\Big\{-\sum_{j=1}^{d} \int_t^\infty g_j(u) + \big[\mathbb{H} * \mathbb{1}_{(-\infty,t]} \boldsymbol{m}_{L_\infty}\big]_j(u)du\Big\},$$

where $\boldsymbol{g}(t) := |\boldsymbol{\eta}_K(t) - \boldsymbol{\eta}_N(t)|$. On the other hand, $\boldsymbol{m}_{L_\infty}$ has already been calculated in Equation (6.17). Therefore

$$\boldsymbol{m}_{L_\infty}(t) \leq \Big[\mathbb{U}^+ * |\boldsymbol{\eta}_K - \boldsymbol{\eta}_N|\Big](t) = [\mathbb{U}^+ * \boldsymbol{g}](t).$$

Together with Equations (4.3) and (4.4), one now obtains:

$$\begin{aligned}\boldsymbol{g}(u) + \Big[\mathbb{H} * \mathbb{1}_{(-\infty,t]}\boldsymbol{m}_{L_\infty}\Big](u) &\leq \boldsymbol{g}(u) + \big[\mathbb{H} * \boldsymbol{m}_{L_\infty}\big](u)\\ &\leq \boldsymbol{g}(u) + \big[\mathbb{H} * \mathbb{U}^+ * \boldsymbol{g}\big](u) = \boldsymbol{g}(u) + \big[\mathbb{U}^- * \boldsymbol{g}\big](u)\\ &= \big[(\mathbb{1} + \mathbb{U}^-) * \boldsymbol{g}\big](u) = \big[\mathbb{U}^+ * \boldsymbol{g}\big](u) = \Big[\mathbb{U}^+ * |\boldsymbol{\eta}_K - \boldsymbol{\eta}_N|\Big](u).\end{aligned}$$

Substitute this in the integral above, and the claimed formula follows. □

Proof (Proposition 6.50): To avoid lengthy expressions, we express thinnings with the thinning operator Θ_ϖ, see the definition in Equation (6.8).

(1) *Equivalence relation between continuations.* Recall that $(\nu_-, \boldsymbol{\eta}_+, N_+)$ is a continuation triple if and only if for all sets $E_+ \subseteq (\chi, \infty)$ the following is satisfied:

$$N(E_+) = \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|N) + \boldsymbol{\eta}_+(\cdot)\Big](E_+), \qquad\qquad N := \nu_- + N_+.$$

Conversely, $(\varnothing, \hat{\boldsymbol{\eta}}_+, N_+)$ is a continuation triple if and only if for all E_+:

$$N_+(E_+) = \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|N_+) + \hat{\boldsymbol{\eta}}_+(\cdot)\Big](E_+).$$

(⇒) We need to show that if $(\nu_-, \boldsymbol{\eta}_+, N_+)$ is a continuation triple, then so is $(\varnothing, \hat{\boldsymbol{\eta}}_+, N_+)$. But, because $\hat{N}$ and N_+ coincide on (χ, ∞), one has for all $E_+ \subseteq (\chi, \infty)$ that

$$\begin{aligned}N_+(E_+) = N(E_+) &= \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|N) + \boldsymbol{\eta}_+(\cdot)\Big](E_+)\\ &= \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|\nu_-) + \boldsymbol{\beta}(\cdot\,|N_+) + \boldsymbol{\eta}_+(\cdot)\Big](E_+)\\ &= \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|N_+) + \hat{\boldsymbol{\eta}}_+(\cdot)\Big](E_+).\end{aligned}$$

(⇐) Conversely, assume $(\varnothing, \hat{\boldsymbol{\eta}}_+, N_+)$ is a continuation triple and we need to show $(\nu_-, \boldsymbol{\eta}_+, N_+)$ is one too. Because N and N_+ coincide on (χ, ∞), one has

$$\begin{aligned}N(E_+) = N_+(E_+) &= \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|N_+) + \hat{\boldsymbol{\eta}}_+(\cdot)\Big](E_+)\\ &= \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|N_+) + \boldsymbol{\beta}(\cdot\,|\nu_-) + \boldsymbol{\eta}_+(\cdot)\Big](E_+)\\ &= \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|N_+ + \nu_-) + \boldsymbol{\eta}_+(\cdot)\Big](E_+) = \Theta_\varpi\Big[\boldsymbol{\beta}(\cdot\,|N) + \boldsymbol{\eta}_+(\cdot)\Big](E_+).\end{aligned}$$

In the last equality, $N_+ + \nu_- = N$ was used.

(2) *Relation to strong solutions.* Recall that a pair $(\boldsymbol{\eta}, N_+)$ is a strong solution and $(\varnothing_-, \boldsymbol{\eta}_+, N_+)$ is a continuation triple if and only if the following two corresponding equations are satisfied:

$$N_+(E) = \Theta_\varpi\big[\boldsymbol{\alpha}(\cdot|N_+)\big](E), \quad N_+(E_+) = \Theta_\varpi\big[\boldsymbol{\alpha}(\cdot|N_+)\big](E_+), \quad (6.31)$$

for all sets $E \subseteq \mathbb{R}$ and $E_+ \subseteq (\chi, \infty)$. Because $\boldsymbol{\beta}$ is $\mathfrak{P}_{\mathbb{X}}(\mathbb{R})$-predictable, $\boldsymbol{\beta}(t|\nu) = \boldsymbol{\beta}(t|\nu|_{(-\infty,t)})$, for all point configurations ν. And since for all $t \leq \chi$ one has $N_+|_{(-\infty,t)} = \varnothing$, and $\boldsymbol{\eta}(t) = \mathbf{0}$, one obtains

$$\boldsymbol{\alpha}(t|N_+) = \boldsymbol{\beta}(t|N_+) + \boldsymbol{\eta}(t) = \boldsymbol{\beta}(t|N_+) = \boldsymbol{\beta}(t|N_+|_{(-\infty,t)}) = \boldsymbol{\beta}(t|\varnothing) = \mathbf{0}.$$

Therefore, for all sets $E_- \subseteq (-\infty, \chi]$:

$$\Theta_\varpi\big[\boldsymbol{\alpha}(\cdot|N_+)\big](E_-) = \Theta_\varpi[\mathbf{0}](E_-) = 0. \quad (6.32)$$

($\Rightarrow$) We need to show the first part of Equation (6.31), given that the second part is satisfied. It suffices to show this for sets $E_- \subseteq (-\infty, \chi]$ and $E_+ \subseteq (\chi, \infty)$ separately: Due to Equation (6.32), one has for sets $E_- \subseteq (-\infty, \chi]$ that

$$N_+(E_-) = \varnothing_-(E_-) = 0 = \Theta_\varpi\big[\boldsymbol{\alpha}(\cdot|N_+)\big](E_-).$$

And for sets $E_+ \subseteq (\chi, \infty)$ the equation is satisfied by assumption.

($\Leftarrow$) This follows immediately, since the first part of Equation (6.31) implies the second part. □

Proof (Corollary 6.51): Let $\hat{\boldsymbol{\eta}}_+$ be the augmented immigration intensity and $\hat{\boldsymbol{\eta}}$ the extension of $\hat{\boldsymbol{\eta}}_+$ to the real line.

(1) *Continuation and strong solution.* Combining the two equivalence relations from Equations (6.19) and (6.20), one has

$$\begin{aligned} (\nu_-, \boldsymbol{\eta}_+, N_+) \text{ continuation} \quad &\Leftrightarrow \quad (\varnothing_-, \hat{\boldsymbol{\eta}}_+, N_+) \text{ continuation} \\ &\Leftrightarrow \quad (\hat{\boldsymbol{\eta}}, N_+) \text{ strong solution.} \end{aligned}$$

And this is exactly the equivalence what needs to be shown.

(2) *Uniqueness of continuation.* Assume $(\nu_-, \boldsymbol{\eta}_+, N_+)$ and $(\nu_-, \boldsymbol{\eta}_+, \tilde{N}_+)$ are two continuation triples. Then, according to the first part, the following two pairs define strong solutions:

$$(\hat{\boldsymbol{\eta}}, N_+) \text{ strong solution,} \qquad (\hat{\boldsymbol{\eta}}, \tilde{N}_+) \text{ strong solution.}$$

Recall the coupling result for strong solutions given in Equation (6.12). Applying this result for $t := \chi$ gives

$$\mathbb{P}[T \leq \chi] \geq \exp\Big\{-\sum_{j=1}^{d} \int_{\chi}^{\infty} \big|\hat{\eta}_j(u) - \hat{\eta}_j(u)\big| + \big[\mathbb{H} * \mathbb{1}_{(-\infty,\chi]} \boldsymbol{m}_L\big]_j(u) du\Big\}.$$

Clearly, the first term in the integrand vanishes. It remains to calculate the function $\mathbb{1}_{(-\infty,\chi]} \boldsymbol{m}_L$. But $\hat{K}$ and $\hat{N}$ are by assumption continuations of the same initial state ν_-, i.e. they coincide on the time interval $(-\infty, \chi]$, so that

$$K|_{(-\infty,\chi]} = N|_{(-\infty,\chi]} = \nu_- \qquad \text{and} \qquad L|_{(-\infty,\chi]} = \varnothing_-.$$

This implies

$$\mathbb{1}_{(-\infty,\chi]}(t) \boldsymbol{m}_L(t) = \mathbf{0}, \qquad \text{and then} \qquad \big[\mathbb{H} * \mathbb{1}_{(-\infty,\chi]} \boldsymbol{m}_L\big]_j(u) = 0.$$

After substitution, this yields $\mathbb{P}[T \leq \chi] = 1$. Hence, the two processes K and N coincide on the interval (χ, ∞). Therefore, K_+ and N_+ coincide too, since they are the restrictions of K and N to this interval. □

Proof (Proposition 6.52): Recall the sufficient conditions for the existence of a minimal solution, see Theorem 6.49.

(1) *Construction of continuation.* Let N_∞ be the minimal solution generated by $\hat{\boldsymbol{\eta}}$. Since every minimal solution is a strong solution, the equivalence in Equation (6.22) applies. Hence, $(\nu_-, \boldsymbol{\eta}_+, N_{\infty,+})$ is a continuation triple, as claimed. The uniqueness of $N_{\infty,+}$ follows from the second part of Corollary 6.51.

(2) *Coupling with two initial conditions.* Let K_∞ and N_∞ be the minimal solutions corresponding to the immigration intensities $\hat{\boldsymbol{\eta}}_\kappa$ and $\hat{\boldsymbol{\eta}}_\nu$. For

minimal solutions, Equation (6.18) applies. Therefore

$$\mathbb{P}[T \le t] \ge \exp\Big\{-\sum_{j=1}^{d}\int_t^\infty \Big[\mathbb{U}^+ * |\hat{\boldsymbol{\eta}}_\kappa - \hat{\boldsymbol{\eta}}_\nu|\Big]_j(u)du\Big\}.$$

Moreover, one has

$$\begin{aligned}\big|\hat{\boldsymbol{\eta}}_\kappa(t) - \hat{\boldsymbol{\eta}}_\nu(t)\big| &= \mathbb{1}_{(\chi,\infty)}(t)\big|\boldsymbol{\eta}_+(t) + \boldsymbol{\beta}(t|\kappa_-) - \boldsymbol{\eta}_+(t) - \boldsymbol{\beta}(t|\nu_-)\big| \\ &= \mathbb{1}_{(\chi,\infty)}(t)\big|\boldsymbol{\beta}(t|\kappa_-) - \boldsymbol{\beta}(t|\nu_-)\big|.\end{aligned}$$

Because $K_{\infty,+}$ and $N_{\infty,+}$ are the restrictions of K_∞ and N_∞ to the interval (χ,∞), the same bound also holds for the coupling time between these two processes. □

6.57 Lemma (Conditional expectations on splitted canonical space). Let $\chi \in \mathbb{R}$ be a fixed time. Consider the univariate canonical space $\mathfrak{N} := \mathscr{N}_{\mathbb{R}_+}(\mathbb{R})$ as explained in Definition 6.53. Let $\mathbb{E}_-$ and $\mathbb{E}_+$ be the expectations with respect to the probability measures $\mathbb{P}_-$ and $\mathbb{P}_+$ on the spaces $\mathfrak{N}_-$ and $\mathfrak{N}_+$, respectively. For a non-negative random variable f one has:

$$\mathbb{E}\big[f(\varpi)\big|\mathfrak{F}_\chi\big] = \mathbb{E}_+\big[f(\varpi_-,\varpi_+)\big]. \tag{6.33}$$

Note that the expectation $\mathbb{E}_+$ is defined with respect to $\omega_+ \in \mathfrak{N}_+$ only and does not affect $\omega_- \in \mathfrak{N}_-$. ♢

Proof (Lemma 6.57): Because ϖ_- is ϖ restricted to $(-\infty,\chi]$, one has $\mathfrak{F}_\chi = \mathscr{F}^{\varpi_-}$, where $\mathscr{F}^{\varpi_-}$ is the σ-algebra generated by ϖ_-. Hence

$$\mathbb{E}\big[f(\varpi)\big|\mathfrak{F}_\chi\big] = \mathbb{E}\big[f(\varpi_-,\varpi_+)\big|\mathscr{F}^{\varpi_-}\big] = \mathbb{E}\big[f(\varpi_-,\varpi_+)\big|\varpi_-\big].$$

Now use the fact that ϖ_- and ϖ_+ are independent to obtain

$$\mathbb{E}\big[f(\varpi_-,\varpi_+)\big|\varpi_-\big] = \mathbb{E}_+\big[f(\varpi_-,\varpi_+)\big].$$ □

Proof (Theorem 6.55): Assume K and N are two strong solutions with bounded first moment measure densities $\boldsymbol{m}_K$ and $\boldsymbol{m}_N$. Recall the decomposition of the canonical probability space, given in Definition 6.53.

(*i*) Define the restrictions of K and N to the time intervals $(-\infty,\chi]$ and

(χ, ∞) by:

$$K_- := K|_{(-\infty,\chi]}, \qquad N_- := N|_{(-\infty,\chi]},$$
$$K_+ := K|_{(\chi,\infty)}, \qquad N_+ := N|_{(\chi,\infty)}.$$

Note that K_- and N_- are both $\mathfrak{F}_-$-measurable and therefore depend only on $\varpi_- \in \mathfrak{N}_-$.

(ii) Below, we will be calculating on the probability space $(\mathfrak{N}_+, \mathfrak{F}_+, \mathbb{P}_+)$. This situation occurs if $\varpi_- \in \mathfrak{N}_-$ is fixed, since then only the second component of $\varpi \equiv (\varpi_-, \varpi_+)$ is random.

Hence fix some $\varpi_- \in \mathfrak{N}_-$, so that the probability space is reduced to the triple $(\mathfrak{N}_+, \mathfrak{F}_+, \mathbb{P}_+)$. Then K_-, N_- are two deterministic point configurations on the time interval $(-\infty, \chi]$. Moreover, K_+, N_+ are the continuations of K_-, N_-, in the sense of Definition 6.35. In other words, one has the two continuation triples:

$$(K_-, \boldsymbol{\eta}_+, K_+) \qquad \text{and} \qquad (N_-, \boldsymbol{\eta}_+, N_+).$$

According to Corollary 6.51, continuations are uniquely determined by their initial state. Moreover, by Proposition 6.52, a continuation is the same as a minimal solution if one takes the augmented immigration intensity. Denote the corresponding minimal solutions by $K_{\infty,+}$, $N_{\infty,+}$, see also Equation (6.23). As a consequence, the two continuation triples above coincide with

$$(K_-, \boldsymbol{\eta}_+, K_{\infty,+}) \qquad \text{and} \qquad (N_-, \boldsymbol{\eta}_+, N_{\infty,+}). \tag{6.34}$$

(iii) Consider now again the full probability space $(\mathfrak{N}, \mathfrak{F}, \mathbb{P})$. Recall from Equation (6.11) that the truncated coupling time T_χ is defined as

$$T_\chi := \inf\Big\{ t \geq \chi : K|_{(t,\infty)} = N|_{(t,\infty)} \Big\}.$$

Fix a time $t \geq \chi$. Clearly, if K and N couple before time t then so do K_+ and N_+, and vice versa. Hence, due to Equation (6.33), one has:

$$\mathbb{P}[T \leq t] = \mathbb{P}[T_\chi \leq t] = \mathbb{E}\Big[\mathbb{P}\big[T_\chi \leq t \big| \mathfrak{F}_\chi\big]\Big] = \mathbb{E}\Big[\mathbb{P}_+\big[T_\chi \leq t\big]\Big]. \tag{6.35}$$

To calculate $\mathbb{P}_+[T_\chi \leq t]$, recall from step (ii), that one can consider ϖ_- as fixed. Hence, we are dealing with the two continuation triples from

Equation (6.34). Therefore, according to Equation (6.24), one has:

$$\mathbb{P}_+\big[T_\chi \le t\big] \ge \exp\Big\{-\sum_{j=1}^{d}\int_t^\infty \big[\mathbb{U}^+ * \boldsymbol{g}_\chi\big]_j(s)ds\Big\},$$

where $\boldsymbol{g}_\chi$ depends on K_- and N_- and is defined as

$$\boldsymbol{g}_\chi(t) := \mathbb{1}_{(\chi,\infty)}(t)\Big|\boldsymbol{\beta}(t|K_-) - \boldsymbol{\beta}(t|N_-)\Big|,$$

for $t \in \mathbb{R}$. After substitution of $\mathbb{P}_+\big[T_\chi \le t\big]$ into Equation (6.35), and by Jensen's inequality, one obtains:

$$\begin{aligned}\mathbb{P}[T \le t] &\ge \mathbb{E}\Big[\exp\Big\{-\sum_{j=1}^{d}\int_t^\infty [\mathbb{U}^+ * \boldsymbol{g}_\chi]_j(s)ds\Big\}\Big] \\ &\ge \exp\Big\{-\sum_{j=1}^{d}\int_t^\infty \mathbb{E}\big[\mathbb{U}^+ * \boldsymbol{g}_\chi\big]_j(s)ds\Big\}. \qquad (6.36)\end{aligned}$$

By Fubini's theorem:

$$\begin{aligned}\int_t^\infty \mathbb{E}\big[\mathbb{U}^+ * \boldsymbol{g}_\chi\big]_j(s)ds &= \int_t^\infty \mathbb{E}\Big[\sum_{k=1}^{d}\int_0^\infty U_{jk}^+(u)g_{\chi,k}(s-u)du\Big]ds \\ &= \sum_{k=1}^{d}\int_t^\infty \Big[\int_0^\infty U_{jk}^+(u)\,\mathbb{E}\big[g_{\chi,k}(s-u)\big]du\Big]ds \\ &= \sum_{k=1}^{d}\int_{t-\chi}^\infty \Big[\int_0^\infty U_{jk}^+(u)\,\mathbb{E}\big[g_{\chi,k}(\chi+s-u)\big]du\Big]ds. \qquad (6.37)\end{aligned}$$

In the last equality, $s \twoheadrightarrow \chi+s$ was substituted. According to the definition of $\boldsymbol{g}_\chi(t)$, one has, in vector notation:

$$\mathbb{E}\big[\boldsymbol{g}_\chi(t)\big] = \mathbb{E}\Big[1_{(\chi,\infty)}(t)\big|\boldsymbol{\beta}(t|K_-) - \boldsymbol{\beta}(t|N_-)\big|\Big].$$

Let $\boldsymbol{m}_\chi$ be the first moment density of $|K_- - N_-|$. Then by Equa-

tion (6.25) and (6.26), one obtains:

$$\begin{aligned}\mathbb{E}\big[\boldsymbol{g}_\chi(t)\big] &= \mathbb{1}_{(\chi,\infty)}(t)\,\mathbb{E}\Big[\big|\boldsymbol{\beta}(t|K_-) - \boldsymbol{\beta}(t|N_-)\big|\Big] \\ &\leq \mathbb{1}_{(\chi,\infty)}(t)\,\mathbb{E}\Big[\boldsymbol{\beta}\big(t\big||K_- - N_-|\big)\Big] \leq \mathbb{1}_{(\chi,\infty)}(t)\big[\mathbb{H} * \boldsymbol{m}_\chi\big](t).\end{aligned}$$

Since $K_- = K|_{(-\infty,\chi]}$ and $N_- = N|_{(-\infty,\chi]}$, this yields:

$$\begin{aligned}\boldsymbol{m}_\chi(t) &= \mathbb{1}_{(-\infty,\chi]}(t)\boldsymbol{m}_{|K_- - N_-|}(t) \leq \mathbb{1}_{(-\infty,\chi]}(t)\boldsymbol{m}_{K+N}(t) \\ &= \mathbb{1}_{(-\infty,\chi]}(t)\Big[\boldsymbol{m}_K(t) + \boldsymbol{m}_N(t)\Big] \leq \mathbb{1}_{(-\infty,\chi]}(t)\bar{\boldsymbol{m}},\end{aligned}$$

where the vector-valued constant $\bar{\boldsymbol{m}}$ is defined as:

$$\bar{\boldsymbol{m}} := \sup_{u\in\mathbb{R}} \boldsymbol{m}_K(u) + \sup_{u\in\mathbb{R}} \boldsymbol{m}_N(u).$$

By assumption, K and N have bounded first moment measures $\boldsymbol{m}_K$ and $\boldsymbol{m}_N$, so that $\bar{\boldsymbol{m}} < \boldsymbol{\infty}$. Combine the previous results to obtain:

$$\mathbb{E}\big[\boldsymbol{g}_\chi(t)\big] \leq \mathbb{1}_{(\chi,\infty)}\Big[\mathbb{H} * \mathbb{1}_{(-\infty,\chi]}\bar{\boldsymbol{m}}\Big](t).$$

(iv) In this step, we calculate a universal bound for $\mathbb{E}\big[\boldsymbol{g}_\chi(t)\big]$, which holds uniformly for all χ. The candidate function is:

$$\psi_j(h) := \mathbb{1}_{(0,\infty)}(h)\sum_{k=1}^{d}\bar{m}_k\int_h^\infty H_{jk}(u)du, \text{ for } h > 0.$$

We claim that $\mathbb{E}\big[\boldsymbol{g}_\chi(t)\big] \leq \boldsymbol{\varphi}(t-\chi)$, for all $t > \chi$. Indeed:

$$\begin{aligned}\mathbb{E}\big[g_{\chi,j}(t)\big] &\leq \Big[\mathbb{H} * \mathbb{1}_{(-\infty,\chi]}\bar{\boldsymbol{m}}\Big]_j(t) = \sum_{k=1}^{d}\int_0^\infty H_{jk}(u)\mathbb{1}_{(-\infty,\chi]}(t-u)\bar{m}_k du \\ &= \sum_{k=1}^{d}\bar{m}_k\int_{\mathbb{R}}\mathbb{1}_{[t-\chi,\infty)}(u)H_{jk}(u)du = \psi_j(t-\chi).\end{aligned}$$

(v) Continuing with Equation (6.37), one finds:

$$\int_t^\infty \mathbb{E}\big[\mathbb{U}^+ * \boldsymbol{g}_\chi\big]_j(s)ds \le \sum_{k=1}^d \int_{t-\chi}^\infty \Big[\int_0^\infty U_{jk}^+(u)\,\mathbb{E}\big[g_{\chi,k}(\chi+s-u)\big]du\Big]ds$$
$$= \int_{t-\chi}^\infty \Big[\sum_{k=1}^d \int_0^\infty U_{jk}^+(u)\psi_k(s-u)du\Big]ds \int_{t-\chi}^\infty [\mathbb{U}^+ * \boldsymbol{\psi}]_j(s)ds.$$

We have used that $\mathbb{E}\big[\boldsymbol{g}_\chi(\chi+s-u)\big] \le \boldsymbol{\psi}(s-u)$. Hence, due to Equation (6.36), one has for all $t \ge \chi$ that:

$$\mathbb{P}[T \le t] \ge \exp\Big\{-\sum_{j=1}^d \int_{t-\chi}^\infty [\mathbb{U}^+ * \boldsymbol{\psi}]_j(s)ds\Big\}. \tag{6.38}$$

(vi) Next calculate:

$$\int_0^\infty \psi_k(h)dh = \int_0^\infty \Big[\sum_{l=1}^d \bar{m}_l \int_h^\infty H_{kl}(u)du\Big]dh$$
$$= \sum_{l=1}^d \bar{m}_l \int_0^\infty \Big[\int_h^\infty H_{kl}(u)du\Big]dh = \sum_{l=1}^d \bar{m}_l \int_0^\infty \Big[\int_0^u H_{kl}(u)dh\Big]du$$
$$= \sum_{l=1}^d \bar{m}_l \int_0^\infty uH_{kl}(u)du.$$

Recall that for any two non-negative functions f, g on $\mathbb{R}_+$:

$$\int_0^\infty (f*g)(s)ds = \Big[\int_0^\infty f(w)dw\Big]\Big[\int_0^\infty g(u)du\Big].$$

Therefore:

$$\int_0^\infty [\mathbb{U}^+ * \boldsymbol{\psi}]_j(s)ds \le \sum_{k=1}^d \Big[\int_0^\infty U_{jk}^+(h)dh\Big]\Big[\int_0^\infty \psi_k(h)dh\Big]$$
$$= \sum_{k=1}^d \Big[(\mathbb{1}_d - Q)^{-1}_{jk} \sum_{l=1}^d \bar{m}_l \int_0^\infty uH_{kl}(u)du\Big] < \infty.$$

We have used that $\bar{\boldsymbol{m}}$ and $\int_0^\infty uH_{kl}(u)du$ are finite, according to the assumptions.

(*vii*) Finally, take the limit in Equation (6.38):

$$\begin{aligned}\mathbb{P}[T \leq t] &\geq \lim_{\chi \to -\infty} \exp\Big\{ -\sum_{j=1}^{d} \int_{t-\chi}^{\infty} [\mathbb{U}^{+} * \boldsymbol{\psi}]_j(s) ds \Big\} \\ &= \exp\Big\{ -\sum_{j=1}^{d} \lim_{h \to \infty} \int_{h}^{\infty} [\mathbb{U}^{+} * \boldsymbol{\psi}]_j(s) ds \Big\} = 1.\end{aligned}$$

This shows that $T = -\infty$, almost surely. Therefore, K and N are identical, except possibly on a set of probability zero. □

Conclusion

This thesis presents linear Hawkes processes in the style of a textbook giving more details than would be common in an academic article. This makes the exposition self-contained and hopefully more accessible. It introduces and describes in detail some of the techniques used to analyze Hawkes processes. Often, these techniques are presented in a more abstract form than is strictly necessary, without specifically referring to Hawkes processes. I hope this exhibits their power more clearly.

Software for Hawkes Processes. Even though the material in the thesis is largely theoretical, I constantly had the applicability of the theory in mind when I wrote the thesis. This resulted in a software package I developed in parallel to the theoretical work. The software is not part of my thesis, but it has still contributed indirectly since implementing it has helped me to better understand some aspects of Hawkes processes I would probably not have otherwise.

Implementing estimation and simulation procedures in the generality described in the first chapter is much more involved than it would seem at first. The introduction discusses some of the problems one encounters and points out some remedies, but there are several more issues one has to deal with, which I could not include in this thesis.

This thesis is exclusively concerned with linear Hawkes processes, but the software I wrote can also cope with more general Hawkes processes, which are not discussed here. It seems that certain non-linear extensions can be very beneficial if fitted to financial data.

Further Research. Hawkes processes are a large class of point processes. Indeed, one could consider Hawkes processes more as a framework than a class of point processes. The only common feature is their self-exciting behavior but, apart from that, the other characteristics can be chosen quite freely. Thus,

if one needs to model a phenomenon that exhibits some sort of self-exciting behavior, Hawkes processes are certainly a serious candidate model.

The theoretical foundation of Hawkes processes is certainly father developed than empirical experience in using them. Hopefully there is more progress in this direction in the future. On the theoretical side, time series models inspired by Hawkes processes could be an interesting addition. The paper [BM96] introduces a class of non-linear extensions to the classical Hawkes process. But also other types of non-linear extensions could be conceived and may have interesting features. As long as one is careful not to generalize linear Hawkes processes too much, the techniques discussed in this thesis can most likely be carried over and adapted in order to prove analogous results.

Bibliography

[AA06] H. Albrecher and S. Asmussen. Ruin probabilities and aggregrate claims distributions for shot noise Cox processes. *Scandinavian Actuarial Journal*, 2:86–110, 2006.

[Ada75] L. Adamopoulos. Some counting and interval properties of the mutually-exciting processes. *Journal of Applied Probability*, 12:78–86, 1975.

[Ada76] L. Adamopoulos. Cluster models for earthquakes: Regional comparisons. *Mathematical Geology*, 8(4):1573–8868, 1976.

[AS64] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, 1964.

[Bel43] R. Bellman. The stability of solutions of linear differential equations. *Duke Mathematical Journal*, 10:643–647, 1943.

[BK02] A. Brix and W. S. Kendall. Simulation of cluster point processes without edge effects. *Advances in Applied Probability*, 34(2):267–280, 2002.

[BM94] P. Brémaud and L. Massoulié. Imbedded construction of stationary sequences and point processes with a random memory. *Queueing Systems*, 17:213–234, 1994.

[BM96] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3):1563–1588, 1996.

[BM01] P. Brémaud and L. Massoulié. Hawkes branching point processes without ancestors. *Journal of Applied Probability*, 38(1):122–135, 2001.

[BM02] P. Brémaud and L. Massoulié. Power spectra of general shot noises and Hawkes point processes with a random excitation. *Advances in Applied Probability*, 34(1):205–222, 2002.

[BNT02] P. Brémaud, G. Nappo, and G. L. Torrisi. Rate of convergence to equilibrium of marked Hawkes processes. *Journal of Applied Probability*, 39(1):123–136, 2002.

[Bou63] N. Bourbaki. *Éléments de Mathématique: Intégration, Chapitres 7 et 8*. Springer, 1963.

[Bow07] C. G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.

[Boy71] A. V. Boyd. Formal power series and the total progeny in a branching process. *Journal of Mathematical Analysis and Applications*, 34(3):565–566, 1971.

[Bré81] P. Brémaud. *Point Processes and Queues: Martingale Dynamics*. Springer, 1981.

[CDDM05] V. Chavez-Demoulin, A. C. Davison, and A. J. McNeil. Estimating Value-at-Risk: A point process approach. *Quantitative Finance*, 5(2):227–234, 2005.

[CF06] P. C. Consul and F. Famoye. *Lagrangian Probability Distributions*. Birkhäuser, 2006.

[CS72] P. C. Consul and L. R. Shenton. Use of Lagrange expansion for generating discrete generalized probability distributions. *SIAM Journal on Applied Mathematics*, 23(2):239–248, 1972.

[DPS00] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

[DVJ03] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume I*. Springer, 2nd edition, 2003.

[Dwa69] M. Dwass. The total progeny in a branching process and a related random walk. *Journal of Applied Probability*, 6:682–686, 1969.

[EGG07] E. Errais, K. Giesecke, and L. R. Goldberg. Pricing credit from the top down with affine point processes. Working paper, Stanford School of Engineering, 2007.

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications: Volume 1.* Wiley, 1968.

[FT98] S. Foss and R.L. Tweedie. Perfect simulation and backward coupling. *Stochastic Models*, 14(1-2):187–204, 1998.

[FY96] F. Forbes and B. Ycart. The philosophers' process on ladder graphs. *Communications in statistics. Stochastic models*, 12(4):559–583, 1996.

[Gan74] F. R. Gantmacher. *The Theory of Matrices*, volume 2. Chelsea, 1974.

[GG05] K. Giesecke and L. R. Goldberg. A top down approach to multi-name credit. Technical report, Stanford School of Engineering, November 2005.

[Goo49] I. J. Good. The number of individuals in a cascade process. *Mathematical Proceedings of the Cambridge Philosophical Society*, 45:360–363, 1949.

[Goo60] I. J. Good. Generalization to several variables of Lagrange's expansion, with applications to stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 56:367–80, 1960.

[Goo75] I. J. Good. The Lagrange distributions and branching processes. *SIAM Journal on Applied Mathematics*, 28(2):270–275, 1975.

[Gro19] T. H. Gronwall. Note on the derivative with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics*, 20:292–296, 1919.

[GT05] K. Giesecke and P. Tomecek. Dependent events and changes of time. Working paper, Cornell University, July 2005.

[Haw71] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[HJ85] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[HO74] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11:493–503, 1974.

[Jac75] J. Jacod. Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Probability Theory and Related Fields*, 31(3):235–253, 1975.

[Jag75] P. Jagers. *Branching Processes with Biological Applications*. John Wiley and Sons Ltd, 1975.

[Kal97] O. Kallenberg. *Foundations of Modern Probability*. Springer, 1997.

[Kel85] F. P. Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society. Series B*, 47(3):379–395, 1985.

[Ker64] J. Kerstan. Teilprozesse Poissonscher Prozesse. In Jaroslav Kožešník, editor, *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pages 377–403. Czechoslovak Academy of Science, 1964.

[KM00] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces. *Advances in Applied Probability*, 32(3):844–865, 2000.

[Kri74] K. Krickeberg. Moments of point processes. In E. F. Harding and M. G. Kendall, editors, *Stochastic Geometry*, pages 89–113. Wiley, 1974.

[KS96] A. Kwiecinski and R. Szekli. Some monotonicity and dependence properties of self-exciting point processes. *The Annals of Applied Probability*, 6(4):1211–1231, 1996.

[Las93] G. Last. On dependent marking and thinning of point processes. *Stochastic Processes and their Applications*, 45(1):73–94, 1993.

[LS79] P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

[Mas98] L. Massoulié. Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Processes and their Applications*, 75(1):1–30, 1998.

[Mey71] P.-A. Meyer. Démonstration simplifiée d'un théorème de Knight. *Séminaire de probabilités de Strasbourg*, 5:191–195, 1971.

[MFE05] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2005.

[Min88] H. Minc. *Nonnegative Matrices*. Wiley, 1988.

[MKM78] K. Matthes, J. Kerstan, and J. Mecke. *Infinitely Divisible Point Processes*. John Wiley & Sons, 2nd edition, 1978.

[MR05] J. Møller and J. G. Rasmussen. Perfect simulation of Hawkes processes. *Advances in Applied Probability*, 37:629–646, 2005.

[MR06] J. Møller and J. G. Rasmussen. Approximate simulation of Hawkes processes. *Methodology and Computing in Applied Probability*, 8:53–64, 2006.

[MVJ92] F. Musmeci and D. Vere-Jones. A space-time clustering model for historical earthquakes. *Annals of the Institute of Statistical Mathematics*, 44(1):1–11, 1992.

[NLJ05] S. Kotz N. L. Johnson, A. W. Kemp. *Univariate Discrete Distributions*. John Wiley & Sons, 2005.

[OA82] Y. Ogata and H. Akaike. On the linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes. *Journal of the Royal Statistical Society: Series B*, 44(1):102–107, 1982.

[Oak75] D. Oakes. The Markovian self-exciting process. *Journal of Applied Probability*, 12:69–77, 1975.

[OAK82] Y. Ogata, H. Akaike, and K. Katsura. The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process. *Annals of the Institute of Statistical Mathematics*, 34(2B):373–387, 1982.

[Oga78] Y. Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(A):243–261, 1978.

[Oga81] Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

[Oga88] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[Oga98] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

[Oga99] Y. Ogata. Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155(2-4):471–507, 1999.

[Ott48] R. Otter. The number of trees. *The Annals of Mathematics, Second Series*, 49(3):583–599, 1948.

[Ott49] R. Otter. The multiplicative process. *Annals of Mathematical Statistics*, 20:206–224, 1949.

[Oza79] T. Ozaki. Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.

[Pap72] F. Papangelou. Integrability of expected increments of point processes and a related change of scale. *Transactions of the American Mathematical Society*, 165:483–506, 1972.

[PW96] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1&2):223–252, 1996.

[Rub72] I. Rubin. Regular point processes and their detection. *IEEE Transactions on Information Theory*, 18(5):547–557, 1972.

[Rud91] W. Rudin. *Functional Analysis*. McGraw-Hill, 1991.

[ST08] G. Stabile and G. Torrisi. Risk processes with non-stationary Hawkes claims arrivals. *Methodology and Computing in Applied Probability*, DOI 10.1007/s11009-008-9110-6, 2008.

[Tan53] J. C. Tanner. A problem of interference between two queues. *Biometrika*, 40(1-2):58–69, 1953.

[Tor02] G. L. Torrisi. A class of interacting marked point processes: Rate of convergence to equilibrium. *Journal of Applied Probability*, 39(1):137–160, 2002.

[UOM95] T. Utsu, Y. Ogata, and R. S. Matsu'ura. The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, 43(1):1–33, 1995.

[VJ75] D. Vere-Jones. Stochastic models for earthquake sequences. *Geophysical Journal International*, 42(2):811–826, 1975.

[VJ78] D. Vere-Jones. Earthquake prediction: A statistician's view. *Journal of Physics of the Earth*, 26:129–146, 1978.

[VJ05] D. Vere-Jones. A class of self-similar random measure. *Advances in Applied Probability*, 37(4):908–914, 2005.

[VJO82] D. Vere-Jones and T. Ozaki. Some examples of statistical estimation applied to earthquake data. *Annals of the Institute of Statistical Mathematics*, B34:189–207, 1982.

[Wat64] S. Watanabe. On discontinuous additive functionals and Lévy measures of a Markov process. *Japanese Journal of Mathematics*, 34:53–70, 1964.