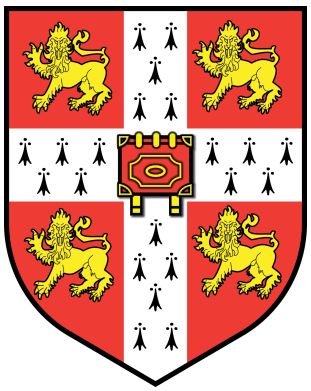
Efficient nonparametric inference for discretely observed compound Poisson processes

Alberto Jesús Coca Cabrero

Darwin College

University of Cambridge



A thesis submitted for the degree of

Doctor of Philosophy

August 2016

Abstract

Compound Poisson processes are the textbook example of pure jump stochastic processes and the building blocks of Lévy processes. They have three defining parameters: the distribution of the jumps, the intensity driving the frequency at which these occur, and the drift. They are used in numerous applications and, hence, statistical inference on them is of great interest. In particular, nonparametric estimation is increasingly popular for its generality and reduction of misspecification issues.

In many applications, the underlying process is not observed directly but at discrete times. Therefore, important information is missed between observations and we face a (non-linear) inverse problem. Using the intimate relationship between Lévy processes and infinite divisible distributions, we construct new estimators of the jump distribution and of the so-called Lévy distribution. Under mild assumptions, we prove Donsker theorems for both (i.e. functional central limit theorems with the uniform norm) and identify the limiting Gaussian processes. This allows us to conclude that our estimators are efficient, or optimal from an information theory point of view, and to give new insight into the topic of efficiency in this and related problems. We allow the jump distribution to potentially have a discrete component and include a novel way of estimating the mass function using a kernel estimator. We also construct new estimators of the intensity and of the drift, and show joint asymptotic normality of all the estimators. Many relevant inference procedures are derived, including confidence regions, goodness-of-fit tests, two-sample tests and tests for the presence of discrete and absolutely continuous jump components.

In related literature, two apparently different approaches have been taken: a natural direct approach, and the spectral approach we use. We show that these are formally equivalent and that the existing estimators are very close relatives of each other. However, those from the first approach can only be used in small compact intervals in the positive real line whilst ours work on the whole real line and, furthermore, are the first to be efficient. We describe how the former can attain efficiency and propose several open problems not yet identified in the field. We also include an exhaustive simulation study of our and other estimators in which we illustrate their behaviour in a number of realistic situations and their suitability for each of them. This type of study cannot be found in existing literature and provides several insights not yet pointed out and solid understanding of the practical side of the problem on which real-data studies can be based. The implementation of all the estimators is discussed in detail and practical recommendations are given.

A mi familia

Acknowledgements

I am all too aware that the completion of my doctoral studies would not have been possible were it not for the generous support of so many. Herein, I express my deep gratitude to all who have contributed to making this journey a unique, meaningful and life-changing experience.

Firstly, I wish to thank my doctoral supervisors —Richard Nickl and L. Christopher G. Rogers. To Richard, my principal supervisor, because he not only introduced me to the fascinating world of nonparametric statistics but his passion for the subject has enthused me and his mentoring has opened up the academic world; I am only so honoured I will continue having the opportunity to learn from his immense mathematical knowledge after the Ph.D in a Post Doc with him. I wish too to thank Chris for sharing his exceptional insights into some challenging probability open problems, as well as for his dedication and personal support.

I would like to express my gratitude to the CCA and its directors for putting their trust in me when they accepted me onto what has been a fantastic Ph.D. programme and to the CCA administrators for their continued support; I am especially thankful to CCA director James Norris for acting as internal examiner in my defence. I am similarly grateful to Markus Reiß for acting as external examiner; the defence was a truly enriching experience in which a number of exciting future lines of research arose. I am also very much indebted to Markus for his hospitality during two visits to Berlin.

I consider myself most fortunate for the rest of the people who have walked this journey with me and who have provided me with a firm personal support of incalculable value. My utmost gratitude must go to my family; especially to my parents, siblings, and grandmothers. They have loved me unconditionally throughout by being the pillar on which I could always rely to rest on and by showing boundless generosity. Most especially, to my parents, Jesús Coca Gradín and Purificación Cabrero Maroto, who have taught me the most important lessons in life and without which I would not have been able to succeed in this enterprise. A few lines herein would never suffice to thank you enough: every achievement of mine is yours.

I am most grateful to my beloved friends; unfortunately, I do not have enough space to thank them all. To Alexandra Gandini who, after my family, has been my greatest influence, both personally and professionally; I wish you and your family the very best. To my childhood friends, Jorge Dávila, Javier González-Adalid, Clara López and Robert

Stocks, who still are my main confidants and with whom my friendship only grows with time despite the distance. To my friends from Cambridge colleges, Erik van Bergen, Mathew Edwards, Douwe Kiela and Kattria van der Ploeg, who have made these years be truly outstanding, not only through memorable experiences but also through the highly stimulating academic discussions I was fortunate to have with them. To Simon Walsh, an exceptional friend and an inspirational man who keeps challenging me intellectually and who was always there in tough times. To the mathematicians, Milana Gatarić, Gil Ramos and Martin Taylor, for their sincere friendships, for the exciting mathematical discussions we have had and for the invaluable advice they have provided me with to navigate through the academic world. And to Julia López Canelada, for her immeasurable patience and generous support whilst writing this thesis, a truly demanding period that she illuminated with her everlasting happiness and radiant smile. Thank you all for giving me the opportunity to share the joy of life and knowledge with you and I can only hope to be honoured to continue doing so in the near future.

I wish also to recognise the generosity and support I received from many organisations during the course of my Ph.D. studies: Fundación "La Caixa", EPSRC, Fundación Mutua Madrileña, Access to Learning Fund, Cambridge Philosophical Society, Lundgren Fund and Darwin College.

And lastly, to luck, for all the above and much more: please, receive my most profound gratitude.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution.

Contents

List of Notation

Δ	The size of the time-increments at which a stochastic process is (discretely) observed
F, $F(\cdot)$	The distribution function of the underlying jumps of a compound Poisson process
$\widetilde{F}, \widetilde{F}(\cdot), \widehat{F}, \widehat{F}(\cdot)$	Estimators of F of Buchmann and Grübel [2003] and Coca [2015], respectively
$\mathcal{F}$, $\mathcal{F}^{-1}$	The Fourier(–Plancherel) transform and its inverse transform
γ	The drift of a compound Poisson process
$\tilde{\gamma}, \hat{\gamma}$	Estimators of γ (naive and spectral of Coca [2015], respectively)
h, h_n	The bandwidth of a kernel function
n	Size of a sample
i.i.d.	Independent and identically distributed
K	A kernel function
λ	The intensity of a compound Poisson process
$\tilde{\lambda}, \hat{\lambda}, \check{\lambda}$	Estimators of λ (naive, spectral mass of Coca [2015] and spectral origin, respectively)
$\ell^\infty(C)$	Space of real-valued uniformly bounded functions on $C \subseteq \mathbb{R}$
$L^\infty(C)$	Equivalence class of real-valued essentially bounded functions on $C \subseteq \mathbb{R}$
$L^r(C)$, $1 \leq r < \infty$	Equivalence class of real-valued, Lebesgue-measurable functions f on $C \subseteq \mathbb{R}$ under the norm $\|f\|_r := \left(\int_C |f(x)|^r dx\right)^{1/r}$
Log, log	The distinguished and the natural logarithms
$\mathcal{L}$, $\mathcal{L}^{-1}$	The Laplace transform and its inverse transform
$\max, \vee$	Maximum
$\min, \wedge$	Minimum
μ	A finite measure
$(N_t)_{t\geq 0}$	A Poisson process
$N(0, \sigma^2)$	A normal distribution with mean zero and variance σ^2

$N, N(\cdot)$	The Lévy distribution function
$\widehat{N}, \widehat{N}(\cdot)$	Estimator of N of Coca [2015]
$N^G, N^G(\cdot)$	The generalised Lévy distribution function
$\widehat{N}^G, \widehat{N}^G(\cdot)$	Estimator of N^G of Nickl and Reiß [2012]
$\mathcal{N}, \mathcal{N}(\cdot)$	A functional of the Lévy measure
$\widehat{\mathcal{N}}, \widehat{\mathcal{N}}(\cdot)$	Estimator of $\mathcal{N}$ of Nickl et al. [2016]
$\mathbb{N}$	The set of natural numbers (nonnegative integer numbers)
ν	The Lévy measure
O, o	Big-O and small-o notation
O_{Pr}	Big-O notation in Pr-probability
$(\Omega, \mathcal{A}, Pr)$	The underlying probability space
P, P_n	The law of an increment of a Lévy process and its empirical counterpart
φ, φ_n	The characteristic function of a Lévy process and its empirical counterpart
Re, Im	The real and the imaginary parts of complex number
$\mathbb{R}$	The set of real numbers
$\mathbb{R}^d$	The d-dimensional Euclidean space
$\operatorname{supp}(\cdot)$	The support of a function
T	The total length of the observation interval
$(X_t)_{t\geq 0}$	A Lévy process (generally compound Poisson process)
$Y_1, Y_2, \ldots$	The underlying jumps of a compound Poisson process
$Z_1, \ldots, Z_n$	The increments of a discretely observed Lévy process
$\mathbb{Z}$	The set of integer numbers
$\lfloor x \rfloor$	The greatest integer less or equal to x
$\|\cdot\|_{\ell^\infty(C)}$	Uniform or supremum norm on $C \subseteq \mathbb{R}$
$\|\cdot\|_\infty$	Short version of $\|\cdot\|_{\ell^\infty(\mathbb{R})}$
$\|\cdot\|_{\infty,\tau}$	The (τ-)exponentially downweighted uniform norm on $\mathbb{R}^+$, $\sup_{x\geq 0} \lvert e^{-\tau x} f(x)\rvert$ with $\tau \geq 0$
$\|\cdot\|_{TV}$	Total variation norm
$\to^d$	Classical convergence in distribution
$\to^{\mathcal{D}}$	Convergence in distribution in the sense introduced by Hoffmann-Jørgensen
$\lesssim, \gtrsim$	Uniform order relations
$\sim$	Uniform equivalence relation

Chapter 1

Introduction

The purpose of this chapter is to introduce the problem we solve in this thesis: to make efficient nonparametric inference on discretely observed compound Poisson processes. In Section 1.1 we introduce these processes and some of their probabilistic properties, which are exploited in subsequent chapters. Based on this, in Section 1.2 we justify why when making statistical inference on them we take a nonparametric approach instead of a more classical parametric one. In Sections 1.3 and 1.4 we introduce nonparametric estimation of compound Poisson processes in two different frameworks depending on the type of data available to the statistician: in the first, the process is fully observed, whilst in the second it is only observed at discrete times. Estimation in the first framework follows from standard results in mathematical statistics. Nonetheless, it allows us to naturally motivate the type of results that are the focus of this thesis and which are developed in the second framework. In Section 1.4 we state the exact problem that concerns us here, and in Section 1.5 we clearly indicate our contributions to it and give an outline of the rest of the thesis.

1.1 Compound Poisson processes

Compound Poisson processes are one of the most basic examples of continuous-time pure-jump stochastic processes. Nevertheless, they provide enough flexibility to model a large number of phenomena observed in numerous fields. They are highly tractable from a mathematical point of view and are building blocks of many more elaborated stochastic processes such as queues and Lévy processes.

The construction of a compound Poisson process can be summarised as follows:

(a) the occurrence of the jumps is determined by a Poisson process;

(b) the size of the jumps is distributed according to a general distribution function;

(c) the interarrival times between jumps and the size of the jumps are mutually independent random variables; and,

(d) the process may increase or decrease steadily due to a drift factor.

In mathematical terms, let $(N_t)_{t\geq 0}$ be a d-dimensional Poisson process with intensity $\lambda > 0$. Let $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables taking values in $\mathbb{R}^d$ with common distribution F. Assume this sequence is independent of the Poisson process and let $\gamma \in \mathbb{R}^d$. Then a d-dimensional compound Poisson process $(X_t)_{t\geq 0}$ with drift γ, intensity λ and jump size distribution F can be written as

$$X_t = \gamma t + \sum_{j=1}^{N_t} Y_j, \qquad t \geq 0, \tag{1.1.1}$$

where an empty sum is zero by convention so, in particular, the process always starts at zero. Figure 1.1 illustrates a typical path of a one-dimensional compound Poisson process without drift. In what follows we concentrate on the case $d = 1$.

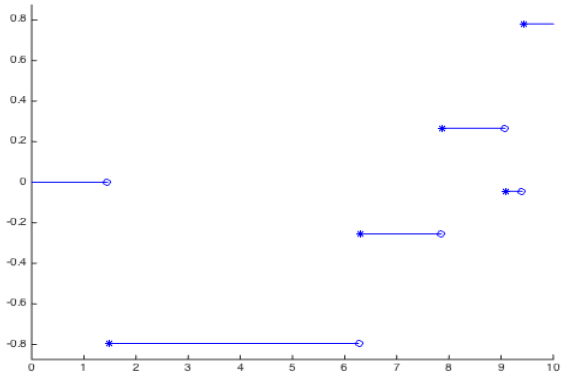


Figure 1.1: $(X_t)_{t\in[0,10]}$ with $\gamma = 0$, $\lambda = 0.5$ and $F = N(0, 1)$

A one-dimensional Poisson process with intensity λ can be constructed as follows. Let $E_1, E_2, \ldots$ be a sequence of independent and identically distributed exponential random variables with parameter λ, with the convention that λ^{-1} is their expected value. Then the Poisson process can be written as

$$N_t = \max\left\{k \in \mathbb{N} : \sum_{j=1}^{k} E_j \leq t\right\}, \qquad t \geq 0,$$

where, again, an empty sum is zero by convention so, in particular, the process always starts at zero. From this construction it is apparent that the interarrival times of the compound Poisson process are exponentially distributed. As pointed out in Section 5.4 in Bingham and Kiesel [2004], the exponential distribution is special in that, subject

to minimal regularity assumptions, it is the only distribution with the so-called *lack-of-memory* property: if $(\Omega, \mathcal{A}, Pr)$ denotes the underlying probability space on which all random elements of this thesis are defined, then

$$Pr(E_1 > s + t \mid E_1 > s) = Pr(E_1 > t), \quad \text{for all } s, t > 0. \tag{1.1.2}$$

This property provides the compound Poisson process with the (strong) Markov property, which makes it very mathematically tractable. In particular, it means that the survival function of the interarrival distributions does not change depending on where we observe them. Intuitively, the random time of occurrence of a jump shows no ageing. Consequently, compound Poisson process are used in numerous applications to model random events that occur 'out of the blue'.

In order to gain more intuition into these processes, let us introduce two applications in which they are used. These also motivate the division between Section 1.3 and Section 1.4. Consider an ATM where deposits can be made. Customers arrive at random and withdraw or deposit a random amount of money. The interest of an ATM manager is to understand how much money is likely to be enough to cover the random transactions without having to risk introducing too much of it in the machine. In normal conditions, the assumptions of mutually independent exponential interarrival times and F-distributed amounts of money are satisfied. However, the homogeneity in the underlying parameters may be violated, so in practice the model is applied after splitting the observations into time-intervals where this is a reasonable assumption. A second example comes from ecology. Consider a nature reserve where groups of birds stop in the course of their migration. Within certain weeks of the year they arrive and leave at random with an approximately constant intensity rate. Furthermore, and in those weeks, the size of their groups also varies randomly according to an approximately homogeneous-in-time distribution. A manager of the park requires to understand the minimum and maximum number of birds that typically occupy a certain area to build appropriate infrastructure. Therefore, in both applications the one observes the aggregated effect of all the random events and this justifies the practical importance of compound Poisson processes. As we mention in Sections 1.3 and 1.4, the difference between the two is the frequency at which the process is observed. In either case, the challenge is to estimate the defining parameters of the underlying process, all of which are generally unknown to the practitioner. Although we discuss estimation of all, the most challenging one is F and this is the focus of this thesis.

Returning to the mathematical properties of the compound Poisson process, the lack-of-memory property and the independence assumptions imply it belongs to the wider class of Lévy processes. In other words, $(X_t)_{t\geq 0}$ as defined above enjoys the following properties (cf. Sato [1999]):

(i) its paths are Pr-almost surely *càdlàg* (right continuous with left limits) starting at

zero; and,

(ii) it has stationary and independent increments.

The second is the key property of Lévy processes. It has been extensively exploited to make statistical inference on them and so do we in subsequent chapters. Furthermore, it gives rise to the intimate relationship that these processes have with infinitely divisible distributions, which were introduced by De Finetti [1929]. As a consequence, and as shown by Lévy [1934], Khintchine [1937] and Itô [1942], the characteristic function of a Lévy process at a certain time is given by the Lévy-Khintchine formula, which explicitly depends on the defining parameters of the process. This can be exploited to build estimators of them, especially when having discrete observations of the process. Indeed, it is the starting point of the so-called spectral approach, which we introduce in Section 2.2 and use in Chapter 3 to show our novel results. We postpone introducing the formula in full generality until Section 2.2.1 and for now we focus on its expression for the case of a compound Poisson process. For this purpose, let $\mathbb{E}$ denote the expectation under Pr and let $\mathcal{F}$ denote the Fourier (–Plancherel) transform acting on finite measures with the convention

$$\mathcal{F}\mu\,(u) = \int_{\mathbb{R}} e^{iux}\mu(dx), \quad u \in \mathbb{R}\,. \tag{1.1.3}$$

Then, by a simple calculation using (1.1.1), the characteristic function of X_t can be written as

$$\varphi_t(u) := \mathbb{E}[e^{iuX_t}] = e^{t\left(iu\gamma + \mathcal{F}\nu\,(u) - \lambda\right)}, \quad t \geq 0, u \in \mathbb{R}, \tag{1.1.4}$$

where ν is generally referred to as the Lévy measure and it satisfies $\nu = \lambda\, dF$. Throughout we denote the measure associated to a distribution function F by dF. Note, in particular, that

$$-2\lambda \leq \mathrm{Re}(\mathcal{F}\nu\,(u) - \lambda) \leq 0 \quad \text{ and } \quad |\,\mathrm{Im}(\mathcal{F}\nu\,(u) - \lambda)| \leq \lambda, \quad u \in \mathbb{R}, \tag{1.1.5}$$

and therefore

$$1 \geq \inf_{u\in\mathbb{R}} |\varphi_t(u)| \geq e^{-2\lambda t} > 0, \quad t > 0. \tag{1.1.6}$$

We repeatedly use these properties in subsequent sections and chapters. We remark that from now on we assume F has no atom at the origin. This is a harmless requirement that guarantees identifiability all of the parameters: otherwise, if a jump can take the value zero, the process does not jump and the mass of such atom cannot be identified. For the same reason, estimation of the intensity cannot be performed. Indeed, if $p_0 \in [0, 1)$ is the mass of dF at the origin, identity (1.1.4) can be written as

$$\varphi_t(u) := e^{t\left(iu\gamma + \mathcal{F}\nu_0\,(u) - \lambda_0\right)}, \quad t \geq 0, u \in \mathbb{R},$$

where $\nu_0 := \lambda_0 dF_0$, $\lambda_0 := \lambda(1-p_0)$ and $dF_0 := (1-p_0)^{-1}(dF - p_0\delta_0)$. Therefore, the zero jumps coming from p_0 and the flat areas arising from λ tangle up to be indistinguishable from a process with intensity λ_0, and this is the only parameter that can be identified. Moreover, all the compound Poisson processes resulting from all possible choices of $p_0 \in [0,1)$ and $\lambda > 0$ for which the product $\lambda(1-p_0)$ remains constant are indistinguishable from each other. Consequently, we may assume $p_0 = 0$ without loss of generality and this way avoid the lack of identifiability of the parameters p_0 and λ.

The statistical relevance of identity (1.1.4) is the following: assume we only have i.i.d. observations of X_t for some $t \geq 0$ fixed. Then, due to the definition of φ_t as an expectation, it is possible to approximate it by its empirical version, the empirical characteristic function. The idea of the spectral approach to construct estimators is to find relationships between each of the parameters and the characteristic function using (1.1.4), and to 'plug in' the empirical version of the latter in place of it. As it should be clear by now, the fact that φ_t is the Fourier transform of the law of X_t, and that we are hence resorting to the frequency domain to construct estimators, gives the whole approach the name of spectral approach. We postpone giving more details of it until the next chapter.

1.2 Why nonparametric estimation?

The most classical branch of statistics is concerned with making conclusions on models that are defined through a finite-dimensional parameter. This is coined parametric statistics and has a long history dating back to J. Bernoulli, A. de Moivre and T. Bayes in the 18^{th} century and further developed by P. S. Laplace and C. F. Gauss in the late 18^{th} and early 19^{th} centuries. Its mathematical foundations were laid in the first half of the 20^{th} century by the remarkably mathematically-gifted biologist R. A. Fisher and, in the second half of the century, the mathematical theory was finally formalised and unified to its modern form by L. M. Le Cam. For an extensive account on its fascinating history we refer the reader to Hald [2007], and for a comprehensive account of its mathematical foundations to van der Vaart [1998]. One of the advantages of parametric procedures is that generally they are simple to formulate and fast to compute. However, in many modern practical situations there is no reason to believe the assumption of the data being generated by a specific finite-dimensional model is correct. In fact, this assumption naturally gives rise to the problem of robustness to the choice of model.

Recall that in the definition of the compound Poisson process above no assumptions were made on the distribution F other than it having no mass at the origin. Consequently, the main estimation problem that concerns us here is that of estimating a family of functionals of a general probability measure and it naturally falls into the realm of inference on infinite-dimensional models or nonparametric statistics. The development of this branch is more contemporary than that of parametric statistics and arguably began in the 1930s

with Glivenko–Cantelli's theorem, independently proved by Glivenko [1933] and Cantelli [1933], and the Kolmogorov–Smirnov statistic studied in Kolmogorov [1933] and Smirnov [1939]. We note in passing that this coincides with the development of the study of Lévy processes mentioned above. As a result of the advent of powerful computers and the new and challenging modelling demands, the last few decades have witnessed the largest development of this field and it currently is a very active area of research. We refer the reader to the recent and unified account of its mathematical foundations given in Giné and Nickl [2016]. In contrast to parametric statistical procedures, nonparametric analogues do not suffer as much from the highly undesirable drawback of model misspecification. For all these reasons herein we focus on nonparametric estimation of compound Poisson processes.

A thoughtful reader may argue that we are making a parametric assumption by letting the interarrival times be exponentially distributed. Nonetheless, we emphasise that we are interested in modelling a process with jumps whose occurrence shows no ageing, because in many applications this is a reasonable approximation. This qualitative property, depicted by (1.1.2), is mathematically equivalent to requiring that the survival function of the interarrival distribution satisfies the Cauchy-functional equation $f(x+y) = f(x)f(y)$, $x, y \in \mathbb{R}$. As mentioned above, the only solution to this functional equation under minimal regularity assumptions is the parametric exponential distribution. The drift term can also be generalised to a time-dependent function but, in the discrete observation setting introduced in Section 1.4 and considered in the rest of the chapters, strong assumptions have to be made to obtain weak conclusions. We therefore lose almost no generality by assuming the relevant-in-practice modelling assumption of a linear deterministic function in (1.1.1).

1.3 Estimation of compound Poisson processes from continuous observations

We begin introducing the problem of nonparametric estimation of compound Poisson process under a simple observation assumption. We do this in Section 1.3.1, where we also give its solution. Then, in Section 1.3.2 we briefly discuss a few concepts from empirical process theory and information theory arising from it. These are repeatedly used in subsequent chapters and the section serves as a natural motivation for the results therein.

1.3.1 The estimation problem

Recall the example of the ATM introduced in Section 1.1. At the end of the day, the ATM manager can access the machine's database and see the time at which each transaction took place and its value, just as in Figure 1.1. Therefore, there exists a $T > 0$ such that they

observe X_t for all $t \in [0,T]$. This is commonly referred to as the continuous observation scheme. In this case, γ is observed directly and so are the independent interarrival times $E_1, \dots, E_{N_T}$ and the independent jumps $Y_1, \dots, Y_{N_T}$. Thus, classical statistical techniques can be used to estimate λ and F (cf. van der Vaart [1998]): if we condition on $N_T = m$, the intensity can be estimated by the maximum likelihood estimator

$$\lambda_m := \frac{1}{\frac{1}{m}\sum_{j=1}^m E_j}$$

and the jump distribution function by the empirical distribution function

$$F_m(x) := \frac{1}{m}\sum_{j=1}^m \mathbb{1}_{(-\infty,x]}(Y_j), \quad x \in \mathbb{R}\,.$$

The asymptotic behaviour of these estimators is well-known: the delta method and the central limit theorem guarantee that

$$\sqrt{m}(\lambda_m - \lambda) \to^d N\left(0, \lambda^2\right) \quad \text{as } m \to \infty, \tag{1.3.1}$$

and Donsker's theorem states that

$$\sqrt{m}(F_m - F) \to^{\mathcal{D}} \mathbb{G}_F \quad \text{in } \ell^\infty(R) \quad \text{as } m \to \infty, \tag{1.3.2}$$

where $\mathbb{G}_F$ is the so-called F-Brownian bridge, i.e. the zero-mean Gaussian process with covariance function given by $\Sigma_{x,y} := F(\min\{x,y\}) - F(x)F(y)$, $x, y \in \mathbb{R}$. The meaning of $\to^d$ and $\to^{\mathcal{D}}$ is discussed in the next section.

These results can be considered as generalisations of the classical central limit theorem, the pillar upon which countless statistical procedures rest. Its statistical importance lies in that it not only guarantees consistent estimation at the parametric rate of convergence $1/\sqrt{m}$, but it also states that if the fluctuations of the estimator around the truth are rescaled by the inverse of this rate, they are asymptotically normally distributed. Thus, the quantiles of the limiting distribution can be calculated, and confidence intervals and statistical tests can be derived from it. Analogous procedures can be derived from (1.3.1) and (1.3.2), and this justifies why in this thesis we focus on discussing and developing this type of results.

1.3.2 A glimpse of empirical process theory and of asymptotic efficiency

In (1.3.1) we have used the notation $\to^d$ to denote convergence in distribution. Recall that, if $\mathbb{D}$ is a metric space, a sequence of random variables $A_m : \Omega \to \mathbb{D}$, $m \in \mathbb{N}$, converges in

distribution or weakly to a random variable $A : \Omega \to \mathbb{D}$ if

$$\mathbb{E}\, f(A_m) \to \mathbb{E}\, f(A) \quad \text{for all } f \in C_b(\mathbb{D}), \tag{1.3.3}$$

where $C_b(\mathbb{D})$ is the space of real-valued continuous and bounded functions on $\mathbb{D}$.

In (1.3.2) we used a different notation, $\to^{\mathcal{D}}$, because, there, convergence in distribution is not necessarily well-defined: throughout we interpret the random elements on both sides as taking values in the space $\ell^\infty(\mathbb{R})$ of real-valued Lebesgue-measurable bounded functions on $\mathbb{R}$ the uniform norm $\|F\|_\infty := \sup_{x\in\mathbb{R}} |F(x)|$; the issue is that, under this interpretation, the left hand side is not necessarily a *bona-fide* random variable because it may not be measurable with respect to the corresponding Borel σ-field as this turns out to be too large (note that $\ell^\infty(\mathbb{R})$ is not even separable). We denote by $\to^{\mathcal{D}}$ convergence in distribution in the sense introduced by Hoffmann-Jørgensen [1984]: in this, the expectation on the left hand side of (1.3.3) is substituted by the more general operation of outer expectation,

$$\mathbb{E}^* B := \inf\{\mathbb{E}\, U : U \geq B, U : \Omega \to \bar{\mathbb{R}} \text{ is measurable and such that } \mathbb{E}\, U \text{ exists}\},$$

where $\bar{\mathbb{R}}$ is the extended real line. Provided that the limiting random element A is Borel measurable, corresponding versions of Portmanteau's theorem, the continuous mapping theorem and Prokhorov's theorem, together with a notion of tightness and tools to show it and to show weak convergence, still hold under this notion of weak convergence. We implicitly but heavily make use of these generalisations in the proofs of our results in Chapter 3 and refer the reader to any of the following excellent accounts for more details: van der Vaart and Wellner [1996], Dudley [1999] and Giné and Nickl [2016]. We remark that this is the modern approach to empirical process theory because of its power, but others were introduced before: Skorokhod [1957] introduced the Skorokhod topology under which the random elements on the left hand side of (1.3.2) are measurable if interpreted as mappings taking values in the space of functions that are right continuous with left limits; Dudley [1966] proposed an alternative weak convergence theory based on a smaller σ-algebra than the one considered above and for which measurability holds — this approach was general enough to be the first to prove the multidimensional version of Donsker's theorem; and Pyke and Shorack [1968] proposed yet another version of weak convergence requiring (1.3.3) to hold only for those functions f for which $f(A_m)$ is measurable. All this fruitful research was sparked by the heuristic introduction of Donsker's theorem in Doob [1949] and by the first attempt to prove it by Donsker [1952]. The latter overlooked the delicate issue of the measurability mentioned above and it was precisely this that triggered all the research in the field.

With the modern interpretation of Donsker's theorem, it can also be understood as a functional central limit theorem —in the same way that the aforementioned Glivenko–Cantelli's theorem can be understood as a functional law of large numbers with the uni-

form norm. Therefore, (1.3.2) is of great statistical interest because it guarantees $1/\sqrt{m}$-consistent uniform estimation, and because confidence bands, goodness-of-fit tests and two-sample tests can be derived from it. To develop all of these, the Kolmogorov–Smirnov statistic, also mentioned in the previous section, plays a major role. This is the distribution of $\|\mathbb{G}\|_\infty$, where $\mathbb{G}$ is a standard Brownian bridge, i.e. $\mathbb{G}_F$ from above with $F = U(0,1)$, the uniform distribution on $[0,1]$, and the content of their result is that $\|\mathbb{G}\|_\infty =^d \|\mathbb{G}_F\|_\infty$ for any continuous F. We note in passing that this was rigorously proved prior to Donsker's theorem. Thus, in view of (1.3.2), to construct these inference procedures for any such F we simply have to calculate the quantiles of this distribution-free statistic.

The last desirable property of F_m that we remark here is that of asymptotic efficiency, which also follows from (1.3.2): if we observe m independent realisations of an F-distributed random variable and no assumption is imposed on F, the covariance function of the F-Brownian bridge coincides with the Cramér–Rao lower bound of the model and therefore F_m is asymptotically efficient from an information-theoretic point of view. This was first shown by Dvoretzky et al. [1956], and the modern formulation of semiparametric efficiency was introduced by Hájek [1970, 1972] and Le Cam [1972]. Intuitively, it means that, in the limit as $m \to \infty$, the empirical distribution function extracts as much information about F as it is possible to extract from the observations at hand. We remark that asymptotic efficiency is relative to the model at hand because so is the Cramér–Rao lower bound, and what we just stated is that F_m has this property when nothing is known about F. A priori knowing some qualitative properties of F may make make the Cramér–Rao lower bound decrease, although this is not always the case: as shown by Kiefer and Wolfowitz [1976], it does not change if F is known to be concave or convex, and thus F_m is still efficient for such model. However, if stronger qualitative assumptions on F are made it does decrease, such as when F is determined by a finite-dimensional parameter in which case the bound is simply given by the Fisher information matrix. Under mild regularity assumptions, asymptotic normality of the maximum likelihood estimator around the true parameter holds and the variance of the limiting normal distribution attains the new information lower bound. This is the case for the exponential distribution, for which the variance in (1.3.1) is the lower bound, and the estimator of the intensity introduced above is also efficient. The dependency on the model is important and will play a crucial role in the discussions about asymptotic efficiency in the next chapter. This property, regardless of the assumptions on F, is important for statistical purposes because it implies optimality of the size of confidence regions around estimates and of the statistical power of the resulting testing procedures. In the context above, the latter means that if a test is constructed using the confidence regions arising from the central limit theorems, the probability that it rejects the null hypothesis when it is not true is maximised.

1.4 Estimation of compound Poisson processes from discrete observations

We can now introduce the more involved framework under which the results of the rest of the chapters are developed. In Section 1.4.1 we describe the estimation problem that concerns us in this thesis and touch upon some of its singularities, and in Section 1.4.2 we give a brief history of the problem which allows us to clearly describe our contributions to it.

1.4.1 The estimation problem

Recall the example of birds migration. Unlike the manager of the ATM, the manager of the nature reserve does not have a database of the time when a group of birds arrived or left and of the size of the group. Instead, they count or estimate the total number of birds on a (discrete) regular basis, such as once a day. This means that $(X_t)_{t\geq 0}$ is not continuously observed up to some time $T > 0$ but rather discretely observed every some $0 < \Delta$-amount of time. In other words, the observations the manager has are $X_\Delta, \dots, X_{n\Delta}$, where $n = \lfloor T/\Delta \rfloor \in \mathbb{N}$. In Figure 1.2 we have included the path in Figure 1.1 when it is discretely observed for $\Delta = 2.5$.

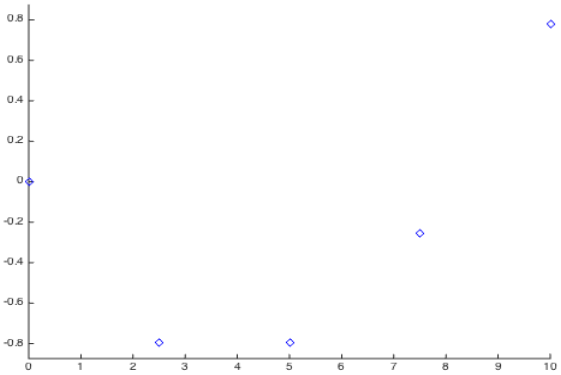


Figure 1.2: $X_\Delta, \dots, X_{n\Delta}$, $\Delta = 2.5$ and $n = 4$ ($\gamma = 0, \lambda = 0.5, F = N(0,1)$)

Unless otherwise stated we assume Δ is fixed and does not change with n, which is generally called the low-frequency observation scheme. This framework is very common in practice, especially when considering storage systems or chemical-biological reactions, where it may be expensive or impossible to take measurements continuously. We can therefore formulate the problem as follows:

> **Problem:** *In this thesis we focus on developing the same type of results as those from Section 1.3 but in the discrete observation regime. More precisely, we consider a one-dimensional compound process observed at discrete times*

for some $\Delta > 0$ fixed and the problem is that of proving optimal (in the sense of asymptotic efficiency introduced in Section 1.3.2) central limit theorems for the parameters F, λ and γ. The estimation of the first is nonparametric and under the standard supremum norm, whilst that of the other two is parametric and under the Euclidean norm. The most challenging part of the problem is estimating F and it is the main focus of the thesis. Yet, we also discuss estimation of the other parameters in detail.

From the figure above, it is apparent that this estimation problem is much more challenging than that considered in Section 1.3: in between two observations we may miss information such as the number of jumps that took place, their size and the time at which they occurred. Therefore, when making inference on the parameters we are confronted with a statistical inverse problem and, a priori, there is no reason to believe estimators with the same type of desirable properties mentioned in Section 1.3.2 exist. Let us elaborate on this focusing on the estimation of the jump distribution. In view of property (ii) in page 5, the increments $Z_k := X_{k\Delta} - X_{(k-1)\Delta}$, $k = 1, \dots, n$, are independent and identically distributed. As it is customary in the literature, we work with these instead of with the observations of the process because of their tractability. Due to the fact that $X_0 \equiv 0$ Pr-almost surely by property (i) in page 5, the common distribution of the increments is then given by that of

$$Z_1 := X_\Delta := \gamma\Delta + \sum_{j=1}^{N_\Delta} Y_j. \tag{1.4.1}$$

By the mutual independence assumption (c) in page 3, we are essentially observing Y_1, the random variable with the distribution of interest, 'corrupted' by a sum of a random number of independent copies of itself plus a drift term. This means the inverse problem is non-linear, which is more precisely depicted by the non-linear relationship between φ_Δ and $\mathcal{F}\nu$ following from (1.1.4). Then, at first it is not clear that estimators attaining the $1/\sqrt{n}$-rate of convergence exist in this inverse setting. Neumann and Reiß [2009] were the first to study the "ill-posedness" of the problem for the more general model of Lévy processes and they showed it directly depends on the decay of the characteristic function of the increments: the faster it decays, the harder it is to invert the problem and the slower the convergence rate of an optimal estimator can be. This parallels the classical deconvolution problem (see Fan [1991]), but with the 'noise' or error being the observation itself and, furthermore, unknown (up to its structural properties). Indeed, and as already pointed out by Nickl and Reiß [2012], the linearised problem is exactly of deconvolution type with such error, and the problem is generally referred to as of auto-deconvolution type. We note in passing that from the last display one already sees the observation is approximately Y_1 plus an independent copy of the increment Z_1. Therefore, to study the ill-posedness of the problem we need to look at the decay of φ_Δ. Taking $t = \Delta > 0$ in

(1.1.6) we observe that

$$\inf_{u\in\mathbb{R}} |\varphi_\Delta(u)| \geq e^{-2\lambda\Delta} > 0, \tag{1.4.2}$$

and the characteristic function does not decay to zero at the tails. Hence, the inverse problem is not ill-posed in terms of convergence rates and estimators of F that converge with the parametric rate $1/\sqrt{n}$ and that satisfy functional central limit theorems under certain norms may exist. Indeed, they do exist and can be found in the works of Buchmann and Grübel [2003] and Coca [2015], on which we focus throughout this thesis with particular emphasis on the latter and on extensions of it. In the following section we give a brief description of their results, and in Chapters 2 and 4 we explore their differences and similarities in depth.

1.4.2 A brief history of the solution to the problem

The first to tackle the problem stated above were Buchmann and Grübel [2003], who partially solved it assuming $\gamma = 0$, knowledge of λ and that $\text{supp}(F) \subseteq \mathbb{R}^+$. Using a direct approach (cf. Section 2.1), they constructed an estimator of F given by an infinite series making no regularity assumptions. Due to this series, the estimator of $F(x)$ may diverge as $x \to \infty$ and to control it they introduced the exponentially downweighted norm $\|F\|_{\infty,\tau} = \sup_{x\geq 0} |e^{-\tau x}F(x)|$, where $\tau \geq 0$ is such that $\int_{\mathbb{R}^+} e^{-\tau x} dF(x) < \log(2)/(\lambda\Delta)$. Then, they proved a functional central limit theorem under this norm, identifying the limiting Gaussian process. It follows that their estimator satisfies a classical Donsker theorem, i.e. under the standard uniform norm $\|\cdot\|_\infty = \|\cdot\|_{\infty,0}$, only if $\lambda\Delta \leq \log(2) \approx 0.69$. Note that $\lambda\Delta$ represents the expected number of jumps the compound Poisson process has given in a Δ-observation-interval, so this condition is quite restrictive in practice. Furthermore, and as discussed in Section 2.1, the introduction of the exponentially downweighted norm is necessary in the rest of the cases. Consequently, in general the resulting estimation is only valid at small positive values and, as observed in our and their simulations, the deterioration of the estimates away from the origin is noticeable and severe in some realistic cases. Nonetheless, the work of Buchmann and Grübel [2003] deserves all merits and has been cited by numerous authors as it is one of the first of its kind. Furthermore, they developed it when the understanding of these type of problem was still very limited, as the first half of Section 6.4 "Concluding remarks" in their work clearly shows:

> *"The theorems in Section 2 show that the decompounding problem can be solved on the usual $n^{-1/2}$-level, a fact that we continue to find slightly surprising. At least in the general case we were initially regarding the problem as "ill-posed", with the corresponding consequences such as a rate lower than $n^{-1/2}$ for the estimates. Of course, the classification of a problem as ill-posed or inverse, etc. depends on the choice of topologies, so our results may be rephrased as saying that there are statistically meaningful choices for the latter where decompound-*

ing can be considered to be a perfectly regular problem."

As mentioned in the previous section, the ill-posedness of the problem was finally understood after the work of Neumann and Reiß [2009], who studied it for more general Lévy processes using the spectral approach first introduced by Belomestny and Reiß [2006]. It follows from their results that, in addition to the compound Poisson case, the parametric rate of convergence $1/\sqrt{n}$ can only be attained for a class of Lévy processes with infinite activity and Blumenthal-Getoor index 0 (i.e. with 'the mildest infinite activity possible'). In this case, the jump measure driving the jumps is not finite, unlike in the compound Poisson case, due to a nonintegrable discontinuity at the origin giving rise to the infinite activity (for more details see Section 2.2.1). Therefore, the classical distribution function F of the jump measure is only well-defined in $\mathbb{R}^-$, and for $\mathbb{R}^+$ one should consider the integral of the right tail instead. Nickl and Reiß [2012] constructed a kernel-type estimator of this generalised distribution function using the spectral approach and, following the insight of Neumann and Reiß [2009], proved a functional central limit theorem under the norm $\|G\|_{\ell^\infty(\mathbb{R}\setminus(-\zeta,\zeta))} := \sup_{|x|\geq\zeta}|G(x)|$, for some $\zeta > 0$ fixed, identifying the limiting process. For clarity, their result also applies to discretely observed compound Poisson processes whose jump measure has a density with a finite second moment.

The machinery they developed and the strategy they took to prove their result was general enough to prove functional central limit theorems under the standard uniform norm in related problems. In particular, Söhl and Trabs [2012] showed such theorem for the simpler linear inverse problem of deconvolution mentioned above. Nickl and Reiß [2012] conjectured that, in the compound Poisson case, a central limit theorem for the classical distribution function F and with the unrestricted standard supremum norm $\|\cdot\|_\infty$ should hold with their estimator and under the same assumptions. Indeed, Coca [2015] proved such a result without any knowledge of $\gamma \in \mathbb{R}$ and $\lambda > 0$, following their strategy but constructing a different estimator and considerably relaxing the regularity assumptions therein. In Chapter 2 we discuss the difficulties faced when trying to use the estimator of Nickl and Reiß [2012] in this case. Regarding the regularity assumptions, Coca [2015] simply assumed a finite logarithmic moment. Moreover, and despite using a kernel estimator, Coca [2015] was able to make inference on a potential discrete component in the jump distribution. To our knowledge, this is the first work in the literature of nonparametric statistics to use such estimators for this purpose, especially when showing functional central limit theorems. More importantly, the strategy he used is general enough to apply to many other kernel estimators. It allowed him to construct statistical tests of each of the components, which are also new. We remark that Buchmann and Grübel [2003] were able to show a central limit theorem for the mass function of the jump measure if this was the only component present in it. In fact, the conclusions of their results are stronger for this estimator than for that of F, although the assumption they make is still stronger than the above-mentioned finite logarithmic moment condition. Coca [2015] also constructed

estimators of λ and γ. For the former he showed a central limit theorem identifying the variance of the asymptotic normal distribution, and for the latter he showed it converges to γ at rate $h_n/\sqrt{n}$, where h_n is the bandwidth of the estimators. Due to the lack of ill-posedness of the problem, h_n can be chosen to decay exponentially fast to zero and the rate of convergence of the estimator of γ is much faster than the parametric rate $1/\sqrt{n}$. This parallels the setting of continuous observations of Section 1.3, in which it can be perfectly learnt Pr-almost surely with $T > 0$ arbitrarily small because it is the slope of the observed process in any region with no jumps. No central limit theorem can be shown for this estimator.

All the estimators of F in the works mentioned so far except for those in Coca [2015] have the drawback that they do not return bona-fide distribution functions because they are not monotonic and do not necessarily end at 1. The estimators in Coca [2015] do have the former undesirable property which, as we argue in the next section, is unavoidable in these inverse problems unless further transformations are made. In the case of a purely atomic jump distribution supported in $\mathbb{R}^+$, Buchmann and Grübel [2004] showed asymptotic normality of two estimators with such transformation, although that do not necessarily end at 1.

Having stated the existence of functional central limit theorems for F in this inverse setting, we are missing to touch upon the asymptotic efficiency of these estimators. Trabs [2015a] recently identified the information lower bounds for this problem and for the more general Lévy processes considered by Nickl and Reiß [2012], assuming the associated Lévy measure is supported in $\mathbb{R}$ (more precisely, in $\mathbb{R}^d$, but we are only concerned with the case $d = 1$). Recall that Buchmann and Grübel [2003] assumed F is supported in $\mathbb{R}^+$ and hence the lower bound may decrease in this restricted model. Nonetheless, and as we show in the Section 2.6, the covariance function of their limiting process does not attain the Cramér–Rao bound of the unrestricted model and therefore their estimator is not efficient. This point was also briefly mentioned by Buchmann and Grübel [2003], who already anticipated that their covariance may not be optimal; we quote the lines after the paragraph quoted above from the Section 6.4 "Concluding remarks" in their work:

> *"However, numerical experiments such as given in Section 3 remind us of the fact that a good rate is not a guarantee for high precision: comparing the left-hand and the right-hand plots in Figure 1 shows that the constant in front of the rate may be rather high."*

On the other hand, the covariance of the limiting process in the limit theorem in Coca [2015], where no assumption on the support of the Lévy measure is made, does attain the information lower bound and our estimator is the first to be asymptotically efficient. Similarly, the estimator of the intensity constructed by Coca [2015] is efficient too as we argue in Section 2.6, and it is also the first to be so in the setting we consider. We note

in passing that the estimator of the generalised distribution function of Nickl and Reiß [2012] does attain the Cramér–Rao bound for that particular object.

1.5 Organisation and contributions of the thesis

This thesis revolves around our manuscript Coca [2015] which, at the time of writing, is under revision. In every chapter we explore the problem introduced in the previous section focusing on different aspects of it, with each being a novel contribution to the solution and understanding of it. No chapter or section in this thesis has arisen from any collaboration with other individuals.

1.5.1 Chapter 2

As mentioned in Section 1.4.2, the problem has been tackled using two different approaches: the direct approach in Buchmann and Grübel [2003] and the spectral approach in Coca [2015]. The former is a natural approach in the more general problem of nonparametric estimation of compound distributions whilst the latter has been extensively used in the also more general problem of nonparametric estimation of discretely observed Lévy processes. In Chapter 2 we discuss the literature in these two fields that is directly related to the estimation problem considered in this thesis. In particular, we construct and carefully analyse existing estimators in different nonparametric problems of discretely observed compound Poisson processes using both approaches. This allows us to conclude that the two are formally equivalent and that all the estimators considered are very close relatives of each other. To our knowledge, this has not yet been pointed out in the literature. In fact, the two approaches have even been described as different alternatives (cf. van Es et al. [2007] and Comte et al. [2014]) or even as "quite different" (cf. Duval [2013a]). We identify the reasons why the direct approach must require stronger assumptions than the spectral approach, hence advocating the use of the latter and giving rise to a number of open problems that have not yet been proposed. We also show the close relationship between existing estimators constructed through the spectral approach. Additionally, we study the efficiency of the estimators, which allows us to further explore their differences. We provide new insight into the topic of efficiency in our setting and, in particular, argue how estimators that are not efficient can achieve this property. All these discussions are novel and give much more understanding into the problem than what there currently seems to be in the literature.

1.5.2 Chapter 3

Chapter 3 contains the theoretical results in Coca [2015]. As mentioned in Section 1.4.2, we are the first solve the general problem stated in Section 1.4.1 and we do so under mild

assumptions. More specifically, under the low-frequency regime of Section 1.4 we construct efficient nonparametric estimators of the jump distribution and of the Lévy distribution (the integrated Lévy measure introduced in Section 1.1). These include a kernel estimator of a potential mass function without knowledge of it, which is a novel addition to the field of nonparametric statistics. Furthermore, the construction generalises to other existing kernel estimators and it gives rise to new procedures to test the presence of an atomic and/or an absolutely continuous component in a distribution. We prove functional central limit theorems using the uniform norm for all quantities, and the limiting Gaussian processes are identified. This allows us to conclude efficiency of both estimators. We also construct novel and efficient estimators of λ and show asymptotic normality of them. These improve upon existing naive estimators in that they are the first to be consistent, and also efficient, in the setting when jumps may cancel each other in the observation (1.4.1). We also construct a novel estimator of γ and prove that its rate of convergence is exponential in the number of observations. This is an unusual rate and has not yet been observed in the literature discussed in Chapter 2. The main result in Coca [2015] and in Chapter 3 also identifies the asymptotic dependence structure of all quantities involved. We are the first to include this, which is of great interest for applications.

1.5.3 Chapter 4

In Chapter 4 we illustrate the applications, implementation and practical performance of the results from Chapter 3. In particular, in Section 4.2 we discuss the construction of confidence regions, goodness-of-fit tests, two-sample tests and tests for the presence of an atomic and of an absolutely continuous jump component. In Section 4.3 we give numerous details of the implementation required to use these in practice and also include the implementation using the estimator of the jump distribution proposed by Buchmann and Grübel [2003]; indeed, they only included the implementation of their estimator of the mass function, which is what they refer to in the second quote we gave in Section 1.4.1. Lastly, in Section 4.3 we illustrate, analyse and compare their practical performance through simulations in several realistic settings. We also include the study of the empirical coverage of the respective confidence regions, both individually and jointly in the parameters, in order to analyse their fitness for testing procedures. These studies are scarce if not inexistent in existing literature on our problem and on closely related ones. Furthermore, due to our detailed analysis, we identify behaviours that have not yet been reported in related literature and, therefore, the section serves as a solid basis for further interesting investigations that we propose there. We also give practical recommendations to circumvent issues arising in the implementation and simulations of the estimators. Most of the discussions of Section 4.2 appeared in the first version of the manuscript Coca [2015] but not in a second one, and the results in the rest of the sections are new contributions to the literature.

Chapter 2

Unifying existing literature

The problem we consider in this thesis falls into two more general problems: nonparametric estimation of compound distributions and of discretely observed Lévy processes. Nonparametric estimators in these problems have generally been constructed through two natural but different approaches, especially to prove functional central limit theorems: what we refer to as the direct approach and the so-called spectral approach. We discuss them in Sections 2.1 and 2.2, respectively, focusing mostly on nonparametric estimation of F and related quantities in the compound Poisson case. We review closely related literature along the way. In Section 2.3 we explicitly identify the links between these two (generally regarded as different) approaches and conclude the existing estimators constructed from them are very close relatives of each other. We also identify why, in the low-frequency regime introduced in Section 1.4, results following from the direct approach do and must require stronger assumptions than those derived from the spectral one, and thus argue that the latter should be the preferred approach when developing results in this area. In Section 2.4 we discuss existing estimators of the intensity and of the drift, and in Section 2.5 we discuss estimation of the process as a whole. Finally, in Section 2.6 we study the efficiency of all the estimators and make several new remarks on the topic. In Sections 2.3 and 2.6 we propose some new open problems. We therefore believe this chapter considerably increases the understanding of the problem and related ones, and we consider it a contribution to the literature in itself.

2.1 The direct approach

The direct approach is a natural one when nonparametrically estimating compound distributions. Therefore, we briefly introduce these prior to introducing the approach.

2.1.1 Compound distributions

In Section 1.4 of the last chapter we set up the framework in which the results of the rest of the thesis are developed. In particular, in (1.4.1) we gave the expression for the quantities we work with throughout. In view of it and if we assume $\gamma = 0$, the increment of a (zero-drift) compound Poisson process can be written as

$$Z = \sum_{j=1}^{M} Y_j, \tag{2.1.1}$$

where M is a Poisson distribution with parameter $\lambda\Delta$ and $Y_1, Y_2, \ldots$ is any sequence of i.i.d. real-valued random variables with distribution F. With this interpretation, it is said that Z follows a compound Poisson distribution and, if we let M be a general distribution supported in $\mathbb{N}$ and drop the i.i.d. assumption of the jumps, it follows a general compound distribution. Therefore, our problem falls into that of estimating the defining parameters of compound distributions from n (independent) observations of them, $Z_1, \ldots, Z_n$. Estimating discretely (and possibly randomly) observed renewal processes also falls into this problem, and that of estimating some queues is intimately related to it. These models are of great interest in a myriad of modern applications and we refer the reader to Daley and Vere-Jones [1998], Asmussen [2008] and Embrechts et al. [2013] for more details. Consequently, in the last decade and a half, research on nonparametric estimation of these distributions, and in particular of compound Poisson distributions, has been very active. In this section we review the literature that is most related to Coca [2015], focusing our attention on those works concerning estimation of compound Poisson distributions. In order of appearance, these are Bøgsted and Pitts [2010], Buchmann and Grübel [2003], Hansen and Pitts [2006], Duval [2013a], Comte et al. [2014] and Buchmann and Grübel [2004]. We heuristically discuss the construction of their estimators and we refer the reader to the respective works for the rigorous results.

2.1.2 Nonparametric estimation with the direct approach

Let ϱ_i denote the probability of M being equal to $i \in \mathbb{N}$ which, for a compound Poisson distribution with parameter $\lambda\Delta$, is $e^{-\lambda\Delta}\frac{(\lambda\Delta)^i}{i!}$. Assume the jumps $Y_1, Y_2 \ldots$ are i.i.d. with distribution F. Then, with probability ϱ_i, Z is the sum of i i.i.d. random variables with distribution F. If we denote the distribution function of Z by G, it means that

$$G(x) = \sum_{i=0}^{\infty} \varrho_i F^{*i}(x), \quad x \in \mathbb{R}, \tag{2.1.2}$$

where $*i$ denotes the i-fold convolution, and we write

$$F^{*i}(x) := \int_{-\infty}^{x} dF^{*i}, \quad x \in \mathbb{R}, \qquad \text{for any distribution function } F. \tag{2.1.3}$$

The mapping taking F to G following from display (2.1.2) is generally called the convolution operator. Hence, most of the literature in nonparametric estimation of the jump distribution of a compound distribution focuses on inverting this operator appropriately under different settings. In essence, this undoes the compounding operation in Z by untangling the series above and gives an expression for F in terms of G. Then, estimators can be constructed by estimating G from the observations $Z_1, \ldots, Z_n$ and plugging this into the corresponding expression for F.

Generally, and partly following the discussions in Bøgsted and Pitts [2010], the first step to invert the convolution operator is to write the equivalent expression to (2.1.2) for some transform such as the Fourier transform:

$$\mathcal{F}dG\,(u) = \sum_{i=0}^{\infty} \varrho_i \left(\mathcal{F}dF(u)\right)^i, \quad u \in \mathbb{R}, \tag{2.1.4}$$

or, more concisely,

$$\mathcal{F}dG = \Gamma(\mathcal{F}dF), \qquad \text{where } \Gamma(z) := \sum_{i=0}^{\infty} \varrho_i z^i, \quad z \in \mathbb{C}\,. \tag{2.1.5}$$

This is a power series with radius of convergence greater than or equal to 1, so (2.1.4) is well-defined because $\sup_{u\in\mathbb{R}} |\mathcal{F}dF(u)| \leq 1$. The idea then is to invert or 'revert' the series and take Fourier inverse transforms: define $\Gamma_0 := \Gamma - \varrho_0$ and $dG_0 := dG - \varrho_0\delta_0$, where δ_0 is Dirac's delta distribution at zero. Then we have that, formally, for some sequence of real numbers $\pi_1, \pi_2, \ldots$,

$$\mathcal{F}dF\,(u) = \sum_{i=1}^{\infty} \pi_i \left(\mathcal{F}dG_0(u)\right)^i =: \Gamma_0^{-1}(\mathcal{F}dG_0(u)), \quad u \in \mathbb{R}, \tag{2.1.6}$$

and the resulting expression is

$$dF = \mathcal{F}^{-1}\left[\Gamma_0^{-1}(\mathcal{F}dG_0)\right] = \sum_{i=1}^{\infty} \pi_i dG_0^{*i}, \tag{2.1.7}$$

where, in line with the definition in (1.1.3),

$$\mathcal{F}^{-1}f := \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\cdot u} f(u)du$$

for any f for which this exists in a distributional sense. Note that we have written dF

in terms of dG_0 and not of dG. In view of (2.1.1), the probability of Z being an empty sum is ϱ_0, so dG_0 represents the component of the distribution function of Z arising from sums of 1 or more jumps. This is the only component containing information about F and hence it is not surprising it features in (2.1.7) in place of dG.

Formal expression (2.1.7) is then the departure point of Buchmann and Grübel [2003], Hansen and Pitts [2006] and Bøgsted and Pitts [2010] to construct estimators of the jump distribution F, and of Duval [2013a] and Comte et al. [2014] to construct estimators of its density when it exists. The first ones to take this approach were Buchmann and Grübel [2003], who developed their results for compound Poisson distributions in order to estimate discretely observed compound Poisson processes, and named the problem *decompounding*. It was later extended for compound geometric distributions by Hansen and Pitts [2006] to estimate $M/G/1$-queues and, finally, Bøgsted and Pitts [2010] solved the problem for more general compound distributions. All three assumed knowledge of the compounding distribution M which, in the compound Poisson case, implies assuming knowledge of the intensity λ. However, we emphasise that this is not needed to use the direct approach and, indeed, Duval [2013a] and Comte et al. [2014] developed their results assuming no knowledge of λ.

Formally, (2.1.7) can be integrated to give

$$F(x) = \Lambda(G_0)(x) := \sum_{i=1}^{\infty} \pi_i G_0^{*i}(x), \quad x \in \mathbb{R}, \tag{2.1.8}$$

where we have used the notation introduced in (2.1.3). Then, if we have knowledge of the sequence $\pi_1, \pi_2, \ldots$, we can estimate G_0 from the observations $Z_1, \ldots, Z_n$, plug its expression into the last display and obtain an estimator of F. In order to estimate G_0 we need to estimate G and ϱ_0. The former can be achieved by the empirical distribution function introduced in Section 1.3.1. Yet, how to estimate ϱ_0 is not so clear: it is the probability of Z being the sum of no jumps, so the first idea is to estimate it by the proportion of zero observations. However, some zero observations may be the result of cancellations of several jumps since the mass of dG at zero can be larger than ϱ_0. It is easy to check that the probability of this happening when dF has no discrete component in $\mathbb{R}^+$ and/or $\mathbb{R}^-$ is zero. This includes the case of dF being supported only in $\mathbb{R}^+$ or $\mathbb{R}^-$ and dF being absolutely continuous with respect to Lebesgue's measure. Therefore, in such situations we can use the naive estimator

$$\tilde{\varrho}_{0,n} := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{0\}}(Z_k). \tag{2.1.9}$$

This is one of the reasons why Buchmann and Grübel [2003], Hansen and Pitts [2006] and Bøgsted and Pitts [2010] assume dF is supported in $\mathbb{R}^+$ although, as we mention later, it

is not the only one. With this assumption, let

$$\widetilde{G}_{0,n}(x) := \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{[0,x]}(Z_k) - \tilde{\varrho}_{0,n} := \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{(0,x]}(Z_k), \quad x \in \mathbb{R},$$

and, thus, F can be estimated by

$$\widetilde{F}_n(x) := \Lambda(\widetilde{G}_{0,n})(x) := \sum_{i=1}^{\infty} \pi_i \widetilde{G}_{0,n}^{*i}(x), \quad x \in \mathbb{R}\,. \tag{2.1.10}$$

Subject to the aforementioned specialisations of the sequence $\pi_1, \pi_2, \ldots$ and setting $\Delta = 1$ without loss of generality, this is the estimator Buchmann and Grübel [2003], Hansen and Pitts [2006] and Bøgsted and Pitts [2010] propose.

A priori, it is not clear that the infinite sum in (2.1.10) converges. In fact, it may well diverge as $x \to \infty$: noting that the mass of $d\widetilde{G}_{0,n}$ is $1 - \tilde{\varrho}_{0,n}$, the formal limit is

$$\lim_{x\to\infty} \widetilde{F}_n(x) := \sum_{i=1}^{\infty} \pi_i (1 - \tilde{\varrho}_{0,n})^i.$$

In general, the convergence of this series is not guaranteed, as it can be easily seen for the compound Poisson distribution with parameter $\lambda\Delta > 0$. In this case, $\pi_i = (-1)^{i+1}\frac{e^{i\lambda\Delta}}{\lambda\Delta i}$, so

$$\lim_{x\to\infty} \widetilde{F}_n(x) := \frac{1}{\lambda\Delta}\sum_{i=1}^{\infty} (-1)^{i+1} \frac{\left(e^{\lambda\Delta}(1 - \tilde{\varrho}_{0,n})\right)^i}{i}. \tag{2.1.11}$$

This is the famous Mercator series and it converges if $e^{\lambda\Delta}(1 - \tilde{\varrho}_{0,n}) \leq 1$. In view of the definition of $\tilde{\varrho}_{0,n}$ in (2.1.9), together with the assumption of dG having mass ϱ_0 at zero and the law of large numbers, $\tilde{\varrho}_{0,n}$ is arbitrarily close to ϱ_0 in sets of probability approaching 1 as $n \to \infty$. Noting that for the compound Poisson distribution $\varrho_0 = e^{-\lambda\Delta}$, we can thus assume the condition for the last display to converge (in sets of probability approaching 1 as $n \to \infty$) is that $e^{\lambda\Delta} - 1 < 1$. Equivalently, $\lambda\Delta < \log 2$, in which case,

$$\lim_{n\to\infty}\lim_{x\to\infty} \widetilde{F}_n(x) := \frac{\log(e^{\lambda\Delta})}{\lambda\Delta} = 1,$$

as expected. Intuitively, $\lambda\Delta$ is the expected number of jumps in an observation of Z. Due to $\log 2 < 1$, condition $\lambda\Delta < \log 2$ means that in many occasions we expect to directly observe the jumps whose distribution we are trying to estimate. These calculations also show that for the more interesting case of larger values of $\lambda\Delta$, (2.1.10) diverges as $x \to \infty$. The assumption of dF being supported in $\mathbb{R}^+$ allows Buchmann and Grübel [2003], Hansen and Pitts [2006] and Bøgsted and Pitts [2010] to circumvent this problem and to justify all of the steps we gave to arrive to (2.1.8) by paying some price: for any $\tau > 0$, they introduce the norm $\|f\|_{\infty,\tau} := \sup_{x\geq 0} e^{-\tau x}|f(x)|$ and the space $D_\tau := \{f :$

f càdlàg and $\lim_{x\to\infty} e^{-\tau x} f(x)$ exists$\}$. The metric space $(D_\tau, \|\cdot\|_{\infty,\tau})$ is Banach and weak convergence in it is with respect to the σ-algebra generated by its closed balls as defined in Pollard [1984], p. 199. Then, assuming τ is large enough, they prove that the series in (2.1.8) converges in D_τ, and so does that in (2.1.10) with probability tending to 1 as $n \to \infty$. Furthermore, they show that, as $n \to \infty$

$$\sqrt{n}\left(\widetilde{F}_n - F\right) \to \mathbb{B}^F \tag{2.1.12}$$

in distribution in $(D_\tau, \|\cdot\|_{\infty,\tau})$, where $\mathbb{B}^F$ is a centred Gaussian process with covariance structure

$$\Sigma^{\mathbb{B}^F}_{x,y} := \int_{[0,x]} \int_{[0,y]} G_0\left((x - z_1) \wedge (y - z_2)\right) H(dz_1) H(dz_2) - G_0 * H(x)\, G_0 * H(y), \quad x, y \geq 0,$$

with

$$H(x) := \sum_{i=1}^{\infty} i\pi_i G_0^{*(i-1)}(x), \qquad x \geq 0.$$

In their proofs they work with the Laplace transform $\mathcal{L}$ in place of Fourier's transform to justify the formal calculations above. For their (damped) Laplace transforms to be smaller than the radius of convergence of the series in (2.1.6), they need to assume τ is large enough. In the case of a compound Poisson distribution with parameter $\lambda\Delta$, the condition is

$$\mathcal{L}dF(\tau) := \int e^{-\tau x} dF(x) < \frac{\log 2}{\lambda\Delta}. \tag{2.1.13}$$

This agrees with our discussion above in which we argued that if $\lambda\Delta < \log 2$ then $\widetilde{F}_n$ is well-defined without having to introduce any dampening.

The results of Buchmann and Grübel [2003], Hansen and Pitts [2006] and Bøgsted and Pitts [2010] are important for several reasons: they were the first of their kind; they apply to very general compounding distributions M and make no further assumptions on F than it being supported in $\mathbb{R}^+$; and, in addition to resulting in $1/\sqrt{n}$-consistent estimation of F, confidence regions and goodness-of-fit tests for F can be developed. Nevertheless, they have some undesirable practical drawbacks: they assume knowledge of the sequence $\pi_1, \pi_2, \ldots$, or in particular of λ in the compound Poisson case; they require F to be supported in $\mathbb{R}^+$; and the convergence under the exponentially downweighted supremum norm $\|\cdot\|_{\infty,\tau}$ implies the resulting estimation for F is not valid away from the origin. In line with this, the confidence regions derived from their results are not 'bands' but they diverge exponentially away from the origin in general. In addition, and as we argue in Section 2.6 below, the covariance $\Sigma^{\mathbb{B}^F}_{x,y}$ does not coincide with the lower bound developed by Trabs [2015a] for the compound Poisson case. The qualitative assumption of the support of F being in $\mathbb{R}^+$ means the lower bound in the model of Buchmann and Grübel [2003] cannot be larger and hence their estimator is not asymptotically efficient. As mentioned in Section

1.3.2, this has the drawbacks that the size of the resulting confidence regions and the power of the goodness-of-fit tests following from their results are not optimal. Imposing some alternative but mild assumptions on F, and in the compound Poisson setting, in Section 2.2.2 we construct an estimator that avoids all these limitations and that is asymptotically efficient.

Duval [2013a] and Comte et al. [2014] make inference on a discretely observed (zero-mean) compound Poisson process with unknown, although bounded, intensity λ assuming dF is absolutely continuous with respect to Lebesgue's measure. Furthermore, they develop their results under the so-called high-frequency observation regime, in which $\Delta = \Delta_n \to 0$ and $n\Delta_n \to \infty$ as $n \to \infty$. This means that, as more data becomes available, jumps are directly observed in the limit and the number of these tends to infinity. Therefore the ill-posedness of the problem vanishes asymptotically and the problem should simplify to the classical density estimation problem from i.i.d. observations. Indeed, the recent results of Mariucci [2016] imply both experiments are asymptotically equivalent (in the Le Cam sense), which was already hinted at by the statement of the results of Duval [2013a] and Comte et al. [2014].

Their estimators are built upon the following observations. Under the assumption of dF having a density f and in view of (2.1.2), $dG = \varrho_0\delta_0 + (1-\varrho_0)g_0$ for a density g_0 and these densities are explicitly related by formal expression (2.1.7). Hence, recalling that $\pi_i = (-1)^{i+1}\frac{e^{i\lambda\Delta}}{\lambda\Delta i}$ and $\varrho_0 = e^{-\lambda\Delta}$, and due to $\Delta = \Delta_n \to 0$, we have that for a finite $I \in \mathbb{N}$

$$f(x) \approx \sum_{i=1}^{I+1} \frac{(-1)^{i+1}}{i} \frac{(e^{\lambda\Delta}-1)^i}{\lambda\Delta} g_0^{*i}(x), \quad x \in \mathbb{R}\,. \tag{2.1.14}$$

Two of the standard ways to estimate a density from i.i.d. observations are through wavelet and kernel estimators. Duval [2013a] and Comte et al. [2014] take these two approaches, respectively, and they input the resulting density estimates of g_0 on the right hand side of the last display to obtain an estimator of f. They also estimate the coefficients in the linear combination: Duval [2013a] plugs-in the estimator

$$\tilde{\lambda}_n := -\frac{1}{\Delta}\log(\tilde{\varrho}_{0,n}) \tag{2.1.15}$$

in them motivated by the fact that $\varrho_0 = e^{-\lambda\Delta}$ and that the naive estimator (2.1.9) is consistent because they assume dF has no atoms; and Comte et al. [2014] propose an alternative quantity that directly estimates the coefficients.

Below we make a comparison with the estimator of f of Comte et al. [2014], so we give a few details of it now. They estimate g_0^{*i} using the kernel estimators for convolutions of densities proposed in Chesneau et al. [2013]. Note that g_0 is the density of the conditional distribution of the increments given they are not zero. Thus, conditioning on the number of non-zero increments being $m = m(n) \in \mathbb{N}$ and denoting them by $\widetilde{Z}_1, \ldots, \widetilde{Z}_m$, the

empirical version of its characteristic function is

$$\varphi_{0,n}(u) := \frac{1}{m}\sum_{k=1}^{m} e^{iu\widetilde{Z}_k}, \quad u \in \mathbb{R}\,.$$

Then, for any $i \geq 1$ they propose the estimator

$$\widetilde{g_{0,n}^{*i}}(x) := \mathcal{F}^{-1}\left[(\varphi_{0,n})^i \mathbb{1}_{[-h_n^{-1},h_n^{-1}]}\right](x), \quad x \in \mathbb{R}, \tag{2.1.16}$$

for some appropriate choice of h_n, the so-called bandwidth, such that $h_n \to 0$ as $n \to \infty$. Note that the indicator function can be introduced into the i-th power and merged with $\varphi_{0,n}$. By the property that products of functions in the frequency domain become convolutions in the 'space' domain, we see that their estimator of f is essentially a plug-in estimator where g_0 in (2.1.14) is substituted by a kernel estimator.

After constructing these estimators, Duval [2013a] and Comte et al. [2014] assume the unknown density f belongs to a Besov and to a Sobolev ball, respectively, and show they are adaptive to its regularity with the $L^p(C)$, $p \geq 1$, and $L^2(\mathbb{R})$ losses, respectively, where $C \subseteq \mathbb{R}$ is compact. As anticipated, the minimax rates of the problem coincide with those of the classical density estimation problem. Notice that the estimators obtained by setting $I = 0$ are the classical density estimators, which is a natural choice since in the limit we are in the continuous observations regime of Section 1.3. Yet, this estimator only achieves the minimax rates of the problem if Δ_n decays sufficiently fast and this condition can be relaxed by larger choices of I. Lastly, note that due to $\lambda\Delta_n < \log 2$ for n large enough, the tail of the series in (2.1.8) and (2.1.10) are negligible, so there is no need to consider the spaces D_τ for those expressions to be well-defined. Therefore, in this observation regime, the estimator of Buchmann and Grübel [2003] and the estimators of F resulting from integrating the density estimators of Duval [2013a] and Comte et al. [2014] are very similar for finite samples.

The approach of inverting the convolution operator has been used by other authors to solve related problems. Antunes and Pipiras [2011] extend the results of Bøgsted and Pitts [2010] to unknown compounding distribution functions M. For this they need to assume they have i.i.d. observations of M, which departs from the 'know-nothing' approach we take here. Concerning density estimation, Duval [2013b] extends the results of Duval [2013a] to renewal processes and Comte et al. [2015] extend those in Comte et al. [2014] to mixed compound Poisson process. The difference of the latter with our setting is that the intensity λ is random and therefore, instead of assuming they have n observations of a path, they assume they have one discrete observation of n independent paths. This falls into the problem of estimating renewal processes that have been observed discretely at random. Also with the direct approach, Duval [2014] studies the identifiability of a discretely observed compound Poisson process when $\Delta = \Delta_n \to \infty$ as $n \to \infty$.

In addition to the estimator of F discussed above, Buchmann and Grübel [2003] constructed an estimator of the jump mass function assuming dF is purely discrete and supported in $\mathbb{N}\setminus\{0\}$. Its expression can be found in Section 4.3.2 and here we limit ourselves to giving the general idea to build it. Note that in this setting, dG is also purely atomic but with support in $\mathbb{N}$. The construction of their discrete estimator is similar to that of the general estimator of F discussed above: they depart from the so-called Panjer recursion, which expresses the mass of the atoms of dG recursively and in terms of those of dF (we emphasise that this recursion can be derived from (2.1.2)); they invert the series and recursively write the mass of the atoms of dF in terms of those of dG; finally, they input the empirical value of the mass of each atom of dG. We remark that, if in $\pi_i = (-1)^{i+1}\frac{e^{i\lambda\Delta}}{\lambda\Delta i}$ in (2.1.10) λ is substituted by $\tilde{\lambda}_n$, then, crucially, the expression for the estimator of the mass of an atom of dF we just constructed corresponds to the evaluation of (2.1.10) at the corresponding integer. Therefore, the cumulative sum of these corresponds to their general estimator of F in Buchmann and Grübel [2003] with the empirical version of its coefficients. Assuming the mass function of dF is in the space

$$\left\{ a \in [0,1]^{\mathbb{N}\setminus\{0\}} : \sum_{j=1}^{\infty} a_j = 1 \ \text{ and } \ \sum_{j=1}^{\infty} a_j^{1/2} < \infty \right\},$$

they prove a central limit theorem in this space with the norm $\sum_{j=1}^{\infty}|a_j|$ and identify the limiting centred Gaussian process. In Section 2.6 we show that, unlike $\Sigma^{\mathbb{B}^F}$, its covariance coincides with the lower bound developed by Trabs [2015a] as a consequence of the substitution of λ by $\tilde{\lambda}_n$ in the coefficients π_i. Efficiency of the estimator does not necessarily follow because we ignore whether the qualitative assumption of the distribution being discrete and supported in $\mathbb{R}^+$ makes the bound decrease. Note that, due to the connection with (2.1.10), whose series has alternating signs, the estimator of the mass of each atom is not necessarily contained in $[0,1]$. Subsequently, in Buchmann and Grübel [2004] they propose a transformation to circumvent this and a very similar estimator combining likelihood ideas with this transformation. They show the estimators converge at the parametric rate, and identify the limiting processes of the finite-dimensional distributions. These are no longer Gaussian, except when dF does not have compact support, in which case they coincide with the limiting Gaussian process in the discrete case considered by Buchmann and Grübel [2003].

2.2 The spectral approach

The spectral approach is a natural one when nonparametrically estimating discretely observed Lévy processes. Therefore, we briefly introduce these processes prior to introducing the approach.

2.2.1 Lévy processes

The definition of a Lévy process has already been introduced in Section 1.1 so we do not give it again. Instead, here we focus on giving intuition about them and on the general Lévy-Khintchine formula for their characteristic function.

Roughly speaking, a Lévy process is a generalisation of a compound Poisson process with drift (recall its expression in (1.1.1)) in that it contains an independent Brownian motion and an independent pure jump process with infinitely many small jumps in any finite interval. Due to this rich structure and their tractability they have proven very useful when modelling systems with more complicated random shocks than those observed in the applications mentioned in Section 1.1. They have been used in many fields and are particularly important in mathematical finance (cf. Cont and Tankov [2003] and references therein for instance).

The celebrated Lévy-Khintchine representation (cf. Chapter 2.8 in Sato [1999]) states that if $(X_t)_{t\geq 0}$ is a one-dimensional real-valued Lévy process then its characteristic function can be written as

$$\varphi_t(u) = \exp\left(t\left(ibu - \frac{\sigma^2 u^2}{2} + \int_{\mathbb{R}\setminus\{0\}} (e^{iux} - 1 - iux\mathbf{1}_{|x|\leq 1})\,\nu(dx)\right)\right), \tag{2.2.1}$$

where $b \in \mathbb{R}$ is a drift parameter, $\sigma^2 \geq 0$ is a diffusion parameter and ν, the Lévy measure, is a σ-finite nonnegative measure concentrated on $\mathbb{R}\setminus\{0\}$ satisfying

$$\int_{\mathbb{R}\setminus\{0\}} \left(1 \wedge x^2\right)\,\nu(dx) < \infty. \tag{2.2.2}$$

In view of (1.1.4), we readily see that our compound Poisson setting corresponds to $\sigma = 0$, ν being a finite measure, and $b = \gamma + \int (1 + x\mathbf{1}_{|x|\leq 1})\,\nu(dx)$. In fact, by the definition of a compound Poisson process, see Chapter 1.4 in Sato [1999], and by the Lévy-Khintchine representation above, ν has finite mass if and only if the jump component of the Lévy process it arises from is a compound Poisson process; this is called the finite activity case. The reason why ν is not a finite measure in the rest of Lévy processes is because in general it has a nonintegrable singularity at the origin controlled by condition (2.2.2). Indeed, it is this singularity that gives rise to the infinitely many small jumps in any finite interval mentioned above. This is referred to as the infinite activity case and is generally split into processes of (locally) finite variation and of (locally) infinite variation, depending on whether

$$\int_{\mathbb{R}\setminus\{0\}} (1 \wedge |x|)\,\nu(dx) \tag{2.2.3}$$

is finite or infinite, respectively. This division of the degree of activity of the small jumps can be further refined by introducing the Blumenthal–Getoor index of the process, defined

as

$$\inf\left\{p \geq 0 : \int_{\mathbb{R}\setminus\{0\}} (1 \wedge |x|^p)\ \nu(dx) < \infty\right\} = \inf\left\{p \geq 0 : \lim_{|u|\to\infty} \frac{\Psi(u)}{|u|^p} = 0\right\},$$

where Ψ is the so-called Lévy exponent and satisfies that $\varphi_t(u) =: \exp\left(t\Psi(u)\right), u \in \mathbb{R}$. Whilst the first expression is the usual definition of the index, the second characterisation provides us with more insight for statistical purposes: if the Blumenthal–Getoor index is strictly positive or if there is a Gaussian component in the Lévy process, the tails of the characteristic function decay exponentially fast in the sense of the last display. If the index and σ are zero, they decay polynomially fast if the infimum is not attained, and they are bounded below if it is attained. The last case of course corresponds to the compound Poisson process. The Blumenthal–Getoor index can be defined for the more general Feller processes and, for more details, we refer the reader to Sections 5.2 and 5.3 in Böttcher et al. [2014], and to Section 5 in Schilling [1998] for a proof of the second characterisation in the last display.

2.2.2 Nonparametric estimation with the spectral approach

In this section we are concerned with nonparametric estimation of the Lévy measure, or of quantities closely related to it, of a Δ-discretely observed Lévy process. We therefore assume we have observations $X_\Delta, \dots X_{\Delta n}$ for some $\Delta > 0$ and $n \in \mathbb{N}$, and, by property (ii) in page 5, we work with the i.i.d. increments $Z_k := X_{k\Delta} - X_{(k-1)\Delta}$, $k = 1, \dots, n$. Just as above, when we speak of the observations, we are referring to the increments.

The starting point when estimating compound distributions through the direct approach was the explicit expression (2.1.2) of the distribution of the increments in terms of the jump distribution. However, due to the richer structure of Lévy processes, no such general expression for their increments exists. Instead, a more natural way of relating the law of the increments with the defining parameters of the process is through the Lévy-Khintchine formula for its characteristic function given by (2.2.1) with $t = \Delta$. Intuitively, this expression has untangled at once the parameters implicitly contained in the observations. In the compound Poisson case, we immediately see this by comparing it to the much-less-informative relationship between them of the direct approach in (2.1.2). Consequently, the focus is no longer on decompounding (at least explicitly) but on how to isolate the influence that each parameter has on the characteristic function, or equivalently on the observations, and on how to invert it. As we conclude in Section 2.3, in the compound Poisson case this approach gives similar estimators but with more desirable properties that hold under milder assumptions. Prior to embarking on the construction of the estimators for this particular case, we discuss estimation for more general Lévy processes. We remark that the spectral approach we use in this section was first introduced by Belomestny and Reiß [2006].

In Section 2.2.1 we saw that, in general, the Lévy measure has a nonintegrable singular-

ity at the origin. Therefore, the only Lévy processes for which the cumulative distribution function of this measure exists are those whose jump component is a compound Poisson process. From now on, let us denote this quantity by

$$N(x) := \int_{-\infty}^{x} d\nu := \lambda F(x), \qquad x \in \mathbb{R}$$

(this is capital ν and should not be confused with the Poisson process $(N_t)_{t\geq 0}$ in the previous chapter or with a normal distribution $N(\cdot,\cdot)$). In the rest of the cases, existing literature in the field has considered functionals of the Lévy measure that in one way or another avoid its singularity at the origin. For instance, the generalised Lévy distribution

$$N^G(x) := \begin{cases} \int_{-\infty}^{x} d\nu & \text{if } x < 0, \\ \int_{x}^{\infty} d\nu & \text{if } x > 0, \end{cases}$$

and more general functionals that smooth out the singularity such as

$$\mathcal{N}(x) := \int_{-\infty}^{x} \rho(y) y^2 \nu(dy), \quad x \in \mathbb{R},$$

where, for some $C > 0$, $0 < \rho(x) \leq C\,(1 \wedge x^{-2})$ satisfies some regularity assumptions. Neumann and Reiß [2009] studied the ill-posedness of the inverse problem of estimating quantities of this type and showed its auto-deconvolution structure mentioned in Section 1.4.1. We note in passing that Chen et al. [2010] also studied the optimal rates of convergence in a related problem, obtaining the same results and others that complement them. Therefore, and in view of the remark about the decay of the characteristic function made at the end of Section 2.2.1, the parametric rate of convergence $1/\sqrt{n}$ may only be attained when σ and the Blumenthal–Getoor index are 0. Indeed, Nickl and Reiß [2012] proved a Donsker theorem for the generalised Lévy distribution N^G for processes whose characteristic function decays at most with a low polynomial order. These, of course, include the compound Poisson process so we now heuristically construct their estimator. By the remarks at the end of Section 2.2.1, these processes are (locally) of bounded variation, in which case the characteristic function of the increments can be written as

$$\varphi := \varphi_\Delta(u) = \exp\left(\Delta\left(ibu + \int_{\mathbb{R}\setminus\{0\}} (e^{iux} - 1)\,\nu(dx)\right)\right), \tag{2.2.4}$$

for some $b \in \mathbb{R}$ that for simplicity of the exposition we assume to be zero from now on. We note in passing that, in the compound Poisson case, φ corresponds to $\mathcal{F}dG$ of Section 2.1.2 and we use both notations interchangeably throughout. In view of condition (2.2.3) and assuming ν has finite first moment, the Lévy measure can be reached if we differentiate

the last display once: we have that

$$\frac{\varphi'(u)}{\varphi(u)} = i\Delta \int_{\mathbb{R}} e^{iux}\, x\, \nu(dx) =: i\Delta \mathcal{F}\left[\,\cdot\, \nu\right](u), \quad u \in \mathbb{R},$$

so, heuristically,

$$N^G(x) = \frac{1}{i\Delta} \int_{\mathbb{R}} f_x^{(G)}(y)\, \mathcal{F}^{-1}\left[\frac{\varphi'}{\varphi}\right](dy), \quad \text{where} \quad f_x^{(G)}(y) = \begin{cases} 1/y\, \mathbf{1}_{(-\infty,x]}(y) & \text{if } x < 0, \\ \\ 1/y\, \mathbf{1}_{[x,\infty)}(y) & \text{if } x > 0. \end{cases}$$

Nickl and Reiß [2012] assumed ν is absolutely continuous with respect to Lebesgue's measure and that $\mathcal{F}[\cdot\nu]$ is square integrable. Thus, for $|x| \geq \zeta$ with $\zeta > 0$ fixed, this expression for N^G is well-defined by Plancherel's theorem. The idea to construct the estimator is to substitute the characteristic function and its derivative by the empirical counterparts

$$\varphi_n(u) := \frac{1}{n} \sum_{k=1}^{n} e^{iuZ_k} \quad \text{and} \quad \varphi_n'(u) := \frac{i}{n} \sum_{k=1}^{n} Z_k e^{iuZ_k}, \quad u \in \mathbb{R}\,.$$

Indeed, Neumann and Reiß [2009] showed that φ_n and φ_n' converge uniformly to φ and φ' in compact intervals of slow enough increasing size if $\int (1 \vee |\cdot|^{\beta})\, d\nu < \infty$ and $\int (1 \vee |\cdot|^{2+\beta})\, d\nu < \infty$, respectively, for some $\beta > 0$. With this in mind, Nickl and Reiß [2012] introduced a kernel function K such that

$$\int_{\mathbb{R}} K(x)\, dx = 1 \quad \text{and} \quad \operatorname{supp}(\mathcal{F}K) \subseteq [-1, 1], \tag{2.2.5}$$

and a bandwidth $h_n \to 0$ as $n \to \infty$. Let $K_{h_n} := 1/h_n K(\cdot/h_n)$, which is called an approximation to the identity because in the limit as $n \to \infty$ it acts as the identity of the convolution operation or, equivalently, as the identity of the product in the frequency domain. With the conditions above, $\operatorname{supp}(\mathcal{F}K_{h_n}) = \operatorname{supp}(\mathcal{F}K(h_n\cdot)) \subseteq [-h_n^{-1}, h_n^{-1}]$, and their estimator is

$$\widehat{N}_n^G(x) = \frac{1}{i\Delta} \int f_x^{(G)}(y) \mathcal{F}^{-1}\left[\frac{\varphi_n'}{\varphi_n} \mathcal{F}K_{h_n}\right](y)\, dy, \qquad |x| \geq \zeta.$$

Assuming the finite $2+\beta$ moment condition, Nickl and Reiß [2012] showed this estimator is well-defined (asymptotically in probability) and, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{N}_n^G - N^G\right) \to^{\mathcal{D}} \mathbb{G}^G \quad \text{in } \ell^{\infty}\left(\mathbb{R} \setminus (-\zeta, \zeta)\right), \tag{2.2.6}$$

where $\mathbb{G}^G$ is a centred Gaussian process with covariance structure

$$\Sigma^{\mathbb{G}^G}_{x,y} := \frac{1}{\Delta^2}\int_{\mathbb{R}} \left(\left(\cdot f^{(G)}_x\right) * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](z)\right)\left(\left(\cdot f^{(G)}_y\right) * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](z)\right) P(dz),$$

with $|x|, |y| \geq \zeta$, $\varphi^{-1} := 1/\varphi$ and where P is the law of the increments (for the compound Poisson case it corresponds to dG of Section 2.1.2 and use both notations interchangeably). As Nickl and Reiß [2012] remarked, convolution with $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ is mathematically equivalent to deconvolution with P and hence the former is the key quantity in $\Sigma^{\mathbb{G}^G}_{x,y}$.

By the arguments above, the class of Lévy processes Nickl and Reiß [2012] considered is, essentially, the largest class for which such a result can hold. This has a few consequences: they had to stay away from the origin and, furthermore, they had to differentiate the characteristic function, hence having to assume ν has a density such that

$$\mathcal{F}[\,\cdot\,\nu] \in L^2(\mathbb{R}) \quad \text{and} \quad \int (|\cdot| \vee |\cdot|^{2+\beta})\, d\nu < \infty \ \text{ for some } \beta > 0.$$

In Section 3.2 they conjectured that for the compound Poisson case, where ν has no singularity at the origin, it should be possible to take $\zeta = 0$ and avoid the first condition of this display by adapting their method. We were not able to do this with their estimator because when $|x| \geq \zeta = 0$, $f^{(G)}_x$ is no longer square integrable so issues arise in the proofs. Instead, we constructed a different estimator which, moreover, allowed us to avoid the condition of ν having a density (although this can also be avoided with their estimator by the same idea we introduce in Chapter 3) and the moment condition. To construct it, we do not differentiate the characteristic function and only have to control φ_n so, from the arguments above, we only need $\int (1 \vee |\cdot|^{\beta})\, d\nu < \infty$ for some $\beta > 0$. In fact, in Chapter 3, we refine the result of Neumann and Reiß [2009] to only require a finite logarithmic moment. We remark that this simplification can only be attained when controlling φ_n and not any of its derivatives (at least with the arguments in Neumann and Reiß [2009]).

Prior to focusing on the compound Poisson case, we briefly discuss the other existing functional central limit theorems for general Lévy processes. Recall the high-frequency regime introduced in Section 2.1.2, in which $\Delta = \Delta_n \to 0$ and $n\Delta_n \to \infty$ as $n \to \infty$. In this setting, an increasingly large number of jumps is observed as $n \to \infty$ so, intuitively, the ill-posedness of the problem should decrease considerably in the limit and it should be possible to attain the parametric rate of convergence for a much larger class of processes than that considered in Nickl and Reiß [2012]. Indeed, Kappus and Reiß [2010] extended the results of Neumann and Reiß [2009] to this observation regime showing that, even when $\sigma^2 > 0$, the parametric rate can be attained if Δ_n is chosen appropriately. With this insight and the machinery developed in Nickl and Reiß [2012], Nickl et al. [2016] have recently constructed estimators of $\mathcal{N}$, and showed a functional central limit theorem under the standard supremum norm for, essentially, the largest class of processes for which such

a result can hold. They also include results for N^G but, for simplicity, here we only discuss those for $\mathcal{N}$. They construct two estimators, a naive one, or linear, based on the intuitive fact that the jumps are observed in the limit as $\Delta_n \to 0$, and a non-linear one based on the spectral approach. For the first, they use that the law of the increments $P = P_\Delta$ satisfies

$$\int_{\mathbb{R}} \rho(x) x^2 \frac{P_\Delta(dx)}{\Delta} \to \sigma^2 \rho(0) + \int_{\mathbb{R}} \rho(x) x^2 \nu(dx) \quad \text{as } \Delta \to 0,$$

for any continuous and bounded ρ, and, assuming $\sigma = 0$, the estimator is thus given by

$$\widetilde{\mathcal{N}}_n(x) := \frac{1}{n\Delta_n} \sum_{k=1}^{n} \rho(Z_k) Z_k^2 \mathbb{1}_{(-\infty,x]}(Z_k), \quad x \in \mathbb{R}\,.$$

For the second they do not assume $\sigma = 0$ and, noting that that the Lévy exponent satisfies

$$\Psi''(u) = \frac{\varphi_\Delta''(u)\varphi_\Delta(u) - (\varphi_\Delta)^2(u)}{\Delta(\varphi_\Delta)^2(u)} = -\sigma^2 - \mathcal{F}\left[\,.^2\nu\right](u),$$

the spectral estimator is given by

$$\widehat{\mathcal{N}}_n(x) := -\int_{-\infty}^{x} \rho(y) \mathcal{F}^{-1}\left[\left(\Psi_n'' + \hat{\sigma}_n^2\right) \mathcal{F} K_{h_n}\right](y) dy, \quad x \in \mathbb{R},$$

where Ψ_n'' is the empirical version of the fraction in the second to last display and $\hat{\sigma}_n^2$ is a consistent estimator of σ^2 with certain properties. Then, under some stronger assumptions than in Nickl and Reiß [2012], they show that, as $n \to \infty$,

$$\sqrt{n\Delta_n}\left(\widetilde{\mathcal{N}}_n - \mathcal{N}\right) \to^{\mathcal{D}} \mathbb{G}^{\mathcal{N}} \quad \text{and} \quad \sqrt{n\Delta_n}\left(\widehat{\mathcal{N}}_n - \mathcal{N}\right) \to^{\mathcal{D}} \mathbb{G}^{\mathcal{N}} \quad \text{in } \ell^\infty(\mathbb{R}),$$

where $\mathbb{G}^{\mathcal{N}}$ is a centred Gaussian process with covariance structure

$$\Sigma_{x,y}^{\mathbb{G}^{\mathcal{N}}} := \int_{-\infty}^{x \wedge y} z^4 \rho(z)^2 \nu(z) dz, \quad x, y \in \mathbb{R}\,. \tag{2.2.7}$$

In subsequent sections we also compare our results to this one. We note in passing that, prior to the results of Nickl and Reiß [2012] and Nickl et al. [2016], Buchmann [2009] showed a functional central limit theorem for general Lévy processes when continuous but perhaps incomplete observations of the process are available. For this, he also constructed a naive or linear estimator, staying away from the origin like in Nickl and Reiß [2012], and used the exponentially downweighted norms in Buchmann and Grübel [2003]. The limit process coincides, up to the modifications required by the weighting, with that in the second to last display when $\rho = 1$.

Let us now focus on the compound Poisson case and the low-frequency observation regime, i.e. when $\Delta > 0$ is fixed. As mentioned above, the key observation of the spectral

approach is that, in view of (1.1.4) taking $t = \Delta$ and $\gamma = 0$, (2.1.5) can be written as

$$\mathcal{F}dG(u) = \varphi_\Delta(u) = e^{\Delta(\mathcal{F}\nu(u)-\lambda)}, \quad u \in \mathbb{R}, \tag{2.2.8}$$

where we recall that $\nu := \lambda dF$. Therefore, this expression explicitly shows the relationship between G and F, and, just as in the estimators we just constructed, the challenge is to rigorously invert it. Crucially, and unlike Nickl and Reiß [2012] and Nickl et al. [2016], we can separate the terms in (2.2.4) due to the finiteness of ν and we do not need to differentiate the last display to isolate the Lévy measure. To do this, we take logarithms on both sides and, rearranging terms and taking the Fourier inverse, it can be formally inverted to give

$$\nu(dx) = \mathcal{F}^{-1}\left[\frac{1}{\Delta}\operatorname{Log}(\mathcal{F}dG) + \lambda\right](dx), \quad x \in \mathbb{R}\,. \tag{2.2.9}$$

Here, Log denotes the distinguished logarithm: if $\phi : \mathbb{R} \to \mathbb{C}\setminus\{0\}$ is a continuous function such that $\phi(0) = 1$ then $\operatorname{Log}(\phi)$ is the unique continuous function satisfying $\exp\big(\operatorname{Log}(\phi)(u)\big) = \phi(u)$ for all $u \in \mathbb{R}$ (the reader is referred to the proof of Theorem 7.6.2 in Chung [2001] for its construction). In terms of basic operations it enjoys the same properties of the standard logarithm so we postpone its analysis until the proofs in Chapter 3. For now, we mention that the logarithm in the last display is well-defined in view of (1.4.2), and that the reason to choose this logarithm is the following: the exponent on the right hand side of (2.2.8) is continuous; if we instead take any branch of the complex logarithm on both sides, the left side becomes discontinuous; therefore, the corresponding identity cannot be true unless we take the distinguished logarithm. Furthermore, to construct the estimators we input an empirical version of $\mathcal{F}dG$ into the right hand side of (2.2.9) so, if we take a branch of the complex logarithm, the resulting expression may end in a different winding number than that of the right hand side of the last display. This is avoided by use of the distinguished logarithm.

The first to use the spectral approach to estimate discretely observed compound Poisson processes were van Es et al. [2007]. They assume dF is absolutely continuous with respect to Lebesgue's measure and construct a kernel estimator for the jump density f as follows. Recall from Section 2.1.2 that, under this assumption, $dG = \varrho_0\delta_0 + (1-\varrho_0)g_0$, where $\varrho_0 = e^{-\lambda\Delta}$ and g_0 is the density of the non-zero observations. Then, for any $x \in \mathbb{R}$, the last display can be rewritten as

$$\begin{aligned} f(x) &= \mathcal{F}^{-1}\left[\frac{1}{\lambda\Delta}\operatorname{Log}\left(e^{\lambda\Delta}\mathcal{F}dG\right)\right](x) \\ &= \mathcal{F}^{-1}\left[\frac{1}{\lambda\Delta}\operatorname{Log}\left(1 + \left(e^{\lambda\Delta}-1\right)\mathcal{F}g_0\right)\right](x). \end{aligned} \tag{2.2.10}$$

In view of the first expression, the modulus of the arguments of the logarithms are con-

tained in $[e^{-\lambda\Delta}, e^{\lambda\Delta}]$, so the last expression is well-defined if we cut-off the tails of the logarithm. This can achieved by using the kernel function introduced above and, for all $x \in \mathbb{R}$,

$$\begin{aligned} f(x) &\approx \mathcal{F}^{-1}\left[\frac{1}{\lambda\Delta}\operatorname{Log}\left(1+\left(e^{\lambda\Delta}-1\right)\mathcal{F}g_0\mathcal{F}K_{h_n}\right)\right](x) \\ &= \mathcal{F}^{-1}\left[\frac{1}{\lambda\Delta}\operatorname{Log}\left(1+\left(e^{\lambda\Delta}-1\right)\mathcal{F}\left[g_0 * K_{h_n}\right]\right)\right](x). \end{aligned} \tag{2.2.11}$$

This suggests to estimate g_0 by the classical kernel density estimator using the non-zero observations, i.e. again conditioning on the number of non-zero increments being $m \in \mathbb{N}$ and denoting them by $\widetilde{Z}_1, \ldots, \widetilde{Z}_m$,

$$\tilde{g}_{0,n}(x) := \frac{1}{m}\sum_{k=1}^{m} K_{h_n}(x - \widetilde{Z}_k), \quad x \in \mathbb{R}\,.$$

Indeed, van Es et al. [2007] proceed in this way assuming knowledge of λ and that f possesses enough regularity, including belonging to a ball of Hölder's and Nikol'ski's functions. They show pointwise weak consistency and asymptotic normality of the resulting estimator, paralleling the classical results of kernel density estimation.

Formal expression (2.2.9) also leads to estimators of the distribution function F. This has recently been exploited by Coca [2015] as we describe next. If we assume dF has a density, one way of approximating $F(x)$ is by integrating (2.2.11) over $(-\infty, x]$. However, we take a slightly different route not assuming this on dF: notice that (2.2.9) can be formally rewritten as

$$\frac{1}{\Delta}\mathcal{F}^{-1}[\operatorname{Log}(\mathcal{F}dG)]\,(dx) = \nu(dx) - \lambda\delta_0(dx), \quad x \in \mathbb{R}\,. \tag{2.2.12}$$

The second summand on the right carries the influence of the observations of Z in which no jump took place. Therefore, if this display is integrated excluding the origin, it can be ruled out at once. Furthermore, due to the ongoing assumption of dF having no atom at the origin neither does $\nu := \lambda dF$ and, formally,

$$\frac{1}{\Delta}\int_{-\infty}^{x} \mathbb{1}_{\mathbb{R}\setminus\{0\}}\,\mathcal{F}^{-1}[\operatorname{Log}(\mathcal{F}dG)]\,(dy) = \int_{-\infty}^{x}\nu(dy) =: N(x), \quad x \in \mathbb{R}\,. \tag{2.2.13}$$

Using the same arguments as above, the logarithm is bounded above and below. However, there is no a priori reason to believe the Fourier inverse is well-defined. To circumvent this, we use a kernel function K with the properties in (2.2.5), although, due to its compact support, it cannot be introduced into the logarithm directly. Consequently, we estimate

the last display by

$$\widehat{N}_n(x) := \frac{1}{\Delta}\int_{-\infty}^{x} \mathbb{1}_{\mathbb{R}\setminus\{0\}}\, \mathcal{F}^{-1}[\mathrm{Log}(\mathcal{F}dG_n)\mathcal{F}K_{h_n}]\,(y)dy, \quad x \in \mathbb{R}, \tag{2.2.14}$$

where G_n is the empirical distribution function of G from the observations $Z_1, \ldots, Z_n$ and hence

$$\mathcal{F}dG_n(u) = \varphi_n := \frac{1}{n}\sum_{k=1}^{n} e^{iuZ_k}, \quad u \in \mathbb{R}\,.$$

Then, in view of $\lim_{x\to\infty} N(x) = \lambda$, Coca [2015] proposes the estimators

$$\hat{\lambda}_n := \widehat{N}_n(\infty) \quad \text{and} \quad \widehat{F}_n(x) := \frac{\widehat{N}_n(x)}{\hat{\lambda}_n},\ x \in \mathbb{R}\,. \tag{2.2.15}$$

Dividing $\widehat{N}_n$ by $\hat{\lambda}_n$ guarantees the desirable property of $\lim_{x\to\infty} \widehat{F}_n(x) = 1$ for any $n \in \mathbb{N}$. More importantly, and as we discuss in Section 2.6, it guarantees the resulting estimator of F is efficient, whilst dividing by the naive estimator $\tilde{\lambda}_n$ does not when we allow for discrete jumps that can cancel each other. We remark that the estimators Coca [2015] proposes, included in the next chapter, differ slightly from the quantities we just proposed and for rigorous and practical purposes should be used instead: notice that, in practice, implementing $\mathbb{1}_{\mathbb{R}\setminus\{0\}}$ is not possible because of the truncation at a zero measure set and the infinite integration interval it results in. Therefore, this indicator is substituted by the more realistic quantity $\mathbb{1}_{[-H_n,H_n]\setminus(-\varepsilon_n,\varepsilon_n)}$, where $H_n^{-1}, \varepsilon_n \to 0$ as $n \to \infty$. Furthermore, we modify the resulting estimator to accommodate a potential atomic component in dF but omit it for now for the sake of clarity of exposition. In a few paragraphs we give details of how we estimate the discrete component, and the full estimator including both is only introduced in Chapter 3. We note in passing that it allows us to construct the first test of whether there is a discrete and/or an absolutely continuous component. Similarly, the estimators of Duval [2013a], Comte et al. [2014] and van Es et al. [2007] slightly differ from those we constructed above. For the properties that concern us here it suffices to take these simpler versions.

To prove our results we require the following mild assumptions on F. We assume $F = F_{ac} + F_d$, where dF_{ac} and dF_d are absolutely continuous and purely atomic with respect to Lebesgue's measure, respectively. On the absolutely continuous function F_{ac}, we assume it is uniformly Hölder continuous over $\mathbb{R}$, and on the discrete measure dF_d we assume it is supported in $\varepsilon \times \mathbb{Z} \setminus \{0\}$ for some $\varepsilon > 0$ fixed. We can relax the former to F_{ac} having a modulus of continuity as slow as a power of a logarithm as found in Coca [2015]. However, this comes at the expense of more involved statements of the results and more cumbersome proofs, so we choose to present this assumption instead. The exclusion of the origin in the latter goes in line with the ongoing and unavoidable assumption that dF has no atom at zero, as otherwise the problem lacks identifiability. Taking $\varepsilon = 1$ we include the

important-in-practice case of jumps taking values in the integers. We emphasise that the estimators can be used when one of the two components F_{ac} and F_d is zero and without prior knowledge of this. Lastly, we assume all the jumps have finite $\log^\beta$-tail moment for some $\beta > 2$. We remark that one of the reasons to be able to prove our results under this mild tail condition instead of a β-polynomial one, $\beta > 0$, is the truncation of the tails resulting from inputting $\mathbb{1}_{[-H_n,H_n]\setminus(-\varepsilon_n,\varepsilon_n)}$ in place of $1_{\mathbb{R}\setminus\{0\}}$ in the definition of $\widehat{N}_n$.

Our results are closest to those of Buchmann and Grübel [2003], so we compare our assumptions to theirs. We do not assume knowledge of λ —furthermore, we can include and estimate an unknown drift term, although we postpone this until Section 2.4— and we do not require one-sidedness in the jump distribution. However, we do need the mild assumptions of the previous paragraph. On the other hand, just with these we are able avoid using the exponentially downweighted norms they use and to develop our results under the norm $\|\cdot\|_\infty := \|\cdot\|_{\infty,0}$. In particular, we show that, as $n \to \infty$, our estimators are well-defined in sets of probability approaching 1 and

$$\sqrt{n}\left(\widehat{F}_n - F\right) \to^{\mathcal{D}} \mathbb{G}^F \quad \text{ in } \ell^\infty(\mathbb{R}), \tag{2.2.16}$$

where, under the notation $\varphi = \varphi_\Delta = \mathcal{F}dG$, $\mathbb{G}^F$ is a centred Gaussian process with covariance structure

$$\Sigma^{\mathbb{G}^F}_{x,y} := \frac{1}{\Delta^2}\int_{\mathbb{R}} \left(f^{(F)}_x * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](z)\right)\left(f^{(F)}_y * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](z)\right) dG(z), \quad x, y \in \mathbb{R},$$

and $f^{(F)}_x := \lambda^{-1}\left(\mathbb{1}_{(-\infty,x]} - F(x)\right)\mathbb{1}_{\mathbb{R}\setminus\{0\}}$, $x \in \mathbb{R}$. As pointed out by Nickl and Reiß [2012] in the compound Poisson case, $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ is a finite signed measure on $\mathbb{R}$ satisfying

$$\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right] = e^{\lambda\Delta}\sum_{i=0}^{\infty}(-1)^i\frac{(\lambda\Delta)^i}{i!}d\bar{F}^{*i}, \tag{2.2.17}$$

where $d\bar{F}(A) := dF(-A)$ for all $A \subseteq \mathbb{R}$ Borel measurable. This, and the fact that it has mass equal to 1, follows by similar arguments to those used to justify

$$dG = e^{-\lambda\Delta}\sum_{i=0}^{\infty}\frac{(\lambda\Delta)^i}{i!}dF^{*i}, \tag{2.2.18}$$

(see Remark 27.3 in Sato [1999]).

Similarly to classical Donsker's theorem and to (2.1.12), this result allows us to make $1/\sqrt{n}$-consistent estimation of F, and to derive confidence regions and goodness-of-fit tests for it. In contrast with those following from (2.1.12), these are for the supremum norm so the convergence of $\widehat{F}_n$ is uniform in $\mathbb{R}$ and the confidence regions are bands valid for the whole of $\mathbb{R}$. The relationship between the limiting processes $\mathbb{G}^F$ and $\mathbb{B}^F$ in (2.1.12) is

made clear in Section 2.6. For now we just mention that, due to the choice of estimator of λ and to the truncation at the origin in the definition of $\widehat{N}_n$, which is carried to $f_x^{(F)}$, $\Sigma_{x,y}^{\mathbb{G}^F}$ coincides with the lower bound developed by Trabs [2015a]. Therefore, our estimators are asymptotically efficient and they are the first to be so in the context of estimating the jump distribution function F of a compound Poisson distribution. As mentioned in Section 1.3.2, the importance of this is that the size of the confidence regions and the statistical power of the inference procedures that follow from the functional central limit theorem in (2.2.16) are optimal.

In Chapter 3, we construct estimators of the mass of the potential atoms in F. The idea to estimate the mass at εj, $j \in \mathbb{Z}\setminus\{0\}$, is simple: in the same way that we integrated (the empirical kernel-regularised version of) (2.2.12) up to x to reach $N(x)$ and then divided by $\hat{\lambda}_n$, we integrate (2.2.12) in a small interval $[\varepsilon j - \varepsilon_n, \varepsilon j + \varepsilon_n]$ with $\varepsilon_n \to 0$, divide by $\hat{\lambda}_n$ and introduce the kernel to ensure the resulting expression is well-defined. We note in passing that this shows the relationship between the notation ε_n and ε, which should not be confused, and without loss of generality $\varepsilon_n < \varepsilon$. Therefore, under the assumption of F only having an atomic component in $\varepsilon \times \mathbb{Z}\setminus\{0\}$, the estimator of F obtained by adding up these estimators is equal to $\widehat{F}_n$, up to exponentially negligible terms. In addition to asymptotic normality of the estimator of the mass of each of the atoms, in Chapter 3 we show joint convergence and a functional central limit theorem for the mass function follows. The only regularity assumption we make is the abovementioned finite logarithmic moment condition.

Lastly, we remark that the spectral approach has also been used when estimating the jump densities of discretely observed processes from superclasses of the compound Poisson process. The general idea, again, is to differentiate the characteristic function depending on the complexity of the underlying process and to assume the density belongs to a certain functional space. The most complete reference is Belomestny et al. [2015], which includes some but not all of the following works. Gugushvili [2009, 2012] extended the results of van Es et al. [2007] in several directions, including adding a diffusion component to the unknown compound Poisson process. Chen et al. [2010] generalised the noise component to a symmetric stable process. Assuming the logarithm of financial assets is a compound Poisson process with drift and with a diffusion component, Cont and Tankov [2003], Belomestny and Reiß [2006], Söhl [2014] and Söhl and Trabs [2014] perform density estimation and develop confidence sets using prices of financial options. In the case of a pure jump Lévy process of finite variation, Comte and Genon-Catalot [2009] and Bec and Lacour [2015] make adaptive estimation in the high-frequency regime, and Comte and Genon-Catalot [2010a,b] obtain the same results in the low-frequency regime when equispaced or irregularly sampled and noisy observations are available, respectively. In the high-frequency regime, Comte and Genon-Catalot [2011] extend the results of Comte and Genon-Catalot [2009] to a considerably larger class of Lévy processes. In addition,

the spectral approach has also been used to address other relevant statistical problems such as testing the characteristics of a Lévy process by Reiß [2013] and adaptive quantile estimation by Trabs [2015b].

2.3 Unifying the two approaches

2.3.1 The formal equivalence

We just discussed the two main approaches to estimate the jump distribution of a discretely observed compound Poisson process. As mentioned at the start of each of the sections, in essence, they focus on tackling two fundamentally different problems (although they implicitly tackle both): untangling the parameters, and rigorously inverting the identities that relate them to the data, respectively. For this reason they lead to estimators for which different conclusions are obtained assuming distinct assumptions. However, as it may already be clear, the formal identities from where they are derived are formally equivalent and hence all the estimators are very similar to each other. Let us make this more precise for the Poisson case.

In view of expression (2.1.5) and recalling that $\varrho_i = e^{-\lambda\Delta}\frac{(\lambda\Delta)^i}{i!}$, $i \in \mathbb{N}$, or equivalently due to (2.2.8),

$$\Gamma(z) = e^{\lambda\Delta(z-1)} \quad \text{and} \quad \Gamma_0(z) := \Gamma(z) - \varrho_0 = e^{-\lambda\Delta}\left(e^{\lambda\Delta z} - 1\right), \quad z \in \mathbb{C}\,.$$

Therefore,

$$\Gamma_0^{-1}(z) = \frac{1}{\lambda\Delta}\operatorname{Log}(1 + e^{\lambda\Delta}z),$$

where Log should be the distinguished logarithm if $z = z(\theta)$ is a continuous path in $\mathbb{C}\setminus\{0\}$ for the same reasons as those mentioned after (2.2.9). Inputting $z = \mathcal{F}dG_0$ into the last display, where we recall that $dG_0 := dG - \varrho_0\delta_0$, (2.1.6) can be alternatively written as

$$\mathcal{F}dF(u) = \frac{1}{\lambda\Delta}\operatorname{Log}(1 + e^{\lambda\Delta}\mathcal{F}dG_0(u)), \quad u \in \mathbb{R}\,. \tag{2.3.1}$$

The estimators of the direct approach were derived from a formal transformation of this expression, and so were those of the spectral approach in view of (2.2.10) and of its equivalent form (2.2.9). Therefore, this display shows that both approaches are formally equivalent and it explicitly highlights their distinction: the former differs from the latter in that, prior to taking Fourier inverse transforms, it insists on expanding the logarithm instead of working directly with it. This difference determines the conditions to be imposed and results in the superiority of the spectral approach: crucially, as it stands, the right hand side of the last display is well-defined for all $u \in \mathbb{R}$. This is because the argument of the logarithm equals $e^{\lambda\Delta}\mathcal{F}dG(u)$ by the definition of dG_0, and due to the strictly positive lower bound for $\mathcal{F}dG(u) = \varphi_\Delta$ in (1.4.2). However, the expansion of the logarithm is not always

justified and further assumptions are needed to embark on this route. This is the content of our discussion after (2.1.10) and the last display further clarifies it: taking $t = \Delta$ and $\gamma = 0$ in (1.1.5) it follows that, if $\lambda\Delta < \pi$, the distinguished logarithm in the last display coincides with the principal branch of the complex logarithm[1] because the argument does not wind about the origin. Furthermore, if $\lambda\Delta < \log 2$, $|e^{\lambda\Delta}\mathcal{F}dG_0(u)| \leq 1$ because the mass of dG_0 is $1 - e^{-\lambda\Delta}$, and just in this case the logarithm can be rigorously expanded using Mercator's complex series. For larger values of $\lambda\Delta$ this expansion only makes sense if the Laplace transform $\mathcal{L}$ as defined in (2.1.13) is used assuming the support of dF, or equivalently of dG_0, is in $\mathbb{R}^+$ and taking $u > 0$ large enough so that $|e^{\lambda\Delta}\mathcal{L}dG_0(u)|$ is smaller or equal to the radius of convergence of the series. Indeed, the precise condition is (2.1.13), which Buchmann and Grübel [2003], Hansen and Pitts [2006] and Bøgsted and Pitts [2010] assumed. Notice that, after this, the direct approach still has to invert the last display just as in the spectral approach. This justifies the fact that, at least in the low-frequency regime, the former has to make stronger assumptions than the latter and, thus, the spectral approach should be preferable in these problems. An alternative way of phrasing this is that, even though the approaches are formally equivalent, there is a fundamental difference between them: the direct approach insists on explicitly decompounding the sum in Z or, equivalently, untangling relationship (2.1.3), whilst in the spectral approach this is achieved at once by resorting to the frequency domain, hence requiring milder assumptions.

Consequently, this raises a natural and interesting question: in the context of general compound distributions, is it possible to extend the result in (2.2.16) by working with Γ_0^{-1}, as in the spectral approach, instead of with its expansion or with Λ in (2.1.8), as in the direct approach? Moreover, can (2.2.16) be extended to more general inverse problems in which the distribution of the observations G is related to that of interest F by $\mathcal{F}dG = \Gamma(\mathcal{F}dF)$ for some Γ (the starting point of the last paragraph)? The only result in this direction of which we are aware is that of den Boer and Mandjes [2016]. They work with the Laplace transform assuming the supports of dF and dG are in $\mathbb{R}^+$, and show pointwise consistency of their estimator at rate $\log n/\sqrt{n}$ under much stronger assumptions than those used for (2.2.16). Comparing their results to (2.2.16), they seem far from optimal and, thus, the problems we just proposed remain very much open. We emphasise that some preliminary calculations suggest they should hold under appropriate conditions on Γ. We can further ask under what assumptions on Γ these hold for the multidimensional and/or noisy case. Although we do not include the proofs here, (2.2.16) does extend to the multidimensional and noisy compound Poisson case under the same (multidimensional-analogous) assumptions.

[1] This was missed by van Es et al. [2007], who mention this property holds for $\lambda\Delta < \log 2$ and restrict their simulations to the not-so-interesting case of $\lambda\Delta = 0.3$ to avoid the implementation of the distinguished logarithm. However, we remark that the practical construction of their estimators in Section 3 in van Es et al. [2007] is also valid for $\lambda\Delta < \pi$, at least in sets of probability approaching 1 as $n \to \infty$.

2.3.2 Similarities between existing estimators

With the comparison of the two approaches made above, we can explicitly compare the estimators derived from them. In view of (2.2.13), Coca [2015] exploits the fact that

$$F(x) = \frac{1}{\lambda\Delta}\int_{-\infty}^{x} \mathbb{1}_{\mathbb{R}\setminus\{0\}}\, \mathcal{F}^{-1}[(\mathrm{Log}(\mathcal{F}dG))]\,(dy).$$

Furthermore, and as argued therein, the right hand side equals

$$\begin{aligned}\frac{1}{\lambda\Delta}\int_{-\infty}^{x} \mathcal{F}^{-1}[(\mathrm{Log}(\mathcal{F}dG)+\lambda\Delta)]\,(dy) &\approx \frac{1}{\lambda\Delta}\int_{-\infty}^{x} \mathcal{F}^{-1}[(\mathrm{Log}(\mathcal{F}dG)+\lambda\Delta)\,\mathcal{F}K_h](dy)\\ &= \frac{1}{\lambda\Delta}\int_{-\infty}^{x} \mathcal{F}^{-1}\Big[\mathrm{Log}\left(1+e^{\lambda\Delta}\mathcal{F}dG_0\right)\mathcal{F}K_h\Big](dy),\end{aligned}$$

where the last equality follows by the definition of dG_0. We remark that, unlike its series representation, the last expression is well-defined because (2.3.1) is, due to the spectral cut-off induced by $\mathcal{F}K_h$ and to the fact that $\|K_h\|_1 = \|K\|_1 < \infty$ in view of (2.2.5). Formally expanding the logarithm and rearranging terms, the last display can be written as

$$\begin{aligned}\int_{-\infty}^{x}\sum_{i=1}^{\infty}(-1)^{i+1}\frac{e^{i\lambda\Delta}}{i\lambda\Delta}\mathcal{F}^{-1}\left[\mathcal{F}dG_0^i\mathcal{F}K_h\right](dy) &= \int_{-\infty}^{x}\left(\sum_{i=1}^{\infty}(-1)^{i+1}\frac{e^{i\lambda\Delta}}{i\lambda\Delta}dG_0^{*i}\right)*K_h(dy)\\ &= \left(\sum_{i=1}^{\infty}(-1)^{i+1}\frac{e^{i\lambda\Delta}}{i\lambda\Delta}G_0^{*i}\right)*K_h(x).\end{aligned}$$

The last expression shows that our estimator can be interpreted as a kernel-regularised version of that in Buchmann and Grübel [2003] with the novelty that we can estimate a discrete jump component and that λ is substituted by an (efficient) estimator of it. On the left hand side, assume $dG_0 = (1-e^{-\lambda\Delta})g_0$ for some density g_0 and take $\mathcal{F}K = \mathbb{1}_{[-1,1]}$. Then, in view of (2.1.16), this expression makes clear that our estimator is an integrated low-frequency version of that in Comte et al. [2014], but in which the estimator for the coefficients is a plug-in estimator using (an efficient) one for λ.

Let us comment on the estimator of van Es et al. [2007]. Under the density assumption, they exploit the fact that

$$f(x) = \frac{1}{\lambda\Delta}\mathcal{F}^{-1}\Big[\mathrm{Log}\left(1+(e^{\lambda\Delta}-1)\mathcal{F}g_0\right)\Big](x) \approx \frac{1}{\lambda\Delta}\mathcal{F}^{-1}\Big[\mathrm{Log}\left(1+(e^{\lambda\Delta}-1)\mathcal{F}g_0\mathcal{F}K_h\right)\Big](x),$$

where we again remark that, unlike its series representation, the last expression is also well-defined by for the same reasons the second to last display was. Formally arguing as

before, the last display can be written as

$$\sum_{i=1}^{\infty}(-1)^{i+1}\frac{\left(e^{\lambda\Delta}-1\right)^{i}}{i\lambda\Delta}\left(g_0 * K_h\right)^{*i} = \sum_{i=1}^{\infty}(-1)^{i+1}\frac{e^{i\lambda\Delta}}{i\lambda\Delta}\left(dG_0 * K_h\right)^{*i}.$$

In view of the right hand side, the estimator of van Es et al. [2007] can be interpreted as the density version of the estimator of Buchmann and Grübel [2003], but in which the kernel density estimator is plugged-in in place of the empirical distribution function. Furthermore, in view of the left hand side and by the discussion right after expression (2.1.16), the estimator of Comte et al. [2014] is the high-frequency version of that in van Es et al. [2007] with $\mathcal{F}K = \mathbb{1}_{[-1,1]}$ assuming no knowledge of λ. With this interpretation, the estimator in Duval [2013a] is the wavelet density version of them in the high-frequency regime. Since using a kernel or wavelet estimator does not affect the asymptotic results, we conclude that, formally, all the estimators are very close relatives of each other. When $\lambda\Delta < \log 2$ they behave similarly in practice. As we point out in Section 2.6, the true distinction between them lies in whether they estimate λ or not (regardless of whether knowledge of it is available or not), as their asymptotic properties do depend on this.

Recall that at the end of Sections 2.1.2 and 2.2.2 we also briefly mentioned the construction of the estimators of Buchmann and Grübel [2003] and Coca [2015] for the jump mass function. In particular, assuming dF only has a discrete component, we remarked that the cumulative sum of the former coincides with $\widetilde{F}_n$ when an estimate of λ is plugged-in, and that of the latter is, up to negligible terms, equal to $\widehat{F}_n$. By the arguments of the previous paragraph, they are thus formally equal up to convolution with the kernel introduced in Section 2.2.2. Indeed, as we mention in Section 2.6, the covariance of the limiting processes in their central limit theorems is the same and, furthermore, it coincides with the information lower bound of the estimation problem when no restriction on the support of F is made. Buchmann and Grübel [2003] restricted the support to $\mathbb{R}^+$ and Trabs [2015a] assumed the Lévy measure is absolutely continuous with respect to Lebesgue's measure, so efficiency cannot be concluded but, if such changes in the model do not change the information bounds, both will be efficient.

We remark that, with this unified view, all the estimators can be interpreted as arising from an infinite sum of terms with alternating signs. This corresponds to their attempt to invert the compounding operation and justifies the common feature that the resulting estimates of the distribution are not monotonic. In other words, this feature does not depend on the type of estimator used such as a kernel one and, a priori, seems unavoidable without any further transformation such as the ones Buchmann and Grübel [2004] made in the purely atomic case mentioned at the end of Section 2.1.2. Due to the close relationship between the decompounding problem and the estimation of discretely observed Lévy processes, this also justifies the appearance of this feature in the estimators of functionals of the Lévy measure.

Even though the estimators of Nickl and Reiß [2012] and Nickl et al. [2016] were derived for more general Lévy processes, a natural question is how they compare to that in Coca [2015] when the underlying process is compound Poisson. These also turn out to be the same, except for exponentially negligible quantities, although we postpone the comparison until Section 2.5 for reasons that will become apparent there. Lastly, we remark that a third approach has recently been introduced to nonparametrically estimate the density of a discretely observed compound Poisson process. Gugushvili et al. [2015] and Gugushvili et al. [2016] use a Bayesian approach and are able to tackle the multidimensional case in the low- and the high-frequency regimes, respectively. Their methodology considerably differs from the one we deal with here and therefore refer the reader to their works for more details.

2.4 Estimators of the intensity and the drift

In this section we review existing estimators of the intensity and of the drift of a compound Poisson distribution. As pointed out by Coca [2015] and further expanded in Section 2.6, the estimation of the former plays a key role in the efficiency of the estimators of the jump distribution. In Section 2.4.1, we introduce the estimators of λ, describing their differences and similarities, and discussing their asymptotic properties. In Section 2.4.2 we discuss the estimators of γ, up to which point we still assume $\gamma = 0$.

2.4.1 Estimating the intensity

In the previous section we already intuitively motivated two estimators of the intensity: the naive one in (2.1.15),

$$\tilde{\lambda}_n := -\frac{1}{\Delta}\log\left(\tilde{\varrho}_{0,n}\right) := -\frac{1}{\Delta}\log\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{\{0\}}(Z_k)\right);$$

and the one derived from the spectral approach in (2.2.15) which, including the modifications discussed therein, is given by

$$\hat{\lambda}_n := \widehat{N}_n(\infty) := \frac{1}{\Delta}\int_{\mathbb{R}} \mathbb{1}_{[-H_n,H_n]\setminus(-\varepsilon_n,\varepsilon_n)}(x)\mathcal{F}^{-1}[\mathrm{Log}(\mathcal{F}dG_n)\mathcal{F}K_{h_n}](x)dx,$$

where $\mathcal{F}dG_n := \frac{1}{n}\sum_{k=1}^{n} e^{i\cdot Z_k}$ is the empirical characteristic function and $h_n, \varepsilon_n, H_n^{-1} \to 0$ as $n \to \infty$. Yet, a third estimator can be constructed through the spectral approach: recall that we motivated the form of $\hat{\lambda}_n$ by the fact that the mass of dN is λ. In addition, we argued that in the construction of $\widehat{N}_n$ we ruled out the influence of the observations of Z in which no jump takes place. We achieved this by introducing the truncation at the origin that annihilates the second summand in (2.2.12). Using the same argument, it should be

possible to exploit that second summand to isolate the influence of such observations of Z. Indeed, an alternative estimator for λ is given by

$$\check{\lambda}_n := -\frac{1}{\Delta}\int_{\mathbb{R}} \mathbb{1}_{[-\varepsilon_n,\varepsilon_n]}(x)\mathcal{F}^{-1}[\mathrm{Log}(\mathcal{F}dG_n)\mathcal{F}K_{h_n}](x)dx.$$

Just as the estimators of the jump distribution from the previous section were close relatives of each other, these are so too. Furthermore, the main difference between them also is the asymptotic efficiency. Recall that the expression for $\tilde{\lambda}_n$ was motivated by the fact that the probability of observing an instance of Z with no jumps is $e^{-\lambda\Delta}$. Let us condition on the number of observations with jumps being $m \in \mathbb{N}$ and on the number of non-zero such observations being $\tilde{m} \in \mathbb{N}$. Then, we have that

$$\tilde{\lambda}_n := -\frac{1}{\Delta}\log\frac{n-\tilde{m}}{n} = -\frac{1}{\Delta}\log\frac{n-m}{n} - \frac{1}{\Delta}\log\frac{n-\tilde{m}}{n-m}.$$

This explicitly shows the issue pointed out in the previous section: the first summand on the right hand side is a consistent estimator for λ so, if jumps can cancel out resulting in zero observations, the second summand converges to a non-zero quantity and $\tilde{\lambda}_n$ cannot be consistent. As mentioned when we introduced $\tilde{\varrho}_{0,n}$ in (2.1.9), if dF has no discrete component in $\mathbb{R}^+$ and/or $\mathbb{R}^-$, the probability of such cancellations is zero. Then, $e^{-\lambda\Delta}$ is all the mass of dG at zero and $\tilde{\lambda}_n$ is asymptotically consistent. In any case, due to the intuition behind $\check{\lambda}_n$, this new estimator should be related to $\tilde{\lambda}_n$. Indeed, denoting the observations of Z with jumps by $\tilde{Z}_1, \ldots, \tilde{Z}_m$,

$$\begin{aligned}
\check{\lambda}_n &= -\frac{1}{\Delta}\int \mathbb{1}_{[-\varepsilon_n,\varepsilon_n]}(x)\mathcal{F}^{-1}\left[\mathrm{Log}\left(\frac{n-m}{n}\left(1+\frac{n}{n-m}\frac{1}{n}\sum_{k=1}^{m}e^{i\cdot\tilde{Z}_k}\right)\right)\mathcal{F}K_{h_n}\right](x)dx \\
&= \left(\tilde{\lambda}_n + \frac{1}{\Delta}\log\frac{n-\tilde{m}}{n-m}\right)\int_{-\varepsilon_n/h_n}^{\varepsilon_n/h_n} K(x)dx \\
&\quad - \frac{1}{\Delta}\int \mathbb{1}_{[-\varepsilon_n,\varepsilon_n]}(x)\mathcal{F}^{-1}\left[\mathrm{Log}\left(1+\frac{n}{n-m}\mathcal{F}d\widetilde{G}_{0,n}\right)\mathcal{F}K_{h_n}\right](x)dx. \qquad (2.4.1)
\end{aligned}$$

In Chapter 3 we assume $\varepsilon_n/h_n \to \infty$ as $n \to \infty$ so, by the first property of K in (2.2.5), the integral in the first summand converges to 1 as $n \to \infty$. In view of (2.3.1), the second summand is estimating the mass of $\nu := \lambda dF$ at the origin, which is 0 by assumption, and is therefore negligible. This shows that $\check{\lambda}_n$ gets rid of the summand in $\tilde{\lambda}_n$ that may make it lack consistency. More importantly, it explicitly reflects the power of the spectral approach in which, as mentioned at the beginning of Section 2.2, the untangling problem is solved at once: $\check{\lambda}_n$ is able to distinguish between the zero observations coming from empty sums in Z and those that result from cancellations of jumps; this can only be because $\check{\lambda}_n$ is able to decompound the zero observations, unlike $\tilde{\lambda}_n$. As expected, below we see that, in general, their asymptotic properties are different. Nonetheless, when the naive estimator

is consistent, the two are equal except for negligible as $n \to \infty$ terms. In order to link $\check{\lambda}_n$ with $\hat{\lambda}_n$, notice that

$$\hat{\lambda}_n - \check{\lambda}_n = \frac{1}{\Delta} \int \mathbb{1}_{[-H_n, H_n]}(x) \mathcal{F}^{-1}[\mathrm{Log}(\mathcal{F} dG_n) \mathcal{F} K_{h_n}](x) dx. \tag{2.4.2}$$

The right hand side of converges to 0 as $n \to \infty$ because the integral in (2.2.12) is zero, so the two estimators are essentially the same in all situations.

So what can be said about their asymptotic properties? To answer this question for $\tilde{\lambda}_n$, its expression suggests to use the delta method. Indeed, defining $\psi(x) := -\frac{1}{\Delta} \log x$,

$$\tilde{\lambda}_n = \psi\left(\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{0\}}(Z_k) \right),$$

and, by the delta method combined with the central limit theorem, it follows that, as $n \to \infty$,

$$\sqrt{n}\left(\tilde{\lambda}_n - \psi(dG(\{0\})) \right) \to^d N\left(0, \frac{dG(\{0\})^{-1} - 1}{\Delta^2} \right), \tag{2.4.3}$$

where $dG(\{0\})$ is the mass of dG at the origin. We readily see that $\tilde{\lambda}_n$ is consistent if and only if $dG(\{0\}) = e^{-\lambda\Delta}$, which holds in the situations for which it was proposed, and, in these, the asymptotic variance is $\frac{e^{\lambda\Delta}-1}{\Delta^2}$. Proving the asymptotic properties of $\hat{\lambda}_n$ and $\check{\lambda}_n$ is much more involved and is part of the proofs in Chapter 3. The assumptions needed are the same as those mentioned when constructing our estimator for the jump distribution. Under these and choosing h_n, ε_n and H_n appropriately, the estimators satisfy that, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\lambda}_n - \lambda \right) \to^d N\left(0, \sigma_\lambda^2\right) \quad \text{and} \quad \sqrt{n}\left(\check{\lambda}_n - \lambda \right) \to^d N\left(0, \sigma_\lambda^2\right), \tag{2.4.4}$$

where

$$\sigma_\lambda^2 := \frac{1}{\Delta^2} \int \left(f^{(\lambda)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right] \right)^2 dG$$

with $f^{(\lambda)} := \mathbb{1}_{\mathbb{R}\setminus\{0\}}$ and $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ as in (2.2.17). We note in passing that this, together with (2.4.2), implies the asymptotic correlation between the two is 1. After some algebra and under the assumption that if dF has a non-zero atomic component its support is in $\varepsilon \times \mathbb{Z}\setminus\{0\}$ for some $\varepsilon > 0$,

$$\begin{aligned} \sigma_\lambda^2 &= \frac{1}{\Delta^2} \left(\sum_{j \in \mathbb{Z}} dG(\{\varepsilon j\}) \left(\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{\varepsilon j\}) \right)^2 - 1 \right) \\ &= \frac{dG(\{0\}) \left(\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{0\}) \right)^2 - 1}{\Delta^2} + \sum_{j \in \mathbb{Z}\setminus\{0\}} \frac{dG(\{\varepsilon j\}) \left(\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{\varepsilon j\}) \right)^2}{\Delta^2}. \end{aligned} \tag{2.4.5}$$

In view of the expressions for dG and $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ in (2.2.18) and (2.2.17), respectively, the second summand is zero if dF has no discrete component in $\mathbb{R}^+$ and/or $\mathbb{R}^-$. This has two consequences: in these situations, $dG(\{0\}) = e^{-\lambda\Delta}$, so $\tilde{\lambda}_n$ is consistent, and, furthermore, $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{0\}) = e^{\lambda\Delta} = dG(\{0\})^{-1}$ so the three estimators are asymptotically equivalent with correlation 1 in view of (2.4.1); however, in general they are not because the second summand is strictly positive due to the zero observations of Z arising from cancellations between jumps, as anticipated above. In Section 2.6 we argue that $\hat{\lambda}_n$ and $\check{\lambda}_n$ are efficient, and from the first equality in the last display we can draw an interesting and intuitive conclusion: in the general case, many observations may be repeated and, for an optimal estimator, the null increments are no longer special because the process may have a non-zero drift, so it needs to decompound each and all of them; the contribution of each of them to the difficulty of estimating λ is given by $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{\varepsilon i\})^2$, weighted according to the probability of observing such jump.

Nonetheless, the practical implementation of $\tilde{\lambda}_n$ is trivial whilst implementing $\hat{\lambda}_n$ and $\check{\lambda}_n$ requires approximating an integral and, more importantly, depends on the choice of h_n, ε_n and H_n. Furthermore, as just mentioned, $\tilde{\lambda}_n$ is asymptotically equivalent to the other two estimators in many realistic cases. Therefore, the naive estimator is very attractive in practice and, moreover, as mentioned in Chapter 4, it behaves better than the other two when the problem is mildly inverse. However, it has a very undesirable practical limitation that $\hat{\lambda}_n$ and $\check{\lambda}_n$ do not have: when there are no zero observations it is undefined. If $dG(\{0\}) = e^{-\lambda\Delta}$, the probability of this occurring in a sample of size n is

$$Pr(Z_1 \neq 0 \dots Z_n \neq 0) = (1 - e^{-\lambda\Delta})^n.$$

When $\lambda\Delta \approx 0$ this is approximately $(\lambda\Delta)^n \approx 0$. Nevertheless, for larger values it may not be negligible as shown in Table 2.1.

		$\lambda\Delta$					
		3	4	5	6	7	8
n	100	0.61	15.75	50.86	78.02	91.28	96.70
	250	0	0.98	18.45	53.77	79.61	91.95
	500	0	0.01	3.40	28.91	63.37	84.56
	1000	0	0	0.12	8.36	40.16	71.50

Table 2.1: Probability (in %) of not observing a zero-increment in n Δ-increments of a process with intensity λ and distribution function with no jumps potentially cancelling each other.

Lastly, let us remark that, as $\Delta \to 0$, these three estimators converge to the maximum likelihood estimator proposed in Section 1.3.1. In this regime, the three estimators above are equal up to negligible terms because the jumps are observed directly and no

cancellations can arise. Therefore, we can focus on $\tilde{\lambda}_n$. When $\Delta \to 0$, the number of zero observations increases exponentially and its expression can be approximated by

$$\frac{1}{n\Delta}\sum_{k=1}^{n} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(Z_k). \tag{2.4.6}$$

Note that $n\Delta = \lfloor T/\Delta \rfloor \Delta$ is roughly the length of the total observation interval, which is approximately equal to the sum of the underlying interarrival times, whilst the sum is counting the number of jumps. Therefore, the inverse of the last display converges to the average length of each interarrival time and the conclusion follows. Indeed, the asymptotic covariance of $\tilde{\lambda}_n$, $\frac{e^{\lambda\Delta}-1}{\Delta^2}$, is approximately λ/Δ, so if (2.4.3) is multiplied by $\sqrt{\lambda\Delta}$, which is the square root of the expected number of jumps per observation interval, we recover (1.3.1).

2.4.2 Estimating the drift

Up to now we have assumed $\gamma = 0$. This is because the works that deal strictly with the compound Poisson case mentioned so far, except for Coca [2015], assume so. In order to construct the spectral estimator of γ of Coca [2015], we note that, if $\gamma \neq 0$, formal expression (2.2.12) becomes

$$\frac{1}{\Delta}\mathcal{F}^{-1}[\mathrm{Log}(\mathcal{F}dG)]\,(dx) = \nu(dx) - \lambda\delta_0(dx) + \gamma\mathcal{F}^{-1}[i\cdot]\,(dx), \quad x \in \mathbb{R}\,. \tag{2.4.7}$$

In line with the other estimators, we consider an integral-functional of this display. In particular, we multiply both sides by x and integrate the resulting expression over an arbitrarily small interval around the origin. More precisely, for $\varsigma > 0$ small, we consider

$$\begin{aligned}\frac{1}{\Delta}\int_{\mathbb{R}} \mathbb{1}_{[-\varsigma,\varsigma]}(x)x\mathcal{F}^{-1}[\mathrm{Log}(\mathcal{F}dG)]\,(dx) = &\int_{\mathbb{R}} \mathbb{1}_{[-\varsigma,\varsigma]}(x)x\nu(dx) - \lambda\int_{\mathbb{R}} \mathbb{1}_{[-\varsigma,\varsigma]}(x)x\delta_0(dx)\\ &+ \gamma\int_{\mathbb{R}} \mathbb{1}_{[-\varsigma,\varsigma]}(x)x\mathcal{F}^{-1}[i\cdot]\,(dx).\end{aligned}$$

The first summand is arbitrarily small for two reasons: the finite measure ν has no mass at the origin by assumption; and, if it has any atoms we have assumed they are in $\varepsilon \times \mathbb{Z}\setminus\{0\}$ for some $\varepsilon > 0$, so when $\varsigma < \varepsilon$ the integral is over a density and it is restricted to an arbitrarily small interval. The second summand is exactly zero by the defining property of δ_0 and, thus, we are left with analysing the third term: since, formally, $\left(\mathcal{F}^{-1}\mu\right)' = -\mathcal{F}^{-1}[i\cdot\mu]$ and $\mathcal{F}^{-1}1 = \delta_0$, $\mathcal{F}^{-1}[i\cdot]$ can be interpreted as the distributional derivative of $-\delta_0$. Then, using integration by parts, the last integral is

$$-\int_{\mathbb{R}} \mathbb{1}_{[-\varsigma,\varsigma]}(x)x(\delta_0)'(dx) = \int_{\mathbb{R}} \left((\delta_{-\varsigma}(x) - \delta_{\varsigma}(x))\,x + \mathbb{1}_{[-\varsigma,\varsigma]}(x)\right)\delta_0(dx) = 1.$$

In practice, we use a kernel estimator and therefore the value of this integral is slightly different: substituting δ_0 by K_{h_n}, the second integral becomes

$$\int_{\mathbb{R}} \left((\delta_{-\varsigma}(x) - \delta_{\varsigma}(x))\, x + \mathbb{1}_{[-\varsigma,\varsigma]}(x) \right) K_{h_n}(x)dx = \int_{-\varsigma/h_n}^{\varsigma/h_n} K(x)dx \; - \; \frac{\varsigma}{h_n} \left(K\left(\frac{\varsigma}{h_n}\right) + K\left(-\frac{\varsigma}{h_n}\right) \right).$$

Consequently, the estimator we propose is

$$\hat{\gamma}_n := \frac{1}{c\Delta} \int_{\mathbb{R}} \mathbb{1}_{[-h_n,h_n]}(x) x \mathcal{F}^{-1}[\mathrm{Log}(\mathcal{F}dG_n)\mathcal{F}K_{h_n}](x)dx, \tag{2.4.8}$$

where $c := \int_{-1}^{1} K - (K(1) + K(-1))$. We emphasise that, in view of the second to last expression, $\varsigma = \varsigma_n$ cannot be chosen to converge to zero at a faster rate than h_n for this estimator to work. Furthermore, the estimators of F and λ we constructed in previous sections can still be used when a non-zero drift is present in the data: this is due to identity (2.4.7) and because, since

$$\int \mathbb{1}_{\mathbb{R}\setminus[-\varsigma,\varsigma]}(x) x \mathcal{F}^{-1}[i \cdot K_{h_n}](x)dx = 1 - \int_{\mathbb{R}} \left((\delta_{-\varsigma}(x) - \delta_{\varsigma}(x))\, x + \mathbb{1}_{[-\varsigma,\varsigma]}(x) \right) K_{h_n}(x)dx \tag{2.4.9}$$

if $\lim_{x\to\pm\infty} xK(x) = 0$ and recalling that $\int_{\mathbb{R}} K = 1$, the left hand side of this display vanishes for $\varsigma_n = \varepsilon_n$ when $\varepsilon_n/h_n \to \infty$.

In Chapter 3 we show that, under the same assumptions as above and as $n \to \infty$,

$$h_n^{-1}\sqrt{n}\,(\hat{\gamma}_n - \gamma) \to^{Pr} 0.$$

This, together with the comment right after (2.4.8), shows that the rate of convergence of $\hat{\gamma}_n$ is exactly $h_n/\sqrt{n}$. In the next chapter we choose h_n to vanish at an exponential rate. Therefore, as already mentioned in the previous Chapter, $\hat{\gamma}_n$ converges to γ exponentially fast. This is an unusual property, since the rate of convergence of an estimator is generally upper bounded by the parametric $1/\sqrt{n}$-rate. It can be intuitively explained as follows: unlike in typical statistical problems, here, the quantity of interest, γ, is directly observed with non-zero probability. This corresponds to those observations with no jumps, which were considered before when constructing $\tilde{\lambda}_n$. Now, however, we are not interested in the parameter of the distribution generating them but simply in their value $\gamma\Delta$, where we recall that Δ is known. In fact, in the abovementioned cases when $\tilde{\lambda}_n$ is a consistent estimator for λ, γ can be learnt without error for (large enough) finite samples. This is because if dF has no discrete component in $\mathbb{R}^-$ (and/or $\mathbb{R}^+$), when n is large enough there will always be at least two observations equal to $\gamma\Delta$ and these will be the smallest (only/largest) ones among all of those that are equal. If there is a priori knowledge that dF enjoys such property, for instance if it has a density or if it is supported only in $\mathbb{R}^+$

or $\mathbb{R}^-$, we propose to use the naive estimator

$$\tilde{\gamma}_n := \frac{1}{\Delta} Z_{\tilde{k}(n)}, \tag{2.4.10}$$

where

$$\tilde{k}(n) := \operatorname{argmin}_{k \in \tilde{K}(n)} |Z_k| \quad \text{and} \quad \tilde{K}(n) := \left\{ k = 1, \dots, n : \sum_{l=1}^{n} \mathbb{1}_{Z_k = Z_l} > 1 \right\}.$$

When the set $\tilde{K}(n)$ is empty we take $\tilde{\gamma}_n := 0$ but, again, the probability of this happening tends to zero as $n \to \infty$. In fact, in general, the minimum in the definition of $\tilde{k}(n)$ is attained for more than one k; because the set of these is finite, we can take the minimum of them without loss of generality. Trivially, and naturally paralleling the behaviour of $\tilde{\lambda}_n$ above, $\tilde{\gamma}_n$ is a consistent estimator (almost surely) in the cases mentioned above when $\tilde{\lambda}_n$ is too, but not in general. In these special cases, all the existing estimators of F and λ do not require any modification as we can simply work with $Z_k - \tilde{\gamma}_n \Delta$, $k = 1, \dots, n$, instead of with the original observations. In the rest of the cases, the only safe strategy is to use $\hat{\gamma}_n$, constructed through the spectral approach. In practice it is very well behaved when the problem is mildly inverse, honouring its exponential rate of convergence. However, as soon as the proportion of observations without jumps is small (whose probability is reflected in Table 2.1 above), it behaves erratically. This is consistent with the intuition given before, since these are the observations that truly inform us about γ.

To our knowledge, these are the first estimators of the drift in the strict setting of estimating discretely observed compound Poisson processes. In other settings such as that of estimating more general discretely observed Lévy processes, estimators have been developed by authors such as Gugushvili [2009], Trabs [2014] and Belomestny et al. [2015]. Ours is closer to the estimator of the first, although they still exploit different properties of the spectral approach. Moreover, they require stronger assumptions and the rate of convergence of their estimators cannot be exponential but polynomial or logarithmic. Therefore we feel the estimators for γ proposed above and the discussions arising from them add value and further insight into the problem.

Lastly, let us compute the limit of these estimators as $\Delta \to 0$. That of the naive estimator is trivially equivalent to measuring the slope in the continuous observation regime of Section 1.3.1. To analyse that of $\hat{\gamma}_n$, we work with the quantity inside the logarithm, which is 1 plus $\mathcal{F}dG_n - 1 = (\mathcal{F}dG_n - \mathcal{F}dG) + (\mathcal{F}dG - 1)$. In the compact support of $\mathcal{F}K_{h_n}$, the first term is uniformly negligible by the control of the empirical characteristic function developed by Kappus and Reiß [2010]. So is the second because $e^{-2\lambda\Delta} \leq |\mathcal{F}dG| \leq 1$ by the Lévy–Khintchine representation of $\mathcal{F}dG$ in (2.2.8). Therefore,

the expression for $\hat{\gamma}_n$ is, up to negligible terms, equal to

$$\frac{1}{c\Delta}\int_{\mathbb{R}} \mathbb{1}_{[-h_n,h_n]}(x)x\mathcal{F}^{-1}[(\mathcal{F}dG_n - 1)\,\mathcal{F}K_{h_n}](x)dx = \frac{1}{c\Delta}\int_{\mathbb{R}} \mathbb{1}_{[-h_n,h_n]}(x)xdG_n * K_{h_n}(x)dx - h_n\frac{1}{c\Delta}\int_{\mathbb{R}} \mathbb{1}_{[-1,1]}(x)xK(x)dx.$$

The second summand is exactly zero if K is symmetric, as we will assume, and exponentially negligible otherwise because we will be able to make h_n vanish exponentially fast. And the first term is the kernel-regularised version of

$$\frac{1}{\Delta n}\sum_{k=1}^{n} \mathbb{1}_{[-h_n,h_n]}(Z_k)Z_k. \tag{2.4.11}$$

Since ν has no atom at the origin by assumption and the number of observations with jumps is asymptotically negligible, this is estimating γ in a very intuitive way when taking $\Delta_n = o(h_n)$.

2.5 Estimators of the process

Most of the discussions so far have concerned estimating the parameters separately. To make inference on the whole underlying compound Poisson process, one typically shows joint convergence of the estimators of all the parameters. In the case when dF has an absolutely continuous component and with the exception of the work of Coca [2015], this has not been addressed in the literature discussed above. There are two different reasons: on the one hand, Buchmann and Grübel [2003] and van Es et al. [2007] assumed $\gamma = 0$ and that the intensity is known; and, on the other, Duval [2013a] and Comte et al. [2014] made estimation in the high frequency regime, so as $\Delta \to 0$ the setting tends to the continuous observation regime of Section 1.3 and estimation of the parameters is asymptotically independent. Notice that, due to the asymptotic properties of our estimator for γ included in the previous section, making inference on the process reduces to making joint inference on F and λ. This goes in line with the remark in Trabs [2015a] where it was pointed out that the addition of a drift to the model does not affect the information lower bounds. The exact asymptotic dependence of the estimators of F and λ, and of the estimators for the mass of the atoms, was given by Coca [2015] and is part of the main result of Chapter 3. Therefore, joint estimation is possible using these results and in Section 4.4 we show their joint behaviour in practice. We remark that, in the case when only a discrete component is present, joint estimation in the discrete case was addressed by Buchmann and Grübel [2003] and Buchmann and Grübel [2004], and is asymptotically equivalent to that following from our results.

An alternative to jointly estimating the parameters is to merge them into a single

parameter. This is precisely the role of the Lévy distribution N introduced in Section 2.2, which carries all the information of the process. There, we constructed an estimator of it, namely $\widehat{N}_n$. In Coca [2015] and in Chapter 3, we show that, under the same assumptions considered in Section 2.2 and as $n \to \infty$, the estimator is well-defined in sets of probability approaching 1 and

$$\sqrt{n}\left(\widehat{N}_n - N\right) \to^{\mathcal{D}} \mathbb{B}^N \quad \text{ in } \ell^\infty(\mathbb{R}), \tag{2.5.1}$$

where $\mathbb{B}^N$ is a centred Gaussian process with covariance structure

$$\Sigma^{\mathbb{B}^N}_{x,y} := \frac{1}{\Delta^2}\int_{\mathbb{R}} \left(f_x^{(N)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](z)\right)\left(f_y^{(N)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](z)\right) dG(z), \quad x, y \in \mathbb{R},$$

with $f_x^{(N)} := \mathbb{1}_{(-\infty,x]}\,\mathbb{1}_{\mathbb{R}\setminus\{0\}}$, $x \in \mathbb{R}$. As we see in the next section, the covariance structure of this process coincides with the information lower bound developed by Trabs [2015a], so $\widehat{N}_n$ is asymptotically efficient and optimal estimation of the process can also be achieved this way. Furthermore, and as pointed out in Coca [2015], it turns out to give important insights that allow us to justify why our estimators are efficient and others are not.

Moreover, note the similarity between $\Sigma^{\mathbb{B}^F}_{x,y}$ and the covariance $\Sigma^{\mathbb{G}^G}_{x,y}$ from Nickl and Reiß [2012] given in Section 2.2.2. Indeed, N^G is the generalised version of N and next we see that, under the assumptions in Nickl and Reiß [2012], the two estimators are, up to exponentially negligible terms, the same. The strategy is to convert $\widehat{N}_n(x)$ into $\widehat{N}^G_n(x)$, and, for simplicity, we assume $\gamma = 0$ and $x \le -\zeta < 0$. Then, we can introduce any of the estimators of the intensity, say $\hat{\lambda}_n$, in the Fourier inverse in (2.2.14) by simply adding and subtracting a (uniformly in x) negligible term, and

$$\begin{aligned}
\widehat{N}_n(x) &\approx \frac{1}{\Delta}\int_{-\infty}^{x} \mathcal{F}^{-1}\left[\left(\mathrm{Log}(\mathcal{F}dG_n) + \hat{\lambda}_n\right)\mathcal{F}K_{h_n}\right](y)dy \\
&= \frac{1}{\Delta}\int_{-\infty}^{x} \frac{1}{iy}\mathcal{F}^{-1}\left[iy\left(\mathrm{Log}(\mathcal{F}dG_n) + \hat{\lambda}_n\right)\mathcal{F}K_{h_n}\right](y)dy \\
&= \frac{1}{\Delta}\int_{-\infty}^{x} \frac{1}{iy}\mathcal{F}^{-1}\left[\left(\mathrm{Log}(\mathcal{F}dG_n) + \hat{\lambda}_n\right)'\mathcal{F}K_{h_n} + \left(\mathrm{Log}(\mathcal{F}dG_n) + \hat{\lambda}_n\right)(\mathcal{F}K_{h_n})'\right](y)dy,
\end{aligned}$$

where the equalities hold in sets of probability approaching 1 as $n \to \infty$. The first summand in the Fourier inverse gives rise to exactly $\widehat{N}^G_n(x)$ because of the finite moment assumption, so we need to show that the second is uniformly negligible. In the next chapter, and in Nickl and Reiß [2012], we assume the tails of K decay faster than quadratically. Therefore, $(\mathcal{F}K_{h_n})' = i\mathcal{F}[\cdot K_{h_n}] = ih_n\mathcal{F}[(\cdot K)_{h_n}]$ and this is supported in $[-h_n^{-1}, h_n^{-1}]$, so $(\mathcal{F}K_{h_n})' \in L^r(\mathbb{R})$ for all $r \ge 1$. In this compact interval, $\left(\mathrm{Log}(\mathcal{F}dG_n) + \hat{\lambda}_n\right)$ is bounded above and below in sets of probability approaching 1 and its product with $(\mathcal{F}K_{h_n})'$ has $L^2(\mathbb{R})$ norm of order $h_n^{1/2}$ if $\cdot K \in L^2(\mathbb{R})$, which we will assume. Furthermore, the supremum in $x \le -\zeta < 0$ of the $L^2(\mathbb{R})$ norm of $1/y\,\mathbb{1}_{(-\infty,x]}(y)$ is bounded above and therefore the second summand in the last display is of order h_n uniformly in x, which we will take

to decay to zero exponentially fast. We note that, if $\widehat{N}_n^G(x)$, $x \leq -\zeta < 0$, is rescaled by $\hat{\lambda}_n$ then, by the same arguments, it is equal to $\widehat{F}_n(x)$ up to uniformly and exponentially negligible additive terms.

Another interesting question is whether, as $\Delta \to 0$ and under the assumptions of Nickl et al. [2016], the appropriate modification of $\widehat{N}_n$ is, up to uniformly negligible terms, equal to the estimators $\widetilde{\mathcal{N}}_n$ and $\widehat{\mathcal{N}}_n$. This is indeed the case, and for the latter can be argued as in the previous paragraph assuming K decays fast enough. For the former, we argue as follows: the appropriate modification of $\widehat{N}_n$ is

$$\frac{1}{\Delta}\int_{-\infty}^{x} \rho(y) y^2 \mathcal{F}^{-1}[\mathrm{Log}(1 + \mathcal{F}dG_n - 1)\mathcal{F}K_{h_n}](y)dy.$$

Arguing as at the end of Section 2.4.2, this can be shown to be, up to uniformly in x negligible terms, equal to

$$\begin{aligned}\frac{1}{\Delta}\int_{-\infty}^{x} \rho(y) y^2 \mathcal{F}^{-1}[(\mathcal{F}dG_n - 1)\,\mathcal{F}K_{h_n}](y)dy =& \frac{1}{\Delta}\int_{-\infty}^{x} \rho(y) y^2 dG_n * K_{h_n}(y)dy \\ &- \frac{1}{\Delta}\int_{-\infty}^{x} \rho(y) y^2 K_{h_n}(y)dy.\end{aligned}$$

As expected, the first summand is the kernel-regularised version of $\widetilde{\mathcal{N}}_n$, and the second is uniformly negligible because $\rho(y) \leq C\,(1 \wedge y^{-2})$ for some $C > 0$ by assumption on ρ. Hence, we also conclude that the existing estimators developed from the spectral approach are equal up to uniformly negligible terms. We remark that, in the compound Poisson case, the Lévy measure has no singularity at the origin and the same calculation shows that, as $\Delta \to 0$, our estimator $\widehat{N}_n$ is asymptotically equivalent to the cumulative function of the linear Lévy density estimator

$$\frac{1}{\Delta n}\sum_{k=1}^{n} K_{h_n}(\cdot - Z_k)\,\mathbb{1}_{\{0\}^c}(Z_k)$$

(note the similarity of this estimator with the intensity and drift estimators in (2.4.6) and (2.4.11)). This estimator and its wavelet versions resemble those of Comte et al. [2014] and Duval [2013a] with $I = 1$ (cf. (2.1.14) and the discussions after it), with the distinction that the use the non-zero observations conditioning on the total number of them while the last display takes all and automatically discards the null ones. We note in passing that in this form they have not been studied in the literature and preliminary calculations indicate they enjoy the same features (minimax optimality, concentration of measure, etc.) as the respective estimators in the standard i.i.d. case. Furthermore, if we rescale our estimator $\widehat{N}_n$ by $\hat{\lambda}_n$, the resulting estimator is asymptotically equivalent to the empirical distribution function as $\Delta \to 0$.

2.6 Asymptotic efficiency of the estimators

In the preceding sections we have focused on analysing the existing estimators of the defining quantities of a discretely observed compound Poisson process. Moreover, we have shown that they are all very similar estimators and, consequently, one would expect the limiting quantities in their central limit theorems to be very similar too. Here we show that this is indeed the case and that the differences arise in the attainment of efficiency. We also show how to transform estimators so that the limiting processes associated to them have smaller covariances with the objective of making them efficient. We mainly focus on the estimators of Buchmann and Grübel [2003] and Coca [2015], and then extend the conclusions to the rest of estimators by adapting the arguments. We also give some insights into the topic that are not present in existing literature.

First we make some remarks that apply to the whole section. In (2.2.17) and (2.2.18) we gave explicit expressions for $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ and dG in terms of F, λ and Δ assuming that $\gamma = 0$. From now on in this thesis, let us write P for dG. Then, keeping the rest of the notation fixed, if $\gamma \neq 0$ the expressions change to

$$\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big] = e^{\lambda\Delta}\,\delta_{\Delta\gamma} * \sum_{k=0}^{\infty} d\bar{F}^{*k}\frac{(-\lambda\Delta)^k}{k!} \tag{2.6.1}$$

and

$$P = e^{-\lambda\Delta}\,\delta_{\Delta\gamma} * \sum_{k=0}^{\infty} dF^{*k}\frac{(\lambda\Delta)^k}{k!}. \tag{2.6.2}$$

It still holds that, for any bounded functions f, g on $\mathbb{R}$,

$$\int_{\mathbb{R}} \Big(f * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)\Big(g * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)\,P(dx) < \infty. \tag{2.6.3}$$

Furthermore, if $\mathcal{F}^{-1}\left[\varphi_0^{-1}(-\cdot)\right]$ and P_0 represent the finite measures above when $\gamma = 0$, the integral in the last display can be written as

$$\begin{aligned}\int_{\mathbb{R}} &\Big(f * \mathcal{F}^{-1}\left[\varphi_0^{-1}(-\cdot)\right] * \delta_{\Delta\gamma}(x)\Big)\Big(g * \mathcal{F}^{-1}\left[\varphi_0^{-1}(-\cdot)\right] * \delta_{\Delta\gamma}(x)\Big)P_0 * \delta_{\Delta\gamma}(dx)\\ &= \int_{\mathbb{R}} \Big(f * \mathcal{F}^{-1}\left[\varphi_0^{-1}(-\cdot)\right](x)\Big)\Big(g * \mathcal{F}^{-1}\left[\varphi_0^{-1}(-\cdot)\right](x)\Big)P_0(dx).\end{aligned}$$

Quantities like (2.6.3) are central in the discussions that follow and we therefore lose no generality by assuming $\gamma = 0$ when making conclusions that only depend on them. With some abuse of notation, we drop the subscript 0 from the finite measures in them.

Let us recall the results of Trabs [2015a] that concern Δ-discretely observed compound Poisson processes. Assuming dF is absolutely continuous with respect to Lebesgue's measure, he studied semiparametric efficiency for estimation of regular enough functionals of

the Lévy density $\nu := \lambda dF$. In particular, he showed that if f is regular enough and a central limit theorem holds for a consistent estimator of $\int_{\mathbb{R}} f d\nu$, the variance of the (centred) limiting normal distribution is bounded below by[2]

$$\frac{1}{\Delta^2} \int_{\mathbb{R}} \left(\left(f \mathbb{1}_{\mathbb{R}\setminus\{0\}} \right) * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot) \right](x) \right)^2 P(dx). \tag{2.6.4}$$

Recalling that $N(x) := \int_{\mathbb{R}} \mathbb{1}_{(-\infty,x]}\, d\nu$ and $\lambda = \int_{\mathbb{R}} d\nu$, and in view of (2.5.1) and (2.4.4), we immediately see that their limiting variances coincide with this lower bound. We are considering a larger model than Trabs [2015a] because we allow dF to have an atomic component and hence its lower bound must be bounded below by the last display. Thus, our results imply that its expression does not change and that our estimators for N and λ are efficient. Furthermore, and in view of the covariance of the limiting process in (2.2.6), so is the estimator of Nickl and Reiß [2012] as an estimator of N^G in the compound Poisson case. This agrees with the asymptotic equivalence between the two estimators (in the negative real line) pointed out in Section 2.5. As we show in Chapter 3, from the expressions for $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ and P in (2.6.1) and (2.6.2), we can also see that the last display can be written as

$$\frac{1}{\Delta} \int_{\mathbb{R}} f(x)^2 \nu(dx) + O(1). \tag{2.6.5}$$

Let us multiply this quantity by Δ and take $f(x) = \mathbb{1}_{(-\infty,\cdot]}(x)\rho(x)x^2$. Then, in view of (2.2.7), we again observe the asymptotic equivalence between our estimator and those in Nickl et al. [2016] in the high-frequency regime that we already mentioned at the end of Section 2.5. We note in passing that, if the last display is multiplied by $\lambda\Delta$, i.e. by the effective number of observations per observation interval, and take $f(x) = 1$, we see that the limiting quantity for our estimators of the intensity coincides with that of the maximum likelihood estimator in Section 1.3.2, as anticipated at the end of Section 2.4.1. Indeed, Duval and Hoffmann [2011] studied the optimal limiting variance when estimating the intensity λ in a simple parametric example and under the high-frequency regime, and obtained this same limiting quantity.

At first view, the truncation $\mathbb{1}_{\mathbb{R}\setminus\{0\}}$ appearing in (2.6.4) may seem superfluous, especially if dF has a density. However, the value of that expression does change depending on whether $\mathbb{1}_{\mathbb{R}\setminus\{0\}}$ is included in it. This is due to the atomic measures at the origin in the first summand of the expressions for $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ and P in (2.2.17) and (2.2.18), respectively, which, crucially, carry no information about dF. If these infinite series representations are plugged into (2.6.4), the term arising from the two Dirac deltas is killed by $\mathbb{1}_{\mathbb{R}\setminus\{0\}}$. In turn, it also is the reason why 2.6.5 holds and has no term of order Δ^{-2}, which is key for optimality in the high-frequency regime. Note that, in the larger model

[2]In personal communication with M. Trabs it was confirmed the indicator $\mathbb{1}_{\mathbb{R}\setminus\{0\}}$ should feature in his results in Section 4.2 for them to be correct.

we consider, $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ and P may have atoms at zero coming from other summands. These contain information about F and the terms arising from them in (2.6.4) are not annihilated by the truncation. This insight also extends to the case when $\gamma \neq 0$.

Let us now justify the efficiency of our estimator of the jump distribution F. When there is knowledge of λ, Trabs [2015a] derived the lower bound for the problem of estimating N. This can be found right after Corollary 4.5 and is given by λ^2 times[3]

$$\frac{1}{\Delta^2}\int_{\mathbb{R}}\left(f_x^{(F)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](y)\right)^2 P(dy), \tag{2.6.6}$$

where we recall that $f_x^{(F)} := \lambda^{-1}\left(\mathbb{1}_{(-\infty,x]} - F(x)\right)\mathbb{1}_{\mathbb{R}\setminus\{0\}}$. If the intensity is known, estimating N and F is the same problem and we note that the last display coincides with the covariance of $\mathbb{G}^F$ in (2.2.16). However, we attained it in the harder setting of not knowing λ, and therefore it must also be the lower bound for estimating F in this more general setting. This adds new insight into the problem and allows us to conclude that our estimator of F is efficient.

To show the lack of efficiency of the estimator of F in Buchmann and Grübel [2003], we first give intuition of why our estimator does have this property. As mentioned at the end of Section 2.5, one can show that our estimator converges to the empirical distribution function as $\Delta \to 0$. A first indication of this is that, as shown in Chapter 3, the last display can be written as

$$\frac{1}{\lambda\Delta}F(x)\left(1-F(x)\right) + O(1). \tag{2.6.7}$$

Therefore, at least for small Δ, the limit process in (2.2.16) is approximately tied up to 0 in the limit as $x \to \infty$. Indeed, separating the atomic and absolutely continuous measures in $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ and P, and by the dominated convergence theorem, the limit of (2.6.6) as $x \to \pm\infty$ is 0 for any $\Delta > 0$. We interpret this as meaning that the limiting process is of 'Brownian bridge type'. This was achieved by the division of $\widehat{N}_n$ by $\hat{\lambda}_n$ in the definition of $\widehat{F}_n$ that forced the end point to be exactly 1 for any n. In other words,

$$\lim_{n\to\infty}\lim_{x\to\pm\infty}\sqrt{n}\left(\widehat{F}_n(x) - F(x)\right) \equiv 0.$$

We remark that this behaviour is also guaranteed if we divide by $\check{\lambda}_n$ instead: while it is clear that $\lim_{x\to-\infty}\widehat{N}_n(x)/\check{\lambda}_n = 0$ for any n,

$$\begin{aligned}\lim_{n\to\infty}\lim_{x\to\infty}\sqrt{n}\left(\frac{\widehat{N}_n(x)}{\check{\lambda}_n} - F(x)\right) &= \lim_{n\to\infty}\lim_{x\to\infty}\sqrt{n}\left(\frac{\hat{\lambda}_n}{\check{\lambda}_n}\widehat{F}_n(x) - F(x)\right)\\ &= \lim_{n\to\infty}\sqrt{n}\left(\frac{\hat{\lambda}_n - \check{\lambda}_n}{\check{\lambda}_n}\right).\end{aligned}$$

[3] In personal communication with the author it was confirmed that $\tilde{\chi}_\nu$ defined after the corollary should equal $\lambda f_x^{(F)}$ for the results to be correct.

In general, this may not converge to a null distribution but, right after (2.4.4), we concluded the correlation between $\hat{\lambda}_n$ and $\check{\lambda}_n$ is asymptotically 1 and the conclusion then follows. In fact, there we also argued that in the cases when $\tilde{\lambda}_n$ is consistent, the three have asymptotic correlation 1 and, by the same argument, the naive one can also be used for the rescaling of $\widehat{N}_n$, but only in these cases. On the other hand, and in view of (2.6.5) with $f = \mathbb{1}_{(-\infty,x]}$, a different behaviour for the limiting process in (2.5.1) is expected. Indeed, with this choice of f and by the same arguments as above, the limit of (2.6.4) as $x \to -\infty$ and $x \to \infty$ is 0 and $\sigma_\lambda^2 > 0$, respectively, for any $\Delta > 0$. We refer to a limiting process with this behaviour as of 'Brownian motion type'. It agrees with the above because, while $\lim_{x\to-\infty} \widehat{N}_n(x) = 0$ for all n,

$$\lim_{n\to\infty} \lim_{x\to\infty} \sqrt{n}\left(\widehat{N}_n(x) - N(x)\right) = \lim_{n\to\infty} \sqrt{n}\left(\hat{\lambda}_n - \lambda\right) =^d N\left(0, \sigma_\lambda^2\right).$$

Recall that N is an estimator of the whole process. Thus, intuitively, the division of $\widehat{N}_n$ by $\hat{\lambda}_n$ in the definition of $\widehat{F}_n$ is removing the uncertainty of not knowing λ, and this division is precisely what guarantees that $\widehat{F}_n$ is efficient. This reinforces the remark of the previous paragraph since it means that, if we know λ, we should still divide the estimator of N by $\hat{\lambda}_n$ when estimating F and hence the asymptotic properties do not change. Furthermore, to estimate N we should rescale the resulting $\widehat{F}_n$ by λ and not use $\widehat{N}_n$ directly.

Let us now analyse the behaviour of the limiting process for the estimator of F in Buchmann and Grübel [2003]. Recall that in the definition of $\widetilde{F}_n$ in (2.1.10), Buchmann and Grübel [2003] do not input an estimator of λ. This means $\widetilde{F}_n$ is not forced to end exactly at one for finite samples but at the value in (2.1.11). Hence, and assuming $\lambda\Delta < \log 2$ for simplicity, for n large enough there is a set of probability approaching 1 as $n \to \infty$ in which

$$\begin{aligned}
\lim_{n\to\infty} \lim_{x\to\infty} \sqrt{n}\left(\widetilde{F}_n(x) - F(x)\right) &= \lim_{n\to\infty} \sqrt{n}\left(\frac{1}{\lambda\Delta}\log\left(1 + e^{\lambda\Delta}(1 - e^{-\tilde{\lambda}_n\Delta})\right) - 1\right) \\
&\approx \frac{1}{\lambda\Delta} \lim_{n\to\infty} \sqrt{n}\left(e^{-\lambda\Delta} - e^{-\tilde{\lambda}_n\Delta}\right) \\
&=^d N\left(0, \frac{e^{-\lambda\Delta}(1 - e^{-\lambda\Delta})}{(\lambda\Delta)^2}\right). \qquad (2.6.8)
\end{aligned}$$

We conclude that the limiting process in the central limit theorem in Buchmann and Grübel [2003] is of Brownian motion type and therefore their estimator cannot be efficient[4]. The conclusion extends to $\lambda\Delta \geq \log 2$ by working directly with the covariance of $\mathbb{B}^F$ in (2.1.12): note that, it can be rewritten as

$$\Sigma_{x,x}^{\mathbb{B}^F} = \int_{[0,x]} \int_{[0,x]} \left(G_0\left((x-y) \wedge (x-z)\right) - G_0(x-y)G_0(x-z)\right) H(dy)H(dz),$$

[4]This was also confirmed through personal communication with M. Trabs.

where we recall that $G_0(x) := G(x) - G(0) = \int_{(0,x]} P$ and

$$H(x) := \frac{1}{\lambda\Delta} \sum_{i=1}^{\infty} (-1)^{i+1} e^{i\lambda\Delta} G_0^{*(i-1)}(x), \quad x \geq 0.$$

From this expression it is clear that $\Sigma_{0,0}^{\mathbb{B}^F} = 0$ and, to compute the limit as $x \to \infty$, we rewrite H in terms of the finite measure $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$. To derive expression (2.2.17) for the latter, one uses the series representation of the exponential expression of φ^{-1} arising from the Lévy–Khintchine formula for φ. An alternative expression is obtained by computing the Taylor series of $1/\cdot$ around $P(\{0\})$, which gives

$$\mathcal{F}^{-1}\left[\varphi^{-1}\right] = \mathcal{F}^{-1}\left[\sum_{i=1}^{\infty} (-1)^{i+1} \frac{(\varphi - P(\{0\}))^{i-1}}{P(\{0\})^i}\right].$$

Note that $\mathcal{F}dG_0 = \varphi - P(\{0\})$ and, in the setting of Buchmann and Grübel [2003] of F supported in $\mathbb{R}^+$, $P(\{0\}) = e^{-\lambda\Delta}$ in view of (2.2.18). Then, $H(x) = \frac{1}{\lambda\Delta}\mathcal{F}^{-1}\left[\varphi^{-1}\right]([0,x])$, at least formally, and the variance above can be written in the spectral approach form

$$\Sigma_{x,x}^{\mathbb{B}^F} = \frac{1}{(\lambda\Delta)^2} \int_{[-x,0]} \int_{[-x,0]} \big(G_0\left((x+y) \wedge (x+z)\right) - G_0(x+y)G_0(x+z)\big) \times \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](dy)\, \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](dz).$$

Then, noting that $\lim_{x\to\infty} G_0(x) = 1 - P(\{0\}) = 1 - e^{-\lambda\Delta}$ and recalling that $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ has mass 1,

$$\lim_{x\to\infty} \Sigma_{x,x}^{\mathbb{B}^F} = \frac{e^{-\lambda\Delta}\left(1 - e^{-\lambda\Delta}\right)}{(\lambda\Delta)^2},$$

which agrees with conclusion (2.6.8) for $\lambda\Delta < \log 2$.

Notice that the right hand side of the first line in the calculations leading to (2.6.8) is exactly zero if λ is substituted by $\tilde{\lambda}_n$. In other words, if in the definition of $\widetilde{F}_n$ we input $\tilde{\lambda}_n$ instead of λ, the limiting process is of Brownian bridge type. While the infinite series in the resulting expression also diverges in general and the exponentially downweighted norms are still needed, it corresponds to an estimator of F whose associated limiting process has covariance equal to the bound developed by Trabs [2015a]: from the first three expressions in Section 2.3.2 it follows that this substitution is the same as dividing $\widehat{N}_n$ by $\tilde{\lambda}_n$ instead of $\hat{\lambda}_n$ in the definition of $\widehat{F}_n$; the conclusion follows because in the previous paragraph we argued that, in the setting of Buchmann and Grübel [2003], $\tilde{\lambda}_n$ is an efficient estimator of F. In fact, from (2.4.1) and (2.4.2) we concluded that in this setting $\tilde{\lambda}_n$ is equal to $\hat{\lambda}_n$ up to negligible terms. Therefore, our estimator of F is essentially the kernel-regularised version of the transformed estimator of Buchmann and Grübel [2003] before insisting on expanding the logarithm. For $\lambda\Delta < \log 2$ they are pretty much equal in their setting. We emphasise that, because Buchmann and Grübel [2003] assumed the support of F is in $\mathbb{R}^+$

and Trabs [2015a] only developed bounds for the unrestricted model (he only considered perturbations of the Lévy measure from an exponential family and hence with full support in $\mathbb{R}$), efficiency of the transformation of $\widetilde{F}_n$ cannot be concluded. We ignore whether the bounds change in the setting of Buchmann and Grübel [2003] but if they do not then the remarks herein imply that the proposed transformation provides an efficient estimator of F. Similar arguments justify that any of the spectral estimators of Section 2.4.1 can also be used. For any of these three choices, the covariance of the limiting process is

$$\int_{[0,x]}\int_{[0,y]}\big(G\left((x-z_1)\wedge(y-z_2)\right)-G(x-z_1)G(y-z_2)\big)H(dz_1)H(dz_2)=\Sigma^{\mathbb{G}^F}_{x,y},\quad x,y\in\mathbb{R}^+,$$

where the equality follows because $G(x)=\int \mathbb{1}_{[0,x)}\,P$, $H(x)=\frac{1}{\lambda\Delta}\mathcal{F}^{-1}\big[\varphi^{-1}\big]\left([0,x]\right)$, and by Fubini's theorem.

In Section 2.3.2 we argued that the estimators of the jump mass function of Buchmann and Grübel [2003] and Coca [2015] are equal up to convolution with a kernel and before expanding the logarithm. Furthermore, as remarked at the end of Section 2.2.2, when dF only has an atomic component the estimator of F resulting from adding the estimates of the atoms of the latter is equal to $\widehat{F}_n$ up to negligible terms. Consequently, the covariances of the limiting processes associated to both must attain the bounds developed by Trabs [2015a]. Again, efficiency cannot be concluded since Trabs [2015a] only considered Lévy measures with no support restriction and which are absolutely continuous with respect to Lebesgue's measure. However, if the bounds do not change in the fully discrete case and when the support of F is restricted to $\mathbb{R}^+$, efficiency will follow from our remarks here. The attainment of the bound agrees with the arguments we just gave: in Section 2.1.2 we mentioned that the estimator of F resulting from that of the jump mass function of Buchmann and Grübel [2003] is $\widetilde{F}_n$, but with $\tilde{\lambda}_n$ in place of λ. In fact, this can also be checked directly: in Chapter 3 we show that the asymptotic variance in the central limit theorem for the estimator of the mass of the atom at $x\in\varepsilon\times\mathbb{Z}\setminus\{0\}$, denote it by p_x, is given by (2.6.6), where $f_x^{(F)}$ is replaced by $\lambda^{-1}\left(\mathbb{1}_{\{x\}}-p_x\right)\mathbb{1}_{\mathbb{R}\setminus\{0\}}$ as expected. Recall that in the setting of Buchmann and Grübel [2003], $\varepsilon=1$ and let $i,j\in\mathbb{N}\setminus\{0\}$. Then, the asymptotic covariance of Coca [2015] is $(\lambda\Delta)^{-2}$ times

$$\begin{aligned}
&\int\big(\left(\mathbb{1}_{\{i\}}-p_i\right)\mathbb{1}_{\mathbb{R}\setminus\{0\}}\big)*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)\big(\left(\mathbb{1}_{\{j\}}-p_j\right)\mathbb{1}_{\mathbb{R}\setminus\{0\}}\big)*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)P(dy)\\
&=\int\big(\mathbb{1}_{\{i\}}+p_i(\mathbb{1}_{\{0\}}-1)\big)*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)\left(\mathbb{1}_{\{j\}}+p_j(\mathbb{1}_{\{0\}}-1)\right)*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)P(dy)\\
&=\int\Big(\mathbb{1}_{\{i\}}*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)\,\mathbb{1}_{\{j\}}*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)+p_ip_j\left(\mathbb{1}_{\{0\}}*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)-1\right)^2\\
&\qquad+\big(p_i\mathbb{1}_{\{j\}}*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)+p_j\mathbb{1}_{\{i\}}*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)\big)\,\mathbb{1}_{\{0\}}*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](y)\Big)P(dy),
\end{aligned}$$

where the last equality follows because $\int f\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,P=f(0)$ for any f bounded,

and because the finite signed measure $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ has mass 1. Recalling that P has support in $\mathbb{N}$, noting that $\mathbb{1}_{\{l\}} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right] = \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{\cdot - l\})$ for any $l \in \mathbb{N}$ and defining $r_l := \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{-l\})$, the last display is

$$\sum_{l=0}^{\infty} \left(r_{i-l}r_{j-l} + p_i p_j (r_{-l} - 1)^2 + (p_i r_{j-l} + p_j r_{i-l}) r_{-l}\right) P(\{l\}).$$

From (2.2.17), we notice that $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ has support in $-\mathbb{N}$ and, furthermore, $r_0 P(\{0\}) = 1$. Therefore, and assuming $i \leq j$, the last display is

$$\begin{aligned} p_i p_j (r_0 - 1)^2 P(\{0\}) + p_i r_j + p_j r_i + p_i p_j (1 - P(\{0\})) + \sum_{l=0}^{i} r_{i-l} r_{j-l} P(\{l\}) \\ = p_i r_j + p_j r_i + p_i p_j (r_0 - 1) + \sum_{l=0}^{i} r_l r_{l+j-i} P(\{i-l\}). \end{aligned}$$

Recalling that $\Delta = 1$ in Buchmann and Grübel [2003], the product of the last display with $(\lambda\Delta)^{-2}$ is the asymptotic covariance in the central limit theorem they obtained —compare it for instance with its empirical version in (4.3.4). An analogous calculation justifies that the limiting covariance between the Gaussian process arising from the mass function and the normal distribution arising from $\tilde{\lambda}_n$ coincides with the one we obtain in Chapter 3.

Recall from the end of Section 2.1.2 that the limiting quantities of the modified estimators in Buchmann and Grübel [2004] coincide with those of Buchmann and Grübel [2003], and hence also with those of Coca [2015], when the mass function is not compactly supported. This raises the following interesting open question: is it possible to show that the analogous transformations for the general estimators of F do not change their asymptotic behaviour when dF is not compactly supported? By all of the discussions above, this should be true and is a useful insight for practical applications as it means that we would know the limiting quantities of the monotone transformations. Moreover, by the Brownian bridge property of the limiting processes they would also be approximately tied up to 1 at the positive end and would therefore be bona-fide distribution functions. However, rigorously answering this question must require finer analysis than that performed in Buchmann and Grübel [2004]: they heavily use the recursive structure of their estimators to show convergence of the finite-dimensional distributions but, in turn, this structure stops them from being able to show tightness of the sequence of estimators.

By the same arguments as above, we conclude that the estimator of the distribution obtained by integrating that of van Es et al. [2007] cannot be efficient unless the intensity is substituted by one of the estimators of Section 2.4. After this substitution and choosing the kernel appropriately, it is the same as ours. Analogously, the estimators of the distribution obtained by integrating the density estimators of Duval [2013a] and Comte et al. [2014] are efficient and, in fact, converge uniformly to the empirical distribution function. Regarding

the efficiency of the individual estimators of the mass of potential atoms of dF, in Chapter 3 we see that the variances of the limiting normal distributions in the central limit theorems for them satisfy the same type of expressions as the lower bounds above. Indeed, the mass at some $x \in \mathbb{R}\setminus\{0\}$ is $\int_{\mathbb{R}} \mathbb{1}_{\{x\}}\, d\nu$ and therefore the variance of the limiting normal distributions associated to these estimators also have the expression that we expect for the lower bound of the model. The estimators of Nickl and Reiß [2012] and Nickl et al. [2016] do not estimate F or N but we can draw some conclusions about them that extend to the more general Lévy processes they estimate: the estimator of $N^G(x)$ of Nickl and Reiß [2012] is forced to be zero both when $x \to \pm\infty$ by definition. However, at $x = \pm\zeta$ it is not and at its values at these points satisfies central limit theorems by the arguments of previous paragraphs. Similarly, the estimators in Nickl et al. [2016] must be of Brownian motion type as they are left free at the positive end point. Therefore, the resulting covariances may decrease when dividing them by the total mass of the estimates and, in practice, this may translate into tighter confidence regions and higher statistical power of the tests developed from them.

Lastly, and although it is far from the scope of this thesis, it would be interesting to try to compute the asymptotic covariance of the integrated density estimators of Gugushvili et al. [2015] and Gugushvili et al. [2016] to compare them to ours. On the other hand, with the machinery recently developed in Castillo and Nickl [2013, 2014], it may be possible to prove a Bernstein–von Mises theorem for the distribution F in this setting. In such a case, the resulting estimators of F from Gugushvili et al. [2015] and Gugushvili et al. [2016], which assume no knowledge of the intensity but estimate it instead, probably do lead to efficient estimators and, at least asymptotically, would be equivalent to ours.

Chapter 3

Theory

This chapter contains the theoretical results in Coca [2015], some of which have already been anticipated in Chapter 2 in a number of sections. In Section 3.1 we recall definitions and notation that are used throughout the chapter and introduce new ones. In Section 3.2 we give the precise assumptions we make to prove our results, and the final form of the estimators that were heuristically introduced in Chapter 2. The statements of our central limit theorems are given in Section 3.3, where we make some additional remarks about them. The rest of the chapter is concerned with proving these results and this is the content of Section 3.4.

3.1 Definitions and notation

We denote by $(\Omega, \mathcal{A}, Pr)$ the probability space on which all the stochastic quantities herein are defined, including the compound Poisson process X from Section 1.1. We define P to be the law of X_Δ or, equivalently, the common law of the independent increments $Z_k := X_{k\Delta} - X_{(k-1)\Delta}$, $k = 1, \dots, n$. Let $P_n := n^{-1} \sum_{k=1}^n \delta_{Z_k}$ be the empirical law of the increments and denote by $\mathcal{F}$ and $\mathcal{F}^{-1}$ the Fourier (-Plancherel) transform and its inverse acting on finite measures or on $L^2(\mathbb{R})$ functions (the reader is referred to Folland [1999] for the results from Fourier analysis and convolution theory used throughout). Then, for all $u \in \mathbb{R}$,

$$\varphi(u) := \mathcal{F}P(u) = \int_{\mathbb{R}} e^{iux} P(dx) \qquad \text{and} \qquad \varphi_n(u) := \mathcal{F}P_n(u) = \frac{1}{n} \sum_{k=1}^n e^{iuZ_k}$$

are the characteristic function of Z and its empirical counterpart. A simple calculation starting from the representation for X_Δ following from expression (1.1.1) shows that

$$\varphi(u) = \exp\Big(\Delta \left(i\gamma u + \mathcal{F}\nu(u) - \lambda \right) \Big), \tag{3.1.1}$$

where ν is a finite measure on $\mathbb{R}$ called the Lévy measure. It satisfies that

$$\int_{-\infty}^{t} \nu(dx) = \lambda F(t) =: N(t), \quad t \in \mathbb{R},$$

where F and N are referred to as the jump distribution and the Lévy distribution. For notational purposes, from now on we generally use t (and s) to denote the 'space' variable, especially when it is the argument of these functions, of covariance functions and of the empirical versions of all of them.

The precise assumptions on ν to prove our results are included in the next section. For now let us mention we assume it can be written as

$$\nu = \nu_d + \nu_{ac}, \tag{3.1.2}$$

where ν_d is a discrete or atomic measure and ν_{ac} is absolutely continuous with respect to Lebesgue's measure. The former satisfies that

$$\nu_d = \sum_{j \in \mathcal{J}} q_j \delta_{J_j}, \tag{3.1.3}$$

where $\mathcal{J}$ is countable, $q_j \in [0, \lambda]$ and $J_j \in \mathbb{R} \setminus \{0\}$ for all $j \in \mathcal{J}$, and we define $q := \sum_{j \in \mathcal{J}} q_j \in [0, \lambda]$. In order to work with the actual weights of the jump measure we define

$$p_j := \frac{q_j}{\lambda} \in [0,1], \text{ for all } j \in \mathcal{J}, \qquad \text{and} \qquad p := \frac{q}{\lambda} \in [0,1]. \tag{3.1.4}$$

Throughout we write $A_c \lesssim B_c$ if $A_c \leqslant C B_c$ holds with a uniform constant C in the parameter c and $A_c \sim B_c$ if $A_c \lesssim B_c$ and $B_c \lesssim A_c$. We use the standard notation $A_n = O(B_n)$ and $A_n = o(B_n)$ to denote that A_n/B_n is bounded and A_n/B_n vanishes, respectively, as $n \to \infty$. We write $\xi_n = O_{Pr}(r_n)$ to denote that $\xi_n r_n^{-1}$ is bounded in Pr-probability.

As introduced in Section 1.3.2, convergence in distribution of real-valued random variables is denoted by $\to^d$. The space of bounded real-valued functions on $\mathbb{R}$ equipped with the supremum norm is denoted by $\ell^\infty(\mathbb{R})$ and $\to^{\mathcal{D}}$ denotes convergence in law in this space (see Dudley [1999], p.94, for its definition). Convergence in law in the product space $\mathbb{R}^{\mathbb{N}} \times \ell^\infty(\mathbb{R})^2$ is denoted by $\to^{\mathcal{D}^\times}$ (see Chapter 1.4 in van der Vaart and Wellner [1996] for convergence in law in product spaces).

3.2 Assumptions and estimators

Assumption. *On the unknown Lévy measure ν we make the following assumptions:*

1 its Lebesgue decomposition is as in (3.1.2)*, satisfying that*

(a) ν_d *is given by* (3.1.3)*, where for some* $\varepsilon > 0$ *fixed,* $J_j = \varepsilon j$ *for all* $j \in \mathcal{J} := \mathbb{Z}\backslash\{0\}$*; and,*

(b) *for all* $s, t \in \mathbb{R}$, $|\int_s^t \nu_{ac}(x)\,dx| \lesssim |s-t|^\alpha$ *for some* $\alpha \in (0,1]$*. And,*

2 $\int_{\mathbb{R}} \log^\beta\left(\max\{|x|, e\}\right)\nu(dx) < \infty$ *for some* $\beta > 2$.

Note that ν may be fully discrete or fully absolutely continuous, which is determined by the (unknown) value of p defined in (3.1.4). The assumption of $\mathrm{supp}(\nu_d) \subseteq \varepsilon \times \mathbb{Z}\backslash\{0\}$ can be relaxed to an ε-separation condition (cf. the end of Remark 3.2.1 below) but we present it as above for simplicity. By excluding 0 from $\mathcal{J}$, we are implicitly assuming ν has no atom at the origin, which we recall is necessary to avoid identifiability issues with the intensity λ. Assumption 1b is immediately satisfied if $\nu_{ac} \in L^r(\mathbb{R})$ for some $r \in (1, \infty]$: take $\alpha = 1 - 1/r$; then, by Hölder's inequality we have that for any $s, t \in \mathbb{R}$

$$\left|\int_s^t \nu_{ac}(x)\,dx\right| \leq \|\nu_{ac}\|_{L^r}\,|s-t|^\alpha.$$

We remark that this assumption can be relaxed to a logarithmic modulus of continuity as found in Coca [2015], although the proofs and statements of the results become more cumbersome. For the sake of clarity, we present them under Assumption 1b. Assumption 2 is also very weak and it is required to guarantee negligibility of some errors.

Regarding the kernel function K that features in our estimators, we assume it is symmetric and it satisfies

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \mathrm{supp}(\mathcal{F}K) \subseteq [-1,1] \quad \text{and} \quad |K(x)| \lesssim (1+|x|)^{-\eta} \text{ for some } \eta > 2. \tag{3.2.1}$$

Therefore the functions $K_{h_n} := h_n^{-1}K(\cdot/h_n)$, where h_n is referred to as the bandwidth, are continuous, have Fourier transform $\mathcal{F}K_{h_n}$ supported in $[-h_n^{-1}, h_n^{-1}]$ and provide an approximation to the identity operator as $h_n \to 0$ when $n \to \infty$.

Recall that we constructed our estimators from the formal identity

$$\frac{1}{\Delta}\mathcal{F}^{-1}[\mathrm{Log}(\varphi)] = \gamma\mathcal{F}^{-1}[i\cdot] + \nu - \lambda\delta_0, \tag{3.2.2}$$

which follows by taking the distinguished logarithm on both sides of (3.1.1) and then the Fourier inverse transform. Based on the observation that the mass of ν is λ, the idea to construct the estimator of the intensity is to take the integral of both sides excluding a small interval around the origin. Therefore, we propose to estimate it by the first estimator of Section 2.4.1,

$$\hat{\lambda}_n := \frac{1}{\Delta}\int_{\mathbb{R}} f_n^{(\lambda)}(x)\,\mathcal{F}^{-1}[\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}](x)\,dx, \qquad f_n^{(\lambda)}(x) := \mathbb{1}_{\{\varepsilon_n \leq |x| \leq H_n\}},$$

where $\varepsilon_n \to 0$ and $H_n \to \infty$ are determined in the next section. For now we remark that

$h_n = o(\varepsilon_n)$ is necessary to annihilate the first and last terms in (3.2.2). This, together with the fact that the smallest jumps of ν_d are at least at distance $\varepsilon > 0$ from the origin and with the uniform Hölder continuity of ν_{ac}, guarantees that $\hat{\lambda}_n$ recovers all the mass of ν as $n \to \infty$. In addition, we also require that $h_n = o(H_n^{-1})$ to control some errors. The truncation of the tails is natural because when implementing the estimator the integral has to be truncated. Furthermore, it is key to be able to prove our results under Assumption 2 instead of under a finite polynomial moment condition. This assumption, together with an appropriate choice of the bandwidth h_n, also allows us to control the fluctuations of φ_n around φ in the sense that $\sup_{u\in[-h_n^{-1},h_n^{-1}]} |\varphi_n(u) - \varphi(u)| \to 0$ in Pr-probability. Recalling that

$$0 < e^{-2\lambda\Delta} \leq |\varphi(u)| \leq 1 \quad \text{for all } u \in \mathbb{R},$$

we conclude that the estimator of the intensity we just proposed is well-defined in sets of Pr-probability approaching 1 as $n \to \infty$. These remarks also apply to the rest of the estimators we now propose.

In Section 2.4.2 we proposed to estimate the drift by

$$\hat{\gamma}_n := \frac{1}{c\Delta} \int_{\mathbb{R}} f_n^{(\gamma)}(x)\, \mathcal{F}^{-1}[\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}](x)\, dx, \qquad f_n^{(\gamma)}(x) := \mathbb{1}_{|x|\leq h_n}\, x,$$

where $c := \int_{-1}^{1} K - (K(1) + K(-1)) = 2\big(\int_0^1 K - K(1)\big)$ in view of the symmetry of K. We refer the reader to that section for more details on its construction. In Section 2.2.2 we argued that to construct our estimator of F we first construct an estimator of N and then divide by $\hat{\lambda}_n$. Including the modifications of this chapter, we proposed to estimate the Lévy distribution at $t \in \mathbb{R}$ by

$$\frac{1}{\Delta} \int_{\mathbb{R}} \mathbb{1}_{(-\infty,t]}(x) f_n^{(\lambda)}(x)\, \mathcal{F}^{-1}[\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}](x)\, dx. \tag{3.2.3}$$

Notice that, due to the regularisation induced by K, this quantity is continuous in t in sets of Pr-probability approaching 1 as $n \to \infty$. Recall that we want to show a functional central limit theorem under the standard supremum norm. Hence, if we want to include a potential discrete component, as allowed by Assumption 1a, we need to make the estimator discontinuous. The general idea Coca [2015] proposed is to estimate the mass of each atom separately and modify the last display accordingly and exactly at the potential atoms. To estimate the mass of a potential atom we rewrite (3.2.2) as

$$\frac{1}{\Delta}\mathcal{F}^{-1}[\mathrm{Log}(\varphi)] = \gamma\mathcal{F}^{-1}[i\cdot] + \sum_{j\in\mathbb{Z}\setminus\{0\}} q_j\, \delta_{J_j} + \nu_{ac} - \lambda\, \delta_0. \tag{3.2.4}$$

Therefore, and by the uniform Hölder continuity of ν_{ac}, it is clear that q_j, $j \in \mathbb{Z}\setminus\{0\}$, is recovered by integrating this expression over a small enough interval around J_j. Indeed,

we propose to estimate the mass of a potential atom at J_j by

$$\hat{q}_{j,n} := \frac{1}{\Delta}\int_{\mathbb{R}} f_n^{(q_j)}(x)\,\mathcal{F}^{-1}[\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}]\,(x)\,dx, \qquad f_n^{(q_j)}(x) := \mathbb{1}_{\{|x-J_j|<\varepsilon_n\}},$$

where we recall that $\varepsilon_n \to 0$ and thus lose no generality by assuming it is smaller than $\varepsilon/2$. We note in passing that this is the reason to choose a similar notation for these two quantities. Then, the idea to estimate N is to force it to have a discontinuity at each J_j in the integration interval equal to $\hat{q}_{j,n}$. Consequently, we propose

$$\widehat{N}_n(t) := \frac{1}{\Delta}\int_{-\infty}^{t} \mathbb{1}_{\{|x|\leq H_n\}}\Big(1-\sum_{j\in\mathbb{Z}}\mathbb{1}_{\{|x-J_j|<\varepsilon_n\}}\Big)\mathcal{F}^{-1}\big[\,\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}\big](x)\,dx + \sum_{\substack{|j|\leq H_n/\varepsilon\\ j\leq t/\varepsilon,\, j\neq 0}} \hat{q}_{j,n}. \tag{3.2.5}$$

The first term is estimating the integral of the absolutely continuous component of ν, whilst the second is estimating the cumulative discrete component. If one of the two is zero, the corresponding term is asymptotically negligible uniformly in t. In fact, if we have a priori knowledge of, say, ν being purely atomic, the first term can be discarded when estimating it and the conclusions herein still follow under the same assumptions. If, instead, we know in advance that ν_d is null, we can discard the second term, although in such a case it is simpler to use the expression in (3.2.3) directly. Without loss of generality we can assume that $|H_n - J_j| > \varepsilon_n$ for all $j \in \mathbb{Z}$, and the last display can be alternatively written as

$$\widehat{N}_n(t) = \frac{1}{\Delta}\int_{\mathbb{R}} f_{t,n}^{(N)}(x)\,\mathcal{F}^{-1}\big[\,\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}\big](x)\,dx,$$

where

$$f_{t,n}^{(N)} = \mathbb{1}_{[-H_n,H_n]\setminus(-\varepsilon_n,\varepsilon_n)} \times \begin{cases} \mathbb{1}_{(-\infty,t]} & \text{if} \quad |t-J_j|>\varepsilon_n \quad \text{for all } j\in\mathbb{Z},\\ \mathbb{1}_{(-\infty,J_j-\varepsilon_n]} & \text{if} \quad J_j-\varepsilon_n\leq t<J_j \quad \text{for some } j\in\mathbb{Z},\\ \mathbb{1}_{(-\infty,J_j+\varepsilon_n]} & \text{if} \quad J_j\leq t\leq J_j+\varepsilon_n \quad \text{for some } j\in\mathbb{Z}. \end{cases} \tag{3.2.6}$$

In view of either expression for $\widehat{N}_n(t)$, when t is away from a potential jump it grows continuously and it carries all the estimated mass of ν_{ac} and ν_d up to t. However, when t is close enough to one of these, say J_j, the first term in the first expression for the estimator remains constant whilst the second introduces a jump equal to the estimate of q_j exactly at $t = J_j$. The truncation around the origin guarantees the terms arising from the first and last summand in (3.2.2) are negligible. As discussed in the previous chapter (cf. (2.4.1) and (2.4.2)), it also corresponds to the common practice in the literature of discarding the increments in which no jumps have taken place, as these do not carry any information about the jump distribution. They do however contain information about the

intensity through the total number of them in a sample of size n and this quantity is kept in the estimators as it is used to decompound the observations appropriately. In analogy with the remarks after introducing $\hat{\lambda}_n$ above, the tail truncation is even more natural here because one can only compute a finite number of estimators of the potential discrete component. Finally, we estimate the rest of the parameters as

$$\widehat{F}_n := \hat{\lambda}_n^{-1}\widehat{N}_n \quad \text{and} \quad \hat{p}_{j,n} := \hat{\lambda}_n^{-1}\hat{q}_{j,n}, \quad j \in \mathbb{Z}\setminus\{0\}. \tag{3.2.7}$$

Therefore, for t negative enough $\widehat{F}_n(t) = 0$ and, due to the expressions for $\hat{\lambda}_n$ and $\widehat{N}_n$, $\widehat{F}_n(t) = 1$ for t large enough. We remark that, despite the appearance of distinguished logarithms and Fourier transforms, all the estimators are real valued for any n.

The strategy to estimate a potential discrete component in ν we just proposed has an additional advantage: we can estimate the total mass of the discrete component (and hence of the absolutely continuous component too) and, moreover, we can derive tests for the presence of either component. We postpone the construction of the latter to Chapter 4 and, for now, propose to estimate q by

$$\hat{q}_n := \sum_{\substack{|j|\leq \widetilde{H}_n/\varepsilon \\ j\neq 0}} \hat{q}_{j,n} = \frac{1}{\Delta}\int_{\mathbb{R}} f_n^{(q)}(x)\,\mathcal{F}^{-1}[\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}](x)\,dx, \quad f_n^{(q)}(x) := \sum_{\substack{|j|\leq \widetilde{H}_n/\varepsilon \\ j\neq 0}} \mathbb{1}_{\{|x-J_j|<\varepsilon_n\}},$$

where $\widetilde{H}_n \to \infty$ at a slower rate than H_n explicitly given in the next section. In line with the transformations above, p can then be estimated by

$$\hat{p}_n := \hat{\lambda}_n^{-1}\hat{q}_n.$$

3.2.1 Remark. *The transformation of* (3.2.3) *to accommodate a potential discrete component we just proposed is not using any property of the compound Poisson process. Instead, it is exploiting the fact that we have an empirical version of the target measure, namely* $\frac{1}{\Delta}\mathbb{1}_{(-\varepsilon_n,\varepsilon_n)^C}\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}$. *Therefore, the ideas here can be applied in much more general settings, including inverse and noninverse problems, as long as one has access to such an estimate. To the best of our knowledge, it has only been proposed by Coca [2015].*

It does however have two implicit drawbacks that we would like to discuss. The first is that, even though we do not assume we know in advance whether a discrete component is present in F, *we do assume knowledge of a superset of* $\mathrm{supp}(\nu_d)$, *since any* q_j *may be zero but the corresponding* J_j *is still included in* (3.2.5) *and* (3.2.6). *In fact, the issue is more subtle and hidden behind our assumption of* $\mathcal{J}$ *being equispaced: when dealing with* $\widehat{N}_n(t)$ *in the proofs, the quantity* $\mathbb{1}_{(-\infty,t]} * K_{h_n}$ *appears and in its limit as* $n \to \infty$ *a Gibbs phenomenon naturally features at* t. *This can be ignored in problems where this limit is integrated with respect to absolutely continuous measures but not here: in the asymptotic covariances of the limit theorems below, the integrators are* $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ *and* P, *which*

were written as infinite series in (2.6.1) *and* (2.6.2)*, and these may have atoms at all of the values of finitely many sums and differences of the atoms of* ν_d*. Therefore, we need to deal with this issue at all of these points and by assuming* $\mathcal{J}$ *is equispaced we avoided speaking about the recombinations of the atoms. One of the purposes of the transformation of* $\mathbb{1}_{(-\infty,t]}$ *in* (3.2.6) *is precisely to avoid the aforementioned phenomenon at every* $t = J_j$ *and it therefore guarantees we achieve the optimal covariances below. We emphasise that assuming this prior knowledge may be avoided by purely data-driven procedures based on the tests that arise from our results for the atomic component in the next section. This is a relevant point in fully automated procedures but, nevertheless, out of the scope of this thesis. Furthermore, in practice it may be avoided in a non-rigorous but simple way: the finite-sample implementation of the estimators of* N *and* F *requires the discretisation of the integral in them. This acts as* (3.2.6) *for every interval in the discretisation grid and hence approximately but automatically performs the desired transformation at recombinations of the atoms. Indeed, the acceptable practical behaviour of directly discretising* (3.2.3) *when there is a discrete component instead of using* (3.2.6) *is illustrated in Section 4.4. The second implicit limitation of our strategy is that we require a minimal separation between the atoms* J_j*. This a natural assumption given the modification we make at each potential atom in* (3.2.6) *and is not very restrictive in practice since in the implementation of* N *and* F *we have to choose a finite discrete grid for the integrals in them. We need it in our proofs since we require*

$$\sum_{j\in\mathcal{J}} p_j\, J_j^{-\eta} < \infty \quad \text{and} \quad \sum_{j\in\mathcal{J}\setminus\{l\}} p_j\, |J_j - J_l|^{1-\eta} < \infty,$$

for any $l \in \mathcal{J}$*, any discrete distribution* $(p_j)_{j\in\mathcal{J}}$ *and any* $\eta > 2$*.*

Having discussed these two limitations we can now show how Assumption 1a can be straightforwardly relaxed. Let $\mathbb{J}$ *be equal to or a superset of the support of the discrete component together with the origin, so* $\mathbb{J} := \{J_0 = 0\} \cup \{J_j : j \in \mathcal{J}\}$ *for some* $\mathcal{J} \subseteq \mathbb{Z}\setminus\{0\}$ *(Assumption 1a implies* $\mathbb{J} = \varepsilon \times \mathbb{Z}$*). Additionally, let* $\bar{\mathbb{J}}$ *be the closure of* $\mathbb{J}$ *under arbitrary but finite sums and differences of its elements and assume that for some* $\varepsilon > 0$ *it is such that* $|J_j - J_l| > \varepsilon$ *for any* $J_j, J_l \in \bar{\mathbb{J}}$*. Then, assuming* $\varepsilon_n < \varepsilon/2$ *without loss of generality, we should make the transformation* (3.2.6) *at every element of* $\bar{\mathbb{J}}$ *and the conclusions of the next section remain true. Adapting the proofs to this more general setting only requires cosmetic changes based on the remarks above.*

3.3 Central limit theorems

Our main result concerns joint estimation of all the parameters introduced above. However, to ease the exposition we first specialise it to the individual parameters and interpret the results separately.

3.3.1 Proposition. *Suppose Assumptions 1 and 2 are satisfied for some $\beta > 2$ and that K satisfies* (3.2.1)*. Let $h_n \sim \exp(-n^{\vartheta_h})$, $\varepsilon_n \sim \exp(-n^{\vartheta_\varepsilon})$ and $H_n \sim \exp(n^{\vartheta_H})$, where $0 < \vartheta_\varepsilon < \vartheta_h < 1/4$ and $1/(2\beta) \leq \vartheta_H < \vartheta_h$. Then, under the notation at the end of Section 3.1, we have that as $n \to \infty$*

$$\sqrt{n}\left(\hat{\lambda}_n - \lambda\right) \to^d N\big(0, \sigma^2_\lambda\big) \quad \textit{and} \quad \sqrt{n}\, h_n^{-1}\big(\hat{\gamma}_n - \gamma\big) \to^{Pr} 0,$$

where in what follows $N(0, \sigma^2)$ denotes a zero-mean normal distribution with variance σ^2 and, writing $f^{(\lambda)} := \mathbb{1}_{\mathbb{R}\setminus\{0\}}$ and $\varphi^{-1} = 1/\varphi$ throughout,

$$\sigma^2_\lambda := \frac{1}{\Delta^2}\int_{\mathbb{R}} \Big(f^{(\lambda)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)^2 P(dx).$$

The expressions of $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ and P can be found in (2.6.1) and (2.6.2), respectively. The fact that they are both finite measures guarantees the finiteness of (2.6.3) for any bounded functions f and g. All the variances and covariances of this section, including σ^2_λ above, have that form and are therefore finite. Furthermore, as argued there, any such quantity does not depend on γ and this justifies the fact that in the following discussions γ does not play any role, even though we are not assuming $\gamma = 0$. The next lemma allows us to rigorously justify some of the intuitive remarks we make throughout this thesis.

3.3.2 Lemma. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be bounded functions and let $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ and P be given by* (2.6.1) *and* (2.6.2)*, respectively. Then*

$$\begin{aligned}
&\int_{\mathbb{R}} \Big(f * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)\Big(g * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big) P(dx) \\
&\quad = f(0)g(0) + \Delta\bigg(\int_{\mathbb{R}} f(x)g(x)\,\nu(dx) - f(0)\,g * \bar{\nu}\,(0) - g(0)\,f * \bar{\nu}\,(0) + \lambda\, f(0)g(0)\bigg) \\
&\qquad + O\big((\lambda\Delta)^2\big).
\end{aligned}$$

In particular, if $f(0) = g(0) = 0$, the last display is

$$\lambda\Delta \int_{\mathbb{R}} f(x)g(x)\,F(dx) + O\big((\lambda\Delta)^2\big).$$

In Section 2.4.1 we showed that, when no exact cancellations between jumps can arise from ν, such as when it has no discrete component in $\mathbb{R}^+$ and/or $\mathbb{R}^-$,

$$\sigma^2_\lambda = \frac{e^{\lambda\Delta} - 1}{\Delta^2}.$$

Therefore, when $\lambda\Delta$ (the expected number of jumps in a Δ-observation window) is small,

$$\sqrt{n\lambda\Delta}\left(\hat{\lambda}_n - \lambda\right) \approx N(0, \lambda^2).$$

For a general ν this is justified by Lemma 3.3.2. Indeed, in Section 2.4.1 we showed that, as $\Delta \to 0$, $\hat{\lambda}_n$ tends to the maximum likelihood estimator of Section 1.3.1, which satisfies (1.3.1). The intuition behind the conclusion of Proposition 3.3.1 for the estimator of the drift was given in Section 2.4.2.

The following proposition deals with individual estimation of the potential weights of the discrete component of N and F. This result already hints at the asymptotic differences between estimating these two functions mentioned in Section 2.6.

3.3.3 Proposition. *Under the assumptions and notation of Proposition 3.3.1 we have that for any $j \in \mathbb{Z}\setminus\{0\}$ and as $n \to \infty$*

$$\sqrt{n}\left(\hat{q}_{j,n} - q_j\right) \to^d N\big(0, \sigma^2_{q_j}\big) \quad \text{and} \quad \sqrt{n}\left(\hat{p}_{j,n} - p_j\right) \to^d N\big(0, \sigma^2_{p_j}\big),$$

where, defining $f^{(q_j)} := \mathbb{1}_{\{J_j\}}$ and $f^{(p_j)} := \lambda^{-1}\big(f^{(q_j)} - p_j f^{(\lambda)}\big)$,

$$\sigma^2_{q_j} := \frac{1}{\Delta^2}\int_{\mathbb{R}}\Big(f^{(q_j)} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\Big)^2 P(dx)$$

and

$$\sigma^2_{p_j} := \frac{1}{\Delta^2}\int_{\mathbb{R}}\Big(f^{(p_j)} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\Big)^2 P(dx).$$

Similarly to the expression for σ^2_λ in (2.4.5), we have that, for any $j \in \mathbb{Z}\setminus\{0\}$,

$$\sigma^2_{q_j} = \frac{1}{\Delta^2}\sum_{l\in\mathbb{Z}} P(\{J_l\})\Big(\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{J_l - J_j\})\Big)^2$$

and

$$\sigma^2_{p_j} = \frac{1}{\Delta^2\lambda^2}\bigg(\sum_{l\in\mathbb{Z}} P(\{J_l\})\Big(\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{J_l - J_j\}) + p_j\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](\{J_l\})\Big)^2 - p_j^2\bigg).$$

As expected, these quantities are zero at jumps that cannot be observed and, in particular, when ν is absolutely continuous ($p = 0$) they are zero for all $j \in \mathbb{Z}\setminus\{0\}$. To gain more intuition about them we turn to the case when $\lambda\Delta$ is sufficiently small. From Lemma 3.3.2, we have that

$$\sqrt{n\Delta}\left(\hat{q}_{j,n} - q_j\right) \approx N\big(0, q_j\big) \quad \text{and} \quad \sqrt{n\lambda\Delta}\left(\hat{p}_{j,n} - p_j\right) \approx N\big(0, p_j(1-p_j)\big).$$

In this case, most if not all of the jumps are observed directly in the limit so little or no decompounding is needed and the asymptotic variances do not depend on weights other than the one under consideration. The interpretation of the dependence on p_j of the variances in both displays is the following. Generally, when $p_j \approx 0$ only a small percentage of increments Z_k equal J_j, so the variability in the estimate is not large. When $p_j \approx 1$ we

expect many increments to be equal to J_j and therefore the variability in the estimate of p_j is small too. However, in this case the parameter $q_j = \lambda p_j$ is approximately λ and hence it carries almost the same variability as the estimate for the intensity given in Proposition 3.3.1. This uncertainty is removed when estimating p_j because $\hat{p}_{j,n} := \hat{q}_{j,n}/\hat{\lambda}_n$.

When $\lambda\Delta$ is not small and $p \neq 0$ the inverse nature of the problem changes the dependence of $\sigma^2_{q_j}$ and $\sigma^2_{p_j}$ on p_j. In view of their general expressions above, if $P(\{J_j\})$, the probability of observing a jump J_j, is not zero, they are not necessarily zero even when $p_j = 0$. This is because despite no underlying discrete jumps equal J_j, the increments Z_k do not capture this and some decompounding is required, therefore introducing some unavoidable uncertainty.

We now address the estimation of N and F, the main focus of this thesis.

3.3.4 Theorem. *Under the assumptions of Proposition 3.3.1 and the notation at the end of section 3.1, we have that as $n \to \infty$*

$$\sqrt{n}\left(\widehat{N}_n - N\right) \to^{\mathcal{D}} \mathbb{B}^N \quad \text{in } \ell^\infty(\mathbb{R}) \qquad \text{and} \qquad \sqrt{n}\left(\widehat{F}_n - F\right) \to^{\mathcal{D}} \mathbb{G}^F \quad \text{in } \ell^\infty(\mathbb{R}),$$

where, defining $f_t^{(N)} := \mathbb{1}_{(-\infty,t]}\,\mathbb{1}_{\mathbb{R}\setminus\{0\}}$ for any $t \in \mathbb{R}$ and

$$f_t^{(F)} := \lambda^{-1}\big(f_t^{(N)} - F(t)f^{(\lambda)}\big) = \lambda^{-1}\big(\mathbb{1}_{(-\infty,t]} - F(t)\big)\,\mathbb{1}_{\mathbb{R}\setminus\{0\}},$$

$\mathbb{B}^N$ and $\mathbb{G}^F$ are tight centred Gaussian Borel random variables in $\ell^\infty(\mathbb{R})$ with covariance structures satisfying that for any $s, t \in \mathbb{R}$

$$\Sigma^N_{s,t} := \frac{1}{\Delta^2}\int_{\mathbb{R}} \Big(f_s^{(N)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)\Big(f_t^{(N)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)P(dx) \tag{3.3.1}$$

and[1]

$$\Sigma^F_{s,t} := \frac{1}{\Delta^2}\int_{\mathbb{R}} \Big(f_s^{(F)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)\Big(f_t^{(F)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)P(dx). \tag{3.3.2}$$

In Chapter 2 we discussed this result and its consequences in depth. In particular, in Section 2.6 we pointed out the meaning and necessity of the quantity $\mathbb{1}_{\mathbb{R}\setminus\{0\}}$ in the covariances, the fact that they coincide with the information lower bounds of the problem, and the additional insights they give regarding efficiency. Therefore, we simply mention that in view of Lemma 3.3.2,

$$\Delta\,\Sigma^N_{s,t} = N(s \wedge t) + O(\Delta) \quad \text{and} \quad \lambda\Delta\,\Sigma^F_{s,t} = \big(F(\min\{s,t\}) - F(s)F(t)\big) + O(\Delta).$$

[1]Note that in Chapter 2, the covariances $\Sigma^N_{s,t}$ and $\Sigma^F_{s,t}$ were denoted by $\Sigma^{\mathbb{B}^N}_{s,t}$ and $\Sigma^{\mathbb{G}^F}_{s,t}$. From now on we simplify the notation as no confusion regarding the limiting processes they are connected to may arise.

Therefore, expressions (2.6.7) and (2.6.5) are justified, and, for $\lambda\Delta$ small,

$$\sqrt{n\Delta}\left(\widehat{N}_n - N\right) \quad \text{and} \quad \sqrt{n\lambda\Delta}\left(\widehat{F}_n - F\right)$$

are approximately centred Gaussian processes with covariance functions given by the leading terms of the second to last display. This agrees with our calculation at the end of Section 2.3 in which we showed that $\widehat{N}_n$ converges to the estimators of Nickl et al. [2016] as $\Delta \to 0$. It also agrees with the remark therein stating that the same calculation guarantees our estimator $\widehat{F}_n$ converges to the empirical distribution function as $\Delta \to 0$.

All the results included so far are particular cases of the following, which establishes joint convergence of all the estimators and is our main result. Prior to stating it let us introduce some notation. For any $L \in \mathbb{R}^4 \times \left(\mathbb{R}^{\mathbb{Z}\setminus\{0\}}\right)^2 \times \left(\ell^\infty(\mathbb{R})\right)^2 \cong \mathbb{R}^{\mathbb{N}} \times \ell^\infty(\mathbb{R})^2$ we denote its first four coordinates by $L_\lambda, L_\gamma, L_q, L_p$, its j-th coordinate within the first and second space $\mathbb{R}^{\mathbb{Z}\setminus\{0\}}$ by L_{q_j} and L_{p_j}, respectively, and the evaluation at $t \in \mathbb{R}$ of its penultimate and last coordinate by $L_N(t)$ and $L_F(t)$, respectively. Finally, let $\underline{q}, \underline{p} \in \mathbb{R}_+^{\mathbb{Z}\setminus\{0\}}$ be row vectors with j-th entry equal to q_j and p_j, respectively, and let $\underline{\hat{q}}_n$ and $\underline{\hat{p}}_n$ be their coordinate-wise estimators.

3.3.5 Theorem. *Let $\mathbb{L}$ be a tight centred Gaussian random variable on $\mathbb{R}^{\mathbb{N}} \times \ell^\infty(\mathbb{R})^2$ and note that it is fully characterised by its finite dimensional distributions. For any $c_{(\lambda)}, c_{(\gamma)}, c_{(q)}, c_{(p)} \in \mathbb{R}$, $M_\theta \in \mathbb{N}$, $\underline{C}_\theta \in \mathbb{R}^{M_\theta}$, where $\theta \in \{q, p, N, F\}$, and any $i_1, \ldots, i_{M_q}, j_1, \ldots, j_{M_p} \in \mathbb{N}, s_1, \ldots, s_{M_N}, t_1, \ldots, t_{M_F} \in \mathbb{R}$, assume it satisfies that*

$$\left(c_{(\lambda)}, c_{(\gamma)}, c_{(q)}, c_{(p)}\right)\begin{pmatrix}\mathbb{L}_\lambda\\ \mathbb{L}_\gamma\\ \mathbb{L}_q\\ \mathbb{L}_p\end{pmatrix} + \underline{C}_q\begin{pmatrix}\mathbb{L}_{q_{i_1}}\\ \vdots\\ \mathbb{L}_{q_{i_{M_q}}}\end{pmatrix} + \underline{C}_p\begin{pmatrix}\mathbb{L}_{p_{j_1}}\\ \vdots\\ \mathbb{L}_{p_{j_{M_p}}}\end{pmatrix} + \underline{C}_N\begin{pmatrix}\mathbb{L}_N(s_1)\\ \vdots\\ \mathbb{L}_N(s_{M_N})\end{pmatrix} + \underline{C}_F\begin{pmatrix}\mathbb{L}_F(t_1)\\ \vdots\\ \mathbb{L}_F(t_{M_F})\end{pmatrix}$$

is a one-dimensional normal random variable with mean zero and variance

$$\sigma_{\mathbb{L}}^2 := \frac{1}{\Delta^2}\int_{\mathbb{R}} \left(f^{(\mathbb{L})} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\right)^2 P(dx), \tag{3.3.3}$$

where $f^{(\mathbb{L})} = f^{(\mathbb{L})}(c_{(\lambda)}, c_{(q)}, c_{(p)}, \underline{C}_q, \underline{C}_p, \underline{C}_N, \underline{C}_F)$ is defined as

$$f^{(\mathbb{L})} := \left(c_{(\lambda)}, c_{(q)}, c_{(p)}\right)\begin{pmatrix}f^{(\lambda)}\\ f^{(q)}\\ f^{(p)}\end{pmatrix} + \underline{C}_q\begin{pmatrix}f^{(q_{i_1})}\\ \vdots\\ f^{(q_{i_{M_q}})}\end{pmatrix} + \underline{C}_p\begin{pmatrix}f^{(p_{j_1})}\\ \vdots\\ f^{(p_{j_{M_p}})}\end{pmatrix} + \underline{C}_N\begin{pmatrix}f_{s_1}^{(N)}\\ \vdots\\ f_{s_{M_N}}^{(N)}\end{pmatrix} + \underline{C}_F\begin{pmatrix}f_{t_1}^{(F)}\\ \vdots\\ f_{t_{M_F}}^{(F)}\end{pmatrix}.$$

(a) Under the assumptions of Proposition 3.3.1 we have that as $n \to \infty$

$$\sqrt{n}\left(\hat{\lambda}_n - \lambda,\, h_n^{-1}(\hat{\gamma}_n - \gamma),\, \underline{\hat{q}}_n - \underline{q},\, \underline{\hat{p}}_n - \underline{p},\, \widehat{N}_n - N,\, \widehat{F}_n - F\right) \to^{\mathcal{D}^\times} \mathbb{L}_{-q,-p},$$

where $\mathbb{L}_{-q,-p}$ *denotes the same random variable as* $\mathbb{L}$ *without its third and fourth coordinates.*

(b) Suppose ν *satisfies Assumption 1a and* $\int_{\mathbb{R}} |x|^\beta \nu(dx) < \infty$ *for some* $\beta > 1$*. Assume* K *satisfies* (3.2.1) *and let* $h_n = \exp(-n^{\vartheta_h}), \varepsilon_n = \exp(-n^{\vartheta_\varepsilon}), H_n = \exp(n^{\vartheta_H})$ *and* $\widetilde{H}_n \sim n^{\vartheta_{\widetilde{H}}}$*, where* $1/(2\beta) \leq \vartheta_{\widetilde{H}} < 1/2$*,* $0 < \vartheta_\varepsilon < \vartheta_h < (1-2\vartheta_{\widetilde{H}})/4$ *and* $0 < \vartheta_H < \vartheta_h$*. Then, as* $n \to \infty$*,*

$$\sqrt{n}\left(\hat{\lambda}_n - \lambda,\, h_n^{-1}(\hat{\gamma}_n - \gamma),\, \hat{q}_n - q,\, \hat{p}_n - p,\, \underline{\hat{q}}_n - \underline{q},\, \underline{\hat{p}}_n - \underline{p},\, \widehat{N}_n - N,\, \widehat{F}_n - F\right) \to^{\mathcal{D}^\times} \mathbb{L}.$$

This theorem provides a large number of inference procedures that are described in Chapter 4. The empirical version of $\sigma^2_{\mathbb{L}}$ is a fundamental quantity in their construction. The procedures include estimation and testing of single parameters, which can already be easily devised from the propositions above, and goodness-of-fit and two-sample tests for the whole process. We regard the stronger assumption of part (b) of the theorem as purely technical. We require it to show convergence of the finite-dimensional distributions: in part (a), we need to control a finite fixed number of projections of the quantity considered therein but, in part (b), we effectively have to control a growing number of projections due to the expressions for $\hat{q}_n$ and $\hat{p}_n$.

3.4 Proofs

Joint convergence as described in Theorem 3.3.5 follows from convergence of each of the coordinates of the infinite vectors therein together with joint convergence of the one dimensional parameters and the finite dimensional distributions of the infinite dimensional parameters. Therefore we need to prove joint convergence of finitely and infinitely many one-dimensional distributions. This is the content of Section 3.4.1 and it includes several results that may be of independent interest. Section 3.4.2 is devoted to proving Propositions 3.3.1 and 3.3.3, which follow from the results of Section 3.4.1, and in Sections 3.4.3 and 3.4.4 we prove Theorems 3.3.4 and 3.3.5, respectively. Lastly, we include the proof of Lemma 3.3.2.

3.4.1 Joint convergence of one-dimensional distributions

The last and main result of this section requires to identify the asymptotic limit of some stochastic quantities and to control some non-stochastic quantities. We follow this order throughout and before introducing the generic central limit theorem dealing with the stochastic quantities, namely Theorem 3.4.5, we develop several auxiliary results.

Note that the observations of the compound Poisson process appear in our estimators through the empirical characteristic function φ_n. Since it is always multiplied by $\mathcal{F}K_{h_n}$,

supported in $[-h_n^{-1}, h_n^{-1}]$, we have to control the uniform norm of $|\varphi_n(u) - \varphi(u)|$ when u varies over these sets. Theorem 4.1 in Neumann and Reiß [2009] gives sufficient conditions to control this quantity and derivatives of it. In particular, it guarantees that if $\int |x|^\beta \nu(x)dx < \infty$ for some $\beta > 0$ then, under the notation at the end of Section 3.1,

$$\sup_{|u| \leqslant h_n^{-1}} |\varphi_n(u) - \varphi(u)| = O_{Pr}\big(n^{-1/2} \log^{1/2+\delta}(h_n^{-1})\big) \tag{3.4.1}$$

for any $\delta > 0$. The assumption for this particular case of the result can be slightly refined as follows.

3.4.1 Theorem. *Suppose that $\xi_1, \xi_2, \ldots$ are independent and identically distributed real-valued random variables satisfying $\mathbb{E}[\log^\beta(\max\{|\xi_1|, e\})] < \infty$ for some $\beta > 1$ and let $w : \mathbb{R} \to [0, \infty)$ be a weight function with $w(u) \leq \log^{-1/2-\delta}(e + |u|)$ for some $\delta > 0$. Then we have that*

$$\sup_{n \geq 1} \mathbb{E}\Big[\sup_{u \in \mathbb{R}} \big\{ \sqrt{n}\, |\varphi_n(u) - \varphi(u)|\, w(u) \big\} \Big] < \infty, \tag{3.4.2}$$

and, from Markov's inequality, (3.4.1) *holds.*

Taking $\xi_k = Z_k$, $k = 1, \ldots, n$, the assumption of the theorem is that $\int_{\mathbb{R}} \log^\beta(\max\{|x|, e\})\, P(dx) < \infty$ for some $\beta > 1$. By Theorem 25.3 and Proposition 25.4 in Sato [1999] it is equivalent to assuming the same moment is finite when substituting P by ν in its expression and therefore it is satisfied under Assumption 2. As mentioned in Neumann and Reiß [2009], this result goes in line with the well known fact proved by Csörgő and Totik [1983]: almost sure uniform convergence of φ_n towards φ on increasingly long symmetric intervals $[-U_n, U_n]$ holds whenever $\log(U_n)/n \to 0$.

We show this theorem by slightly refining one of the steps in the proof of Theorem 4.1 in Neumann and Reiß [2009]. Therefore we give a brief overview of their ideas and take no credit for them. We refer the reader to their work for more details.

Proof. The strategy Neumann and Reiß [2009] adopt to prove the result is to control the expectation of the supremum over $u \in \mathbb{R}$ of the empirical process $\sqrt{n} \exp(i \cdot u) w(u)(P_n - P)$ by a maximal inequality from empirical process theory. Namely, they make use of Corollary 19.35 from van der Vaart [1998] and thus they have to control the L^2-bracketing number of a certain class of functions. Inspired by Yukich [1985] they bound this bracketing number by an expression depending on the quantity

$$M = M(\epsilon, k) := \inf\Big\{ m \geq 1 : \mathbb{E}[\xi_1^{2k} \mathbb{1}_{|\xi_1| > m}] \leq \epsilon^2 \Big\},$$

where $k \in \mathbb{N}$ is fixed and we are only concerned with the case $k = 0$. By the lemma after

Theorem 2 in Yukich [1985], the theorem follows if

$$\int_0^1 \sqrt{\log M(\epsilon, k)}\, d\epsilon < \infty. \tag{3.4.3}$$

For $k = 0$ the expectation in the definition of M simplifies to $P(|\xi_1| > m)$. Then, trivially $|y| \leq \max\{|y|, e\}$ for any $y \in \mathbb{R}$ and, by Markov's inequality, we have for any $\beta > 1$

$$P(|\xi_1| > m) \leq \frac{\mathbb{E}[\log^\beta(\max\{|\xi_1|, e\})]}{\log^\beta(m)}.$$

Once Neumann and Reiß [2009] derive an upper bound of this form the result follows because

$$M(\epsilon, 0) \leq \inf\left\{ m \geq 1 : \frac{\mathbb{E}[\log^\beta(\max\{|\xi_1|, e\})]}{\log^\beta(m)} \leq \epsilon^2 \right\}$$

and hence

$$\log M(\epsilon, 0) \leq \epsilon^{-2/\beta}\, \mathbb{E}[\log^\beta(\max\{|\xi_1|, e\})]^{1/\beta}.$$

Therefore, (3.4.3) is satisfied when assuming $\mathbb{E}[\log^\beta(\max\{|\xi_1|, e\})] < \infty$ for some $\beta > 1$. □

The quantity $\mathrm{Log}(\varphi_n/\varphi)$ plays a central role in the proofs as it carries all the stochasticity. To apply standard tools such as the central limit theorem for triangular arrays or tools from empirical process theory we need to linearise it, in the sense of writing it as a function applied to $P_n - P$. The next result guarantees its decomposition into a linear part and a 'remainder term'.

In line with the notation introduced at the end of Section 3.1, we write $\xi_n = o_{Pr}(r_n)$ to denote that $\xi_n r_n^{-1}$ vanishes as $n \to \infty$ in Pr-probability and $\xi_n =^{Pr} \xi_n'$ denotes equality in sets of Pr-probability approaching 1 as $n \to \infty$.

3.4.2 Theorem. *Let ξ_k, $k = 1, \ldots, n$, be independent and identically distributed real-valued random variables with characteristic function φ such that $\inf_{u\in\mathbb{R}} |\varphi(u)| > 0$. Let φ_n be the empirical version of it and take $h_n \to 0$ such that $\log(h_n^{-1})/n^{1/(1+2\delta)} \to 0$ for some $\delta > 0$. If $\mathbb{E}[\log^\beta(\max\{|\xi_1|, e\})] < \infty$ for some $\beta > 1$ the quantity $\mathrm{Log}(\varphi_n(u)/\varphi(u))$, $u \in [-h_n^{-1}, h_n^{-1}]$, is well-defined and finite in sets of Pr-probability approaching one as $n \to \infty$. Furthermore, in these sets*

$$\mathrm{Log}\,\frac{\varphi_n(u)}{\varphi(u)} =^{Pr} (\varphi_n(u) - \varphi(u))\, \varphi^{-1}(u) + R_n(u), \qquad u \in [-h_n^{-1}, h_n^{-1}], \tag{3.4.4}$$

where the remainder term is given by

$$R_n(u) = R(z)|_{z=\varphi_n(u)/\varphi(u)} \qquad \textit{with} \qquad R(z) = \int_1^z \frac{s - z}{s^2} ds$$

and it is such that $\sup_{|u|\leqslant h_n^{-1}}|R_n(u)| = O_{Pr}(n^{-1}(\log h_n^{-1})^{1+2\bar{\delta}})$ *for any* $\bar{\delta} \in (0, \delta]$.

Proof. From the construction of the distinguished logarithm (see Theorem 7.6.2 in Chung [2001]) we know that in the closed disk $|z-1| \leq 1/2$ it coincides with the principal branch of the complex logarithm. Thus, there it satisfies that

$$\operatorname{Log} z = (z-1) + R(z),$$

where $R(z)$, the integral form of the remainder, is as in the statement of the theorem, so

$$|R(z)| \leq \sup_{s\in[1,z]} |s|^{-2} \int_1^z |z-s|ds \leq \max\{1, |z|^{-2}\}|z-1|^2. \tag{3.4.5}$$

Taking $z = \varphi_n(u)/\varphi(u)$, we see that the left hand side of (3.4.4) is well-defined in sets of Pr-probability approaching one as $n \to \infty$ and (3.4.4) holds in them if we show that in such type of sets $|\varphi_n(u)/\varphi(u) - 1| \leq 1/2$ whenever $u \in [-h_n^{-1}, h_n^{-1}]$. In view of the finite logarithmic moment assumption, the asymptotics of h_n and Theorem 3.4.1, we have that for any $\delta > 0$

$$\sup_{|u|\leqslant h_n^{-1}} |\varphi_n(u) - \varphi(u)| = O_{\Pr}\big(n^{-1/2}\log^{1/2+\delta}(h_n^{-1})\big) = o_{\Pr}(1). \tag{3.4.6}$$

The conclusion then follows because, due to the strictly positive lower bounded on $|\varphi|$,

$$\left|\frac{\varphi_n(u)}{\varphi(u)} - 1\right| \mathbb{1}_{|u|\leq h_n^{-1}} \leq \sup_{|u|\leqslant h_n^{-1}} |\varphi_n(u) - \varphi(u)| \Big(\inf_{|u|\leqslant h_n^{-1}} |\varphi(u)|\Big)^{-1} = o_{\Pr}(1). \tag{3.4.7}$$

To show the second claim of the theorem we first note that evaluating (3.4.5) at $z = \varphi_n(u)/\varphi(u)$,

$$|R_n(u)| \leq \max\left\{1, \left|\frac{\varphi(u)}{\varphi_n(u)}\right|^2\right\} \left|\frac{\varphi_n(u)}{\varphi(u)} - 1\right|^2. \tag{3.4.8}$$

In view of (3.4.6), the second term on the right hand side is of the required order. We claim that

$$\inf_{|u|\leqslant h_n^{-1}} |\varphi_n(u)| > \kappa \quad \text{ for every } \kappa \in \Big(0, \inf_{u\in\mathbb{R}} |\varphi(u)|\Big) \tag{3.4.9}$$

in the same Pr-probability sets where (3.4.6) holds. Indeed, by the strictly positive lower

bound on $|\varphi|$ we have that for any $\epsilon > 0$

$$\begin{aligned}\Pr\Big(\sup_{|u|\leqslant h_n^{-1}}|\varphi_n(u)-\varphi(u)|\leq\epsilon\Big) &\leq \Pr\Big(\sup_{|u|\leqslant h_n^{-1}}\Big||\varphi_n(u)|-|\varphi(u)|\Big|\leq\epsilon\Big)\\ &\leq \Pr\Big(|\varphi_n(u)|\geq|\varphi(u)|-\epsilon \text{ for all } u\in[-h_n^{-1},h_n^{-1}]\Big)\\ &\leq \Pr\Big(|\varphi_n(u)|\geq\inf_{u\in\mathbb{R}}|\varphi(u)|-\epsilon \text{ for all } u\in[-h_n^{-1},h_n^{-1}]\Big)\\ &= \Pr\Big(\inf_{|u|\leqslant h_n^{-1}}|\varphi_n(u)|\geq\inf_{u\in\mathbb{R}}|\varphi(u)|-\epsilon\Big)\leq 1,\end{aligned}$$

and the conclusion follows by taking $\epsilon < \inf_{u\in\mathbb{R}}|\varphi(u)|$ and noting that the left hand side converges to 1 as $n\to\infty$ by (3.4.6).

□

The following two lemmas are relatively long to prove or will be repeatedly used after and therefore we include them to improve the flow of other proofs.

3.4.3 Lemma. *Let φ and $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ be as in* (3.1.1) *and* (2.6.1)*, respectively, with ν any finite measure on $\mathbb{R}$ satisfying $\nu(\mathbb{R})=\lambda$. Let K be a kernel function satisfying $K,\mathcal{F}K\in L^1$. Then for any $h>0$ fixed*

$$\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\mathcal{F}K_h\big](x)=\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]*K_h(x)\quad \textit{for all } x\in\mathbb{R} \tag{3.4.10}$$

and, consequently, if $K,g\in L^2(\mathbb{R})$ we have that

$$\mathcal{F}^{-1}\big[\mathcal{F}g\,\varphi^{-1}(-\cdot)\mathcal{F}K_h\big](x)=g*K_h*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\quad \textit{for all } x\in\mathbb{R}\,. \tag{3.4.11}$$

Proof. By the expression for φ given in (3.1.1) we have $\varphi^{-1}(-u) = \exp\left(\Delta(iu\gamma+\lambda-\mathcal{F}\bar{\nu}(u))\right)$ for any $u\in\mathbb{R}$. Then, noting that $\mathcal{F}\delta_{\Delta\gamma}(u)=\exp(iu\Delta\gamma)$ and writing the exponential of $-\Delta\mathcal{F}\bar{\nu}(u)$ as an infinite series, we have for any $u\in\mathbb{R}$

$$\varphi^{-1}(-u)=e^{\lambda\Delta}\mathcal{F}\delta_{\Delta\gamma}(u)\sum_{k=0}^{\infty}(\mathcal{F}\bar{\nu}(u))^k\frac{(-\Delta)^k}{k!}=e^{\lambda\Delta}\mathcal{F}\delta_{\Delta\gamma}(u)\sum_{k=0}^{\infty}\mathcal{F}\left[\bar{\nu}^{*k}\right](u)\frac{(-\Delta)^k}{k!}, \tag{3.4.12}$$

where in the last equality we use that if μ_1 and μ_2 are finite measures on $\mathbb{R}$ so is their convolution and $\mathcal{F}\mu_1(u)\mathcal{F}\mu_2(u)=\mathcal{F}[\mu_1*\mu_2](u)$ for every $u\in\mathbb{R}$. These properties of convolution theory and Fourier analysis are repeatedly used in what follows and at the end of the proof we use that the same identity holds for $L^2(\mathbb{R})$ functions. To introduce the infinite sum into the Fourier transform we use that if μ_m, $m=1,2,\ldots$, and μ are finite measures on $\mathbb{R}$ such that $\|\mu_m-\mu\|_{TV}\to 0$ then $\mathcal{F}\mu_m\to\mathcal{F}\mu$ pointwise. Taking

$$\mu_m:=\sum_{k=0}^{m}\bar{\nu}^{*k}\frac{(-\Delta)^k}{k!}\quad\text{and}\quad\mu:=\sum_{k=0}^{\infty}\bar{\nu}^{*k}\frac{(-\Delta)^k}{k!}$$

we see the condition is readily satisfied because

$$\left\| \sum_{k=m+1}^{\infty} \bar{\nu}^{*k} \frac{(-\Delta)^k}{k!} \right\|_{TV} \leq \sum_{k=m+1}^{\infty} \frac{(\lambda\Delta)^k}{k!} \to 0 \qquad \text{as } m \to \infty$$

because it is the tail of a convergent series. Therefore

$$\begin{aligned} \varphi^{-1}(-u)\mathcal{F}K_h(u) &= e^{\lambda\Delta}\mathcal{F}\delta_{\Delta\gamma}(u)\mathcal{F}\left[\sum_{k=0}^{\infty} \bar{\nu}^{*k} \frac{(-\Delta)^k}{k!}\right](u)\mathcal{F}K_h(u) \\ &= \mathcal{F}\left[\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right] * K_h\right](u), \end{aligned}$$

and the first display in the lemma follows by taking the Fourier inverse on both sides, noting that $\varphi^{-1}(-\cdot)\mathcal{F}K_h, \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right] * K_h \in L^1(\mathbb{R})$ and that the latter is continuous because $K, \mathcal{F}K \in L^1(\mathbb{R})$. The second display is then justified by the remark after (3.4.12) and because (3.4.10) is in $L^2(\mathbb{R})$ if K is by Minkowski's inequality for convolutions.

□

3.4.4 Lemma. *Let φ, $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ and P be given by* (3.1.1), (2.6.1) *and* (2.6.2), *where $\gamma > 0$ and ν is a finite measure satisfying $\nu(\mathbb{R}) = \lambda$. Let K be a symmetric kernel function satisfying $K \in L^1(\mathbb{R})$ and $\operatorname{supp}\mathcal{F}K \subseteq [-1,1]$, and let the bandwidth $h_n \to 0$. Let $g_n^{k)} \in L^2(\mathbb{R})$, $k = 1, 2$, be functions satisfying $\sup_n \sup_{x\in\mathbb{R}} |g_n^{k)}(x)| < \infty$ and such that $g^{k)} := \lim_{n\to\infty} g_n^{k)} * K_{h_n}$ exists at every point. Then $g^{k)}$ is finite everywhere,*

$$\begin{aligned} &\lim_{n\to\infty} \int_{\mathbb{R}} g_n^{1)} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\; g_n^{2)} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\, P(dx) \\ &= \int_{\mathbb{R}} g^{1)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\; g^{2)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\, P(dx) \end{aligned} \tag{3.4.13}$$

exists and so does

$$\lim_{n\to\infty} \int_{\mathbb{R}} g_n^{k)} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\, P(dx) = \lim_{n\to\infty} g_n^{k)} * K_h(0) =: g^{k)}(0) \tag{3.4.14}$$

for any $k = 1, 2$. Furthermore, if ν satisfies Assumption 1a for some $\varepsilon > 0$ and, for each $k = 1, 2$, $\tilde{g}^{k)}$ is a bounded function agreeing with $g^{k)}$ everywhere up to a zero Lebesgue-measure set disjoint from $\varepsilon \times \mathbb{Z}$ then

$$\begin{aligned} &\int_{\mathbb{R}} \tilde{g}^{1)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\; \tilde{g}^{2)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\, P(dx) \\ &= \int_{\mathbb{R}} g^{1)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\; g^{2)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\, P(dx). \end{aligned} \tag{3.4.15}$$

Proof. Let $k = 1, 2$ be fixed throughout. Note that $\|K_h\|_{TV} = \|K\|_{TV} < \infty$ for all $h > 0$, $\sup_n \sup_{x\in\mathbb{R}} |g_n^{k)}(x)| < \infty$ by assumption, and $\|\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]\|_{TV} < \infty$. Then, in view of

Minkowski's inequality for integrals, $\sup_n \sup_{x\in\mathbb{R}} |g_n^{k)} * K_{h_n}(x)| < \infty$ and therefore $g^{k)}$ is finite everywhere and $\sup_n \sup_{x\in\mathbb{R}} |g_n^{k)} * K_{h_n} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x)| < \infty$. The existence of all the displays therefore follows and so does equality (3.4.13) using dominated convergence.

Note that $\|\mathcal{F}K_h\|_2 \lesssim h^{-1/2}\|K\|_1 < \infty$, which in particular implies that $K \in L^2(\mathbb{R})$ used below. Additionally, $\mathcal{F}g \in L^2(\mathbb{R})$ and $\varphi^{-1} \in L^\infty(\mathbb{R})$ so, by Hölder's inequality, $\mathcal{F}g\,\varphi^{-1}(-\cdot)\mathcal{F}K_h \in L^1(\mathbb{R})$ .Then, in view of (3.4.11), the symmetry of K and the finiteness of P, the left hand side in (3.4.14) equals the limit as $n \to \infty$ of

$$\frac{1}{2\pi}\int_{\mathbb{R}} \mathcal{F}g_n^{k)}(-u)\,\mathcal{F}K_{h_n}(u)\,\varphi^{-1}(u)\,\mathcal{F}P(u)\,du = \int_{\mathbb{R}} g_n^{k)}(x)\,K_{h_h}(x)\,dx = g_n^{k)} * K_{h_n}\,(0),$$

where the first equality follows by the symmetry of K and Plancherel's formula due to $g, K \in L^2(\mathbb{R})$. The limit of this quantity is $g^{k)}(0)$ by definition and the first part of the lemma is then justified.

To show the last claim we note that because $\tilde{g}^{k)}$ is bounded we can apply Fubini's theorem to write the left hand side of (3.4.15) as

$$\int_{\mathbb{R}^3} \tilde{g}^{1)}(x-y_1)\,\tilde{g}^{2)}(x-y_2)\,P(dx)\,\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(dy_1)\,\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(dy_2).$$

In view of Lemma 27.1 in Sato [1999] the product of the three measures gives a finite measure on $\mathbb{R}^3$ comprising a (possibly null) atomic component and a (possibly null) absolutely continuous component. Recall that the former may have atoms only at $(\varepsilon\times\mathbb{Z})^3$. In any of these point measures the functions $\tilde{g}^{1)}$ and $\tilde{g}^{2)}$ are evaluated at a value in $\varepsilon\times\mathbb{Z}$ and therefore there they coincide with $g^{1)}$ and $g^{2)}$ by assumption. In the rest of the values they agree up to a set of zero Lebesgue-measure but in these they are integrated with respect to an absolutely continuous measure and thus the conclusion follows.

□

The following is the central result when dealing with the stochastic terms arising from proving joint convergence of one-dimensional parameters.

3.4.5 Theorem. *Let φ and φ_n be as in Section 3.1, i.e. the characteristic function, and its empirical counterpart, of the Δ-increments of a compound Poisson process with law P, empirical measure P_n, drift $\gamma \in \mathbb{R}$ and finite jump measure ν satisfying $\nu(\mathbb{R}) = \lambda$ and Assumption 2. Let K be a symmetric kernel function satisfying $K \in L^1(\mathbb{R})$ and $\operatorname{supp}\mathcal{F}K \subseteq [-1,1]$, and let the bandwidth $h_n \to 0$ be such that $\log(h_n^{-1})/n^{1/(1+2\delta)} \to 0$ for some $\delta > 0$. Finally, let $g_n \in L^2(\mathbb{R})$ be a function satisfying*

$$\int_{-h_n^{-1}}^{h_n^{-1}} \big|\mathcal{F}g_n(u)\big|du = o\big(n^{1/2}(\log h_n^{-1})^{-(1+2\delta)}\big),\quad \sup_n \sup_{x\in\mathbb{R}} |g_n(x)| < \infty \quad \textit{and} \quad g := \lim_{n\to\infty} g_n * K_{h_n} \tag{3.4.16}$$

exists at every point. Then we have that as $n \to \infty$

$$\sqrt{n}\frac{1}{\Delta}\int_{\mathbb{R}} g_n(x)\mathcal{F}^{-1}\left[\mathrm{Log}\,\frac{\varphi_n}{\varphi}\mathcal{F}K_h\right](x)dx = {}^{Pr}\sqrt{n}\frac{1}{\Delta}\int_{\mathbb{R}} g_n * K_h * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)(P_n - P)(dx)$$
$$+\sqrt{n}\frac{1}{2\pi\Delta}\int_{\mathbb{R}} \mathcal{F}g_n(-u)R_n(u)\mathcal{F}K_h(u)du \quad (3.4.17)$$
$$\to^d N(0,\sigma_g^2), \quad (3.4.18)$$

where R_n *is as in Theorem 3.4.2 and* σ_g^2 *is finite and satisfies*

$$\Delta^2\sigma_g^2 = \int_{\mathbb{R}} \left(g * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x) - g(0)\right)^2 P(dx) = \int_{\mathbb{R}} \left(g * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\right)^2 P(dx) - g(0)^2.$$

Proof. We first note that the conditions of Theorem 3.4.2 are readily satisfied when $\xi_k = Z_k$ because $\inf_{u\in\mathbb{R}} |\varphi(u)| \geq \exp(-2\lambda\Delta) > 0$, $\mathrm{supp}(\mathcal{F}K_h) \subseteq [-h^{-1}, h^{-1}]$ and we have assumed the same decay in h_n as well as Assumption 2. Then, using the compact support of $\mathcal{F}K_h$, the quantity in the Fourier inverse transform of

$$\frac{1}{\Delta}\int_{\mathbb{R}} g_n(x)\,\mathcal{F}^{-1}\left[\mathrm{Log}\,\frac{\varphi_n}{\varphi}\mathcal{F}K_h\right](x)dx \quad (3.4.19)$$

is in $L^2(\mathbb{R})$ in sets of Pr-probability approaching 1 as $n \to \infty$ and, in these, Plancherel's formula can be used, noting that $g_n \in L^2(\mathbb{R})$, to write the last display as

$$\frac{1}{2\pi\Delta}\int_{\mathbb{R}} \mathcal{F}g_n(-u)\,\mathrm{Log}\,\frac{\varphi_n(u)}{\varphi(u)}\mathcal{F}K_h(u)du = {}^{Pr}\frac{1}{2\pi\Delta}\int_{\mathbb{R}} \mathcal{F}g_n(-u)\left(\varphi_n(u) - \varphi(u)\right)\varphi^{-1}(u)\mathcal{F}K_h(u)du$$
$$+\frac{1}{2\pi\Delta}\int_{\mathbb{R}} \mathcal{F}g_n(-u)R_n(u)\mathcal{F}K_h(u)du. \quad (3.4.20)$$

Due to $|\mathcal{F}K_h| \leq \|K\|_1\,\mathbb{1}_{[-h^{-1},h^{-1}]}$ pointwise, the second summand can be bounded by $\Delta^{-1}\|K\|_1$ times

$$\sup_{u\in[-h_n^{-1},h_n^{-1}]} |R_n(u)| \int_{-h_n^{-1}}^{h_n^{-1}} |\mathcal{F}g_n(u)|\,du = O_{Pr}\Big(n^{-1}(\log h_n^{-1})^{1+2\delta}\int_{-h_n^{-1}}^{h_n^{-1}} |\mathcal{F}g_n(u)|\,du\Big)$$
$$= o_{Pr}\big(n^{-1/2}\big),$$

where the equalities are justified using Theorem 3.4.2 and the first condition in (3.4.16). Hence this summand is negligible in the asymptotic distribution of the left hand side of (3.4.17). The first summand in (3.4.20) is Δ^{-1} times

$$\frac{1}{2\pi}\int_{\mathbb{R}} \mathcal{F}g_n(-u)\varphi^{-1}(u)\mathcal{F}K_h(u)\mathcal{F}\left[P_n - P\right](du) = \int_{\mathbb{R}} \mathcal{F}^{-1}\left[\mathcal{F}g_n\varphi^{-1}(-\cdot)\mathcal{F}K_h\right](x)\left(P_n - P\right)(dx),$$

where the equality follows by the symmetry of K, because $P_n - P$ is a finite measure

and due to $\mathcal{F}g_n\,\varphi^{-1}(-\cdot)\mathcal{F}K_h \in L^1(\mathbb{R})$ by Hölder's inequality noting that $\mathcal{F}g_n \in L^2(\mathbb{R})$, $\varphi^{-1} \in L^\infty(\mathbb{R})$ and $\|\mathcal{F}K_h\|_2 \leq h^{-1/2}\|K\|_1 < \infty$. Note that (3.4.17) is then justified by Lemma 3.4.3. Therefore we point out that the stochastic quantity (3.4.19) is centred after linearisation and the Lindeberg-Feller central limit theorem applies to

$$\frac{1}{\Delta}\int_{\mathbb{R}} g_n * K_h * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x)\,(P_n - P)\,(dx) \tag{3.4.21}$$

if

$$\sup_{x\in\mathbb{R}}\big|g_n * K_h * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x)\big| = o(n^{1/2}) \tag{3.4.22}$$

and if

$$\lim_{n\to\infty}\int_{\mathbb{R}}\big(g_n * K_{h_n} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x) - g_n * K_{h_n}(0)\big)^2\,P(dx) \tag{3.4.23}$$

exists (see, for example, Proposition 2.27 in van der Vaart [1998]). Note that $\|K_h\|_{TV} = \|K\|_{TV} < \infty$ for all $h > 0$ and, by assumption, $\sup_n \sup_{x\in\mathbb{R}} |g_n(x)| < \infty$ and $\|\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\|_{TV} < \infty$. Then, in view of Minkowski's inequality for integrals, $\sup_n \sup_{x\in\mathbb{R}} |g_n * K_{h_n}(x)| < \infty$ and therefore g is finite everywhere and $\sup_n \sup_{x\in\mathbb{R}} |g_n * K_{h_n} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x)| < \infty$. Consequently, (3.4.22) is satisfied and (3.4.23) follows by dominated convergence because of the third assumption in (3.4.16). This also justifies that (3.4.18) holds for the first expression for σ_g^2. The second expression for this quantity follows by expanding the square in (3.4.23) and using (3.4.14).

□

The following auxiliary result is immediate to prove but it will be subsequently used several times. Therefore, and for the sake of clarity, we formulate it in the form of a lemma.

3.4.6 Lemma. *Let $a, b, U, V \in \mathbb{R}$ such that $U > 1$, $V > 0$ and $|a|, |b| < V$. Then, for $k = 0, 1$,*

$$\int_{-U}^{U}\Big|\mathcal{F}\Big[(\cdot)^k\,\mathbb{1}_{[a,b]}\Big]\Big|\,(u)\,du \lesssim U^{2k-1}V^{k+1} + (1-k)\log U.$$

Proof. Note that for all $u \in \mathbb{R}$

$$\Big|\mathcal{F}\Big[(\cdot)^k\,\mathbb{1}_{[a,b]}\Big]\Big|\,(u) \leq \Big\|(\cdot)^k\,\mathbb{1}_{[a,b]}\Big\|_1 \leq V^k\big(|a| + |b|\big) \lesssim V^{k+1}.$$

Therefore, the conclusion for the case $k = 1$ follows immediately.

Due to $U > 1$, when $k = 0$ we split the integral in the statement as the sum of that over $[-U^{-1}, U^{-1}]$ and $[-U, U] \setminus [-U^{-1}, U^{-1}]$. By the last display the former is bounded above by $U^{-1}V$, up to constants independent of U and V. To bound the latter we note

that

$$\int_{[-U,U]\setminus[-U^{-1},U^{-1}]} \left| \frac{e^{iub} - e^{iua}}{iu} \right| du \lesssim \int_{U^{-1}}^{U} u^{-1}\, du \lesssim \log(U).$$

□

The next lemma guarantees the functions g_n we subsequently input into Theorem 3.4.5 satisfy the necessary conditions therein. We use the notation $\equiv$ between two functions to denote pointwise equality.

3.4.7 Lemma. *Suppose K satisfies (3.2.1) and $h_n, \varepsilon_n \to 0$, $H_n \to \infty$ are such that $h_n \sim \exp(-n^{\vartheta_h})$ with $\vartheta_h < 1/4$, $h_n \varepsilon_n^{-1} = o(1)$ and $h_n H_n = o(1)$, hence satisfying the assumptions of Theorem 3.4.5 on K and h_n. Then, for all $c, \varepsilon > 0, j \in \mathbb{Z}$ and $t \in \mathbb{R}$ the functions $f_n^{(\lambda)}$, $f_n^{(\gamma)}$, $f_n^{(q_j)}$ and $f_{t,n}^{(N)}$ defined in Section 3.2 are bounded, belong to $L^2(\mathbb{R})$ and satisfy (3.4.16) for g_n equal to each of them with the constant hidden in the notation small-o of the first condition not depending on t or j when $g_n = f_{t,n}^{(N)}$ or $g_n = g_{(q_j),n}$. Recalling the definitions of $f^{(\lambda)}$, $f^{(q_j)}$ and $f_t^{(N)}$ in Section 3.3 and defining $f^{(\gamma)} := 0$, we have that*

$$f^{(\lambda)} := \mathbb{1}_{\mathbb{R}\setminus\{0\}} \equiv \lim_{n\to\infty} f_n^{(\lambda)} * K_{h_n}, \quad f^{(\gamma)} := 0 \equiv \lim_{n\to\infty} h_n^{-1} f_n^{(\gamma)} * K_{h_n}, \tag{3.4.24}$$

$$f^{(q_j)} := \mathbb{1}_{\{\varepsilon j\}} \equiv \lim_{n\to\infty} f_n^{(q_j)} * K_{h_n},\ j \in \mathbb{Z}, \tag{3.4.25}$$

and

$$l_t := \lim_{n\to\infty} f_{t,n}^{(N)} * K_{h_n} \equiv f_t^{(N)} - \frac{1}{2}\,\mathbb{1}_{\{t\}}\,\mathbb{1}_{\mathbb{R}\setminus\varepsilon\times\mathbb{Z}}(t). \tag{3.4.26}$$

Furthermore, for any finite set $T \subset \mathbb{R}$, any $C_{(\lambda)}, C_{(\gamma)}, C_{(j)}, C_t \in \mathbb{R}$, $j \in \mathbb{Z}\setminus\{0\}$ and $t \in T$, and any $\widetilde{H}_n \to \infty$ define the linear combination

$$f_{t,n} := C_{(\lambda)} f_n^{(\lambda)} + C_{(\gamma)} f_n^{(\gamma)} + \sum_{\substack{|j|\le \widetilde{H}_n/\varepsilon \\ j\neq 0}} C_{(j)} f_n^{(q_j)} + \sum_{t\in T} C_t f_{t,n}^{(N)}. \tag{3.4.27}$$

Then $f_{t,n}$ enjoys the same abovementioned properties as the individual functions if either

(a) finitely many coefficients $C_{(j)}$ are not zero;

(b) or $\sup_{j\in\mathbb{Z}\setminus\{0\}} |C_{(j)}| < \infty$, $\widetilde{H}_n \sim n^{\vartheta_{\widetilde{H}}}$ with $\vartheta_{\widetilde{H}} < 1/2$ and $\vartheta_h < (1 - 2\vartheta_{\widetilde{H}})/4$.

In both cases

$$\lim_{n\to\infty} f_{t,n} * K_{h_n} \equiv C_{(\lambda)} f^{(\lambda)} + C_{(\gamma)} f^{(\gamma)} + \sum_{j\in\mathbb{Z}\setminus\{0\}} C_{(j)} f^{(q_j)} + \sum_{t\in T} C_t l_t.$$

Proof. The conditions of Theorem 3.4.5 on K are trivially satisfied by those assumed here, and those on h_n hold because

$$\log(h_n^{-1})/n^{1/(1+2\delta)} \sim n^{\vartheta_h - 1/(1+2\delta)} = o\big(n^{(-3+2\delta)/(4+8\delta)}\big) = o(1)$$

for any $\delta \in (0, 3/2)$ due to $\vartheta_h < 1/4$.

To check (3.4.16) for each of the individual functions note that for any n fixed, any $j \in \mathbb{Z}$ and $t \in \mathbb{R}$, the functions $f_n^{(\lambda)}$, $f_n^{(\gamma)}$, $f_n^{(q_j)}$ and $f_{t,n}^{(N)}$ are uniformly bounded by $\max\{1, c^{-1}\} < \infty$ and have bounded supports. Therefore each of them is bounded, square integrable and satisfies the second condition in (3.4.16). To check the first condition we first note that $f_n^{(\lambda)} = \mathbb{1}_{[-H_n, H_n]} - \mathbb{1}_{(-\varepsilon_n, \varepsilon_n)}$. Then, by Lemma 3.4.6 with $k = 0$,

$$\int_{-h_n^{-1}}^{h_n^{-1}} \left|\mathcal{F} f_n^{(\lambda)}(u)\right| du \lesssim h_n H_n + h_n \varepsilon_n + \log(h_n^{-1}) = O(\log h_n^{-1}),$$

where in the last equality we have used the asymptotics of h_n, ε_n and H_n. Taking $k = 1$ in Lemma 3.4.6, the same arguments guarantee that

$$\int_{-h_n^{-1}}^{h_n^{-1}} \left|\mathcal{F} f_n^{(\gamma)}(u)\right| du \lesssim 1,$$

and taking $k = 0$ we have that for any $j \in \mathbb{Z}$

$$\int_{-h_n^{-1}}^{h_n^{-1}} \left|\mathcal{F} f_n^{(q_j)}(u)\right| du = \int_{-h_n^{-1}}^{h_n^{-1}} \left|e^{iu\varepsilon j} \mathcal{F}\left[\mathbb{1}_{[-\varepsilon_n, \varepsilon_n]}\right](u)\right| du \lesssim h_n \varepsilon_n + \log(h_n^{-1}) = O(\log h_n^{-1}).$$

The same asymptotics are attained for $f_{t,n}^{(N)}$. To justify this note that this function can be written as $\mathbb{1}_{[a_1, b_1]} + \mathbb{1}_{[a_2, b_2]}$ for some $a_1, a_2, b_1, b_2 \in \mathbb{R}$ satisfying that their absolute values are bounded above by $H_n + \varepsilon_n$. Then, using Lemma 3.4.6 with $k = 0$ and the asymptotics of h_n, ε_n and H_n, we have that for any $t \in \mathbb{R}$

$$\int_{-h_n^{-1}}^{h_n^{-1}} \left|\mathcal{F} f_{t,n}^{(N)}(u)\right| du \lesssim h_n (H_n + \varepsilon_n) + \log(h_n^{-1}) = O(\log h_n^{-1}).$$

Each of these functions then satisfies the first condition in (3.4.16) because $h_n \sim \exp(-n^{\vartheta_h})$ with $\vartheta_h < 1/4$, so

$$(\log h_n^{-1})^{2(1+\delta)} n^{-1/2} \sim n^{2(1+\delta)(\vartheta_h - \frac{1}{4(1+\delta)})} = o(1)$$

for some $\delta \in (0, 1/(4\vartheta_h) - 1)$, for which we also have that $\log h_n^{-1}/n^{1/(1+2\delta)} = o(n^{-\vartheta_h(1+2\theta_h)/(1-2\theta_h)}) \to 0$. The lack of dependence on j or t of the constants hidden in the notation $\lesssim$ is clear by the arguments we have employed.

We now compute the limits in (3.4.24) in reverse order. Notice that for any $x \in \mathbb{R}$

$$h_n^{-1} \cdot \mathbb{1}_{(-h_n,h_n)} * K_{h_n}(x) = \frac{x}{h_n} \int_{x/h_n-1}^{x/h_n+1} K(y)\, dy - \int_{x/h_n-1}^{x/h_n+1} y\, K(y)\, dy.$$

When $x = 0$ this equals $-\int_{[-1,1]} \cdot K$, which is zero by the symmetry of K, and otherwise it is of order $O(h_n^{\eta-1})$ by the decay of $|K|$. Hence the second limit in (3.4.24) is justified for any $c > 0$ because $\eta > 2$ and $h_n \to 0$. To show the first limit we start by noting that for any function g_n such that $\sup_n \sup_{x\in\mathbb{R}} |g_n(x)| < \infty$ we have that as $n \to \infty$

$$\left|\left(g_n - g_n\, \mathbb{1}_{[-H_n,H_n]}\right) * K_{h_n}(x)\right| \leq \sup_n \sup_{x\in\mathbb{R}} |g_n(x)| \int_{\left[\frac{x-H_n}{h_n}, \frac{x+H_n}{h_n}\right]^C} \left|K(y)\right| dy \to 0 \quad \text{for all } x \in \mathbb{R}, \tag{3.4.28}$$

because $K \in L^1(\mathbb{R})$ and $h_n^{-1}, H_n \to \infty$. Therefore the limit of $f_n^{(\lambda)} * K_{h_n}$ as $n \to \infty$ equals that of $\tilde{f}_n^{(\lambda)} * K_{h_n}$ at every point, where $\tilde{f}_n^{(\lambda)} := \mathbb{1}_{\mathbb{R}\setminus(-\varepsilon_n,\varepsilon_n)}$. Additionally, for any $x \in \mathbb{R}$,

$$\mathbb{1}_{\mathbb{R}\setminus(-\varepsilon_n,\varepsilon_n)} * K_{h_n}(x) = 1 - \mathbb{1}_{(-\varepsilon_n,\varepsilon_n)} * K_{h_n}(x)$$

because $\int_{\mathbb{R}} K = 1$, so we only need to compute the limit of the convolution on the right hand side. This coincides with (3.4.25) when $j = 0$ and hence we now compute (3.4.25) for any $j \in \mathbb{Z}$. For any $x \in \mathbb{R}$ we have that

$$\mathbb{1}_{(\varepsilon j-\varepsilon_n,\varepsilon j+\varepsilon_n)} * K_{h_n}(x) = \int_{(x-\varepsilon j-\varepsilon_n)/h_n}^{(x-\varepsilon j+\varepsilon_n)/h_n} K(y)\, dy.$$

By the decay of $|K|$, if $x \neq \varepsilon\, j$ this is of order $O(\varepsilon_n h_n^{\eta-1})$ and if $x = \varepsilon j$ then 1 minus this display equals $\int_{[-\varepsilon_n/h_n,\varepsilon_n/h_n]^c} K$ due to $\int K = 1$. Using that $h_n, \varepsilon_n \to 0$, $\eta > 2$, $h_n \varepsilon_n^{-1} = o(1)$ and $K \in L^1(\mathbb{R})$ we thus conclude that

$$\lim_{n\to\infty} \mathbb{1}_{(\varepsilon j-\varepsilon_n,\varepsilon j+\varepsilon_n)} * K_{h_n} \equiv \mathbb{1}_{\{\varepsilon j\}}, \tag{3.4.29}$$

and consequently the first limit in (3.4.24) and that in (3.4.25) follow.

To show (3.4.26) note that, in view of (3.4.28) because $\sup_n \sup_{x\in\mathbb{R}} |f_{t,n}^{(N)}(x)| < \infty$, the limit of $f_{t,n}^{(N)} * K_{h_n}$ as $n \to \infty$ equals that of $\tilde{f}_{t,n}^{(N)} * K_{h_n}$ at every point, where

$$\tilde{f}_{t,n}^{(N)} := \mathbb{1}_{\mathbb{R}\setminus(-\varepsilon_n,\varepsilon_n)} \times \begin{cases} \mathbb{1}_{(-\infty,t]} & \text{if} \quad |t - J_j| > \varepsilon_n & \text{for all } j \in \mathbb{Z}, \\ \mathbb{1}_{(-\infty,J_j-\varepsilon_n]} & \text{if} \quad J_j - \varepsilon_n \leq t < J_j & \text{for some } j \in \mathbb{Z}, \\ \mathbb{1}_{(-\infty,J_j+\varepsilon_n]} & \text{if} \quad J_j \leq t \leq J_j + \varepsilon_n & \text{for some } j \in \mathbb{Z}\,. \end{cases} \tag{3.4.30}$$

Therefore, when $t = \varepsilon j$ for some $j \in \mathbb{Z}$ we have

$$\tilde{f}_{t,n}^{(N)} = \mathbb{1}_{(-\infty,\varepsilon j-\varepsilon_n]} + \mathbb{1}_{(\varepsilon j-\varepsilon_n,\varepsilon j+\varepsilon_n)} - \mathbb{1}_{(-\varepsilon_n,\varepsilon_n)}\, \mathbb{1}_{[0,\infty)}(j), \tag{3.4.31}$$

almost everywhere and hence the following arguments are not affected. Due to $\varepsilon_n \to 0$ as $n \to \infty$, for any other t fixed we can assume

$$\tilde{f}_{t,n}^{(N)} = \mathbb{1}_{(-\infty,t]} - \mathbb{1}_{(-\varepsilon_n,\varepsilon_n)}\,\mathbb{1}_{(0,\infty)}(t). \tag{3.4.32}$$

Thus, and in view of (3.4.29), we only need to compute the limit of

$$\mathbb{1}_{(-\infty,y]} * K_{h_n}(x) = \int_{(x-y)/h_n}^{\infty} K(z)\,dz$$

when $y = \varepsilon j - \varepsilon_n$ for some $j \in \mathbb{Z}$ and when $y = t$. For the former case we have $(x-y)/h_n = (x-\varepsilon j)/h_n + \varepsilon_n/h_n$. Hence the limit of the last display is $\int_{\mathbb{R}} K = 1$ if $x < \varepsilon j$ and zero otherwise, which can be written as $\mathbb{1}_{(-\infty,\varepsilon j)}$, and the limit of (3.4.31) convolved with K_{h_n} is $\mathbb{1}_{(-\infty,t]}\,\mathbb{1}_{\mathbb{R}\setminus\{0\}}$. When $y = t$ the same arguments apply to $x \neq t$ but when $x = t$ we obtain $\int_{\mathbb{R}^+} K = 1/2$ by the symmetry of K. This gives the limiting function $\mathbb{1}_{(-\infty,t)} + \frac{1}{2}\,\mathbb{1}_{\{t\}}$ and therefore the limit of (3.4.32) convolved with K_{h_n} is $\mathbb{1}_{(-\infty,t]}\,\mathbb{1}_{\mathbb{R}\setminus\{0\}} - \frac{1}{2}\,\mathbb{1}_{\{t\}}$, thus justifying (3.4.26).

The statement regarding the linear combination follows immediately for case (a). To show it for case (b) note that because $\varepsilon_n \to 0$ we can assume $\varepsilon_n < \varepsilon/2$ and therefore for any n fixed

$$\sup_n \sup_{x\in\mathbb{R}} |f_{t,n}(x)| \leq \operatorname{card}(T) + 1 + 1 < \infty$$

and

$$\|f_{t,n}\|_2 \leq \operatorname{card}(T)(H_n + \varepsilon_n)^{1/2} + \widetilde{H}_n \varepsilon_n^{1/2} + h_n^{1/2} < \infty.$$

Notice that assumptions $\vartheta_{\widetilde{H}} < 1/2$ and $\vartheta_h < (1 - 2\vartheta_{\widetilde{H}})/4$ imply

$$\widetilde{H}_n\, n^{-1/2} (\log h_n^{-1})^{2(1+\delta)} \sim n^{\vartheta_{\widetilde{H}} - 1/2 + 2(1+\delta)\vartheta_h} = o(1)$$

for any $\delta \in \big(0, (1 - 2\vartheta_{\widetilde{H}})/(4\vartheta_h) - 1\big)$. This interval has non-empty intersection with that under which the first display of the proof holds and consequently for some of those $\delta > 0$

$$\int_{-h_n^{-1}}^{h_n^{-1}} |\mathcal{F} f_{t,n}(u)|\,du = O(\widetilde{H}_n \log(h_n^{-1})) = o\big(n^{1/2} (\log h_n^{-1})^{-(1+2\delta)}\big)$$

as required. To compute the limit of $f_{t,n} * K_h$ we only need to work with the second sum in $f_{t,n}$ and we recall that for any $j \in \mathbb{Z}\setminus\{0\}$

$$f_n^{(q_j)} * K_{h_n}(x) = \begin{cases} \int_{-\varepsilon_n/h_n}^{\varepsilon_n/h_n} K & \text{if } x = \varepsilon j, \\ \\ O(\varepsilon_n h_n^{\eta-1}) & \text{otherwise,} \end{cases}$$

where $\eta > 2$. Moreover, note that the assumption on the asymptotics of $\widetilde{H}_n$ implies

$\widetilde{H}_n = o(n^{1/2})$ and therefore $\widetilde{H}_n h_n^{\epsilon} = o(1)$ for any $\epsilon > 0$. Due to $\sup_{j\in\mathbb{Z}\setminus\{0\}} |C_{(j)}| < \infty$ and because $1 - \int_{-\varepsilon_n/h_n}^{\varepsilon_n/h_n} K = o(1)$ the result follows in view of

$$\sum_{\substack{|j|\le \widetilde{H}_n/\varepsilon \\ j\neq 0}} C_{(j)} f_n^{(q_j)} * K_{h_n}(x) = \begin{cases} C_{(j)} + o(1) + O\big((\widetilde{H}_n - 1) h_n^{\eta-2}\varepsilon_n h_n\big) = C_{(j)} + o(1) & \text{if } x = \varepsilon j \text{ for} \\ & \text{some } j \in \mathbb{Z}\setminus\{0\}, \\ O\big(\widetilde{H}_n h_n^{\eta-2}\varepsilon_n h_n\big) = o(1) & \text{otherwise.} \end{cases}$$

□

All the results so far are developed to deal with stochastic quantities but, because of the use of a kernel in the estimators, we also need to control some non-stochastic quantities or bias terms. This is the content of the following lemma.

3.4.8 Lemma. *Let K satisfy (3.2.1) and take $h_n \sim \exp(-n^{\vartheta_h})$, $\varepsilon_n \sim \exp(-n^{\vartheta_\varepsilon})$ and $H_n \sim \exp(n^{\vartheta_H})$ such that $0 < \vartheta_\varepsilon < \vartheta_h < \infty$ and $0 < \vartheta_H < \vartheta_h$. Adopt the setting and notation of Section 3.1 and assume the finite measure ν satisfies Assumption 1a. Suppose that either*

(a) Assumption 2 holds for some $\beta > 2$ and $\vartheta_H \ge 1/(2\beta)$;

(b) or $\int_{\mathbb{R}} |x|^\beta \nu(dx) < \infty$ for some $\beta > 1$.

Using the notation introduced in Lemma 3.4.7, define

$$B_n^{(\lambda)} := \frac{1}{\Delta}\int_{\mathbb{R}} f_n^{(\lambda)}(x)\mathcal{F}^{-1}[\mathrm{Log}\,\varphi\mathcal{F}K_{h_n}](x)\,dx - \lambda,$$

$$B_n^{(\gamma)} := h_n^{-1}\frac{1}{\Delta}\int_{\mathbb{R}} f_n^{(\gamma)}(x)\mathcal{F}^{-1}[\mathrm{Log}\,\varphi\mathcal{F}K_{h_n}](x)\,dx - h_n^{-1}\gamma,$$

$$B_n^{(q_j)} := \frac{1}{\Delta}\int_{\mathbb{R}} f_n^{(q_j)}(x)\mathcal{F}^{-1}[\mathrm{Log}\,\varphi\mathcal{F}K_{h_n}](x)\,dx - q_j, \qquad j \in \mathbb{Z}\setminus\{0\},$$

and

$$B_{t,n}^{(N)} := \frac{1}{\Delta}\int_{\mathbb{R}} f_{t,n}^{(N)}(x)\mathcal{F}^{-1}[\mathrm{Log}\,\varphi\mathcal{F}K_{h_n}](x)\,dx - N(t), \qquad t \in \mathbb{R}\,.$$

Then, all these individual quantities are of order $o\big(n^{-1/2}\big)$ and so are $\sup_{j\in\mathbb{Z}\setminus\{0\}} |B_n^{(q_j)}|$ and $\sup_{t\in\mathbb{R}} |B_{t,n}^{(N)}|$. Furthermore, for any finite set $T \subset \mathbb{R}$, any $C_{(\lambda)}, C_{(\gamma)}, C_{(j)}, C_t \in \mathbb{R}$, $j \in \mathbb{Z}\setminus\{0\}$ and $t \in T$, and any $\widetilde{H}_n \to \infty$ define the linear combination

$$B_n := C_{(\lambda)}B_n^{(\lambda)} + \gamma B_n^{(\gamma)} + \sum_{j\in\mathbb{Z}\setminus\{0\}} C_{(j)}\big((B_n^{(q_j)} + q_j)\,\mathbb{1}_{|j|\le\widetilde{H}_n/\varepsilon} - q_j\big) + \sum_{t\in T} C_t B_{t,n}^{(N)}. \quad (3.4.33)$$

Then, $B_n = o\big(n^{-1/2}\big)$ if either

(a) condition (a) above is satisfied and finitely many coefficients $C_{(j)}$ are not zero;*

(b) or condition (b) above is satisfied,* $\sup_{j\in\mathbb{Z}\setminus\{0\}}|C_{(j)}|<\infty$ *and* $\widetilde{H}_n\sim n^{\vartheta_{\widetilde{H}}}$ *for some* $\vartheta_{\widetilde{H}}\geq 1/(2\beta)$.

Proof. In view of the arguments employed to arrive at the right hand side of (3.2.2) we have that for any bounded function $g=g_n$,

$$\frac{1}{\Delta}\int_{\mathbb{R}} g(x)\mathcal{F}^{-1}[\operatorname{Log}\varphi\mathcal{F}K_h](x)dx = -\gamma\int_{\mathbb{R}} g(x)(K_{h_n})'(x)-\lambda\int_{\mathbb{R}} g(x)K_{h_n}(x)\,dx + \sum_{j\in\mathbb{Z}\setminus\{0\}} q_j\int_{\mathbb{R}} g(x)\,K_{h_n}(x-\varepsilon j)\,dx+\int_{\mathbb{R}} g(x)\,\nu_{ac}*K_{h_n}(x)\,dx, \tag{3.4.34}$$

where we have swapped the infinite sum with the integral using that for any $m\in\mathbb{N}$ and any $x\in\mathbb{R}$

$$\left|\sum_{\substack{|j|\leq m\\ j\neq 0}} q_j g(x)K_{h_n}(x-\varepsilon j)\right|\leq \sup_{y\in\mathbb{R}}|g(y)|\sum_{j\in\mathbb{Z}\setminus\{0\}} q_j\,|K_{h_n}|\,(x-\varepsilon j)$$

and that the right hand side is integrable because $\|K_h(\cdot-y)\|_1=\|K\|_1$ for all $y\in\mathbb{R}$ and $\sum_{j\in\mathbb{Z}\setminus\{0\}} q_j\leq\lambda<\infty$.

We start by computing $B_n^{(\gamma)}$. Taking $g=g_n=h_nf_n^{(\gamma)}$ in (3.4.34), using integration by parts in the first summand and, in view of $\nu_{ac},K\in L^1(\mathbb{R})$, Fubini's theorem in the last one, we have that

$$\begin{aligned} cB_n^{(\gamma)} =\,&\gamma\,h_n^{-1}\bigg(\int_{-1}^{1}K(x)\,dx-\big(K(-1)+K(1)\big)-c\bigg)-\lambda\int_{-1}^{1}x\,K(x)\,dx\\ &+\sum_{j\in\mathbb{Z}\setminus\{0\}} q_j\int_{\mathbb{R}}\frac{h_nx+\varepsilon j}{h_n}\,\mathbb{1}_{(-h_n,h_n)}(h_nx+\varepsilon j)K(x)\,dx\\ &+\int_{\mathbb{R}}K(y)\int_{-h_ny+h_n}^{-h_ny-h_n}h_n^{-1}(x+h_ny)\nu_{ac}(x)\,dx\,dy. \end{aligned}$$

By the symmetry and integrability of K, the decay of $|K|$ in (3.2.1), the fact that $c:=2\big(\int_{[0,1]}K-K(1)\big)$ and using that Assumption 1b is satisfied for some $\alpha>0$ we conclude that

$$|B_n^{(\gamma)}|\lesssim h_n^{\eta}\sum_{j\in\mathbb{Z}\setminus\{0\}} q_j\,j^{-\eta}+h_n^{\alpha}=O\big(h_n^{\alpha}\big)=o\big(n^{-1/2}\big)$$

for some $\eta>2$, where the last equality is justified because $h_n\sim\exp(-n^{\vartheta_h})$ for some $\vartheta_h>0$.

To calculate the order of $B_n^{(\lambda)}$ we write $f_n^{(\lambda)}=1-\mathbb{1}_{[-H_n,H_n]^C}-\mathbb{1}_{(-\varepsilon_n,\varepsilon_n)}$ and analyse (3.4.34) when g equals each of these quantities. When $g=1$ we immediately see it equals

0 because K is symmetric, $\int_{\mathbb{R}} K = 1$ and $\int_{\mathbb{R}} \nu = \lambda$. To analyse the second summand let us first consider (3.4.34) for the more general function $g = g_n = \tilde{g}_n \mathbb{1}_{[-H_n,H_n]^C}$, where $\tilde{g}_n$ is a function satisfying $\sup_n \sup_{x\in\mathbb{R}} |\tilde{g}_n(x)| < \infty$. This generalisation is used later on. In this case the right hand side of (3.4.34) equals

$$-\gamma \int_{[-H_n,H_n]^C} \tilde{g}_n(x)(K_{h_n})'(x)\,dx - \lambda \int_{[-H_n,H_n]^C} \tilde{g}_n(x)K_{h_n}(x)\,dx + \int_{\mathbb{R}} K(y) \int_{[-H_n-h_ny,H_n-h_ny]^C} \tilde{g}_n(x+h_ny)\,\nu(dx)\,dy. \tag{3.4.35}$$

The value of the first integral depends on the function $\tilde{g}_n$. When $\tilde{g}_n = 1$ it equals $K_{h_n}(-H_n) - K_{h_n}(H_n)$, which is zero by the symmetry of K. By the decay of $|K|$ in (3.2.1), the absolute value of the second integral is bounded above by

$$\sup_{x\in\mathbb{R}} |\tilde{g}_n(x)| \int_{H_n/h_n}^{\infty} |K|(x)\,dx = O\big((h_n/H_n)^{\eta-1}\big) = o\big(n^{-1/2}\big)$$

for some $\eta > 2$, where the last equality is justified by the exponential decay of h_n and $H_n \to \infty$. The absolute value of the last term in (3.4.35) can be bounded, up to constants independent of n, by

$$\sup_{x\in\mathbb{R}} |\tilde{g}_n(x)| \Big(\int_0^{h_n^{-1}} |K|(x)\,dx \int_{H_n}^{\infty} \nu(dx) + \lambda \int_{h_n^{-1}}^{\infty} |K|(x)\,dx \Big).$$

If condition (a) of the lemma is satisfied $\int_{H_n}^{\infty} \nu(dx) \le (\log H_n)^{-\beta} \int_{H_n}^{\infty} (\log x)^{\beta} \nu(dx) = o\big(n^{-1/2}\big)$. If instead condition (b) holds $\int_{H_n}^{\infty} \nu(dx) \le (H_n)^{-\beta} \int_{H_n}^{\infty} x^{\beta} \nu(dx) = o\big(n^{-1/2}\big)$. Therefore, and using the decay of $|K|$ and h_n, the last display is $o\big(n^{-1/2}\big)$. Consequently we conclude that the second summand in the decomposition of $f_n^{(\lambda)}$ is $o\big(n^{-1/2}\big)$. To analyse the third one we need to compute (3.4.34) when $g_n = f_n^{(q_0)}$. Hence, we first compute $B_n^{(q_j)}$ for any $j \in \mathbb{Z}$ and, only during this calculation and with some abuse of notation, we take $q_0 = -\lambda$ in its expression for notational purposes. In view of (3.4.34) we have that

$$\begin{aligned} B_n^{(q_j)} =& \gamma\, h_n^{-1} \bigg(K\Big(\frac{\varepsilon j - \varepsilon_n}{h_n}\Big) - K\Big(\frac{\varepsilon j + \varepsilon_n}{h_n}\Big) \bigg) + q_j \int_{(-\varepsilon_n/h_n, \varepsilon_n/h_n)^C} K(x)\,dx \\ &+ \sum_{l\in\mathbb{Z}\setminus\{j\}} q_l \int_{(\varepsilon(j-l)-\varepsilon_n)/h_n}^{(\varepsilon(j-l)+\varepsilon_n)/h_n} K(x)\,dx + \int_{\mathbb{R}} K(y) \int_{\varepsilon j - h_n y - \varepsilon_n}^{\varepsilon j - h_n y + \varepsilon_n} \nu_{ac}(x)\,dx\,dy. \end{aligned}$$

Then, by the decay of $|K|$, the fact that $\int_{\mathbb{R}} K = 1$, Assumption 1b and the integrability

of K, the last display can be bounded, up to constants independent of j and n, by

$$\begin{aligned}|B_n^{(q_j)}| &\lesssim \mathbb{1}_{\mathbb{Z}\setminus\{0\}}(j)\, h_n^{\eta-1} + |q_j|\,(h_n/\varepsilon_n)^{\eta-1} + h_n^{\eta-1}\sum_{l\in\mathbb{Z}\setminus\{0\}} q_{j-l}\, l^{1-\eta} + \varepsilon_n^{\alpha} \\ &\lesssim h_n^{\eta-1} + \lambda\,(h_n/\varepsilon_n)^{\eta-1} + \lambda\, h_n^{\eta-1}\sum_{l\in\mathbb{Z}\setminus\{0\}} l^{1-\eta} + \varepsilon_n^{\alpha} = o\big(n^{-1/2}\big) \qquad (3.4.36)\end{aligned}$$

for some $\eta > 2$ and $\alpha > 0$, where the last equality follows because $h_n \sim \exp(-n^{\vartheta_h})$, $\varepsilon_n \sim \exp(-n^{\vartheta_\varepsilon})$ and $\vartheta_\varepsilon <\leq \vartheta_h$. This ends showing that $B_n^{(\lambda)} = o\big(n^{-1/2}\big)$ and, in addition, it shows that $B_n^{(q_j)}$, $j \in \mathbb{Z}\setminus\{0\}$, and its supremum are of the same order.

To analyse $B_{t,n}^{(N)}$ we work directly with its supremum over $t \in \mathbb{R}$. We first remove the truncation in the tails of $f_{t,n}^{(N)}$, just as we have done for $B_n^{(\lambda)}$. With this in mind, take $\tilde{g}_n = \tilde{g}_{t,n} = \tilde{f}_{t,n}^{(N)}$ in (3.4.35), where $\tilde{f}_{t,n}^{(N)}$ is defined in (3.4.30). Note that the upper bounds for the absolute values of the last two summands of (3.4.35) only depend on t through $\sup_{x\in\mathbb{R}} |\tilde{g}_{t,n}(x)|$. Furthermore $\sup_n \sup_{t\in\mathbb{R}} \sup_{x\in\mathbb{R}} |\tilde{g}_{t,n}(x)| < \infty$, so the supremum of these two terms are of order $o\big(n^{-1/2}\big)$ by the same arguments used therein. Therefore, if we show that the supremum over $t \in \mathbb{R}$ of the absolute value of the first summand is $o\big(n^{-1/2}\big)$, we only need to analyse the supremum of (3.4.34) for $g = g_{t,n} = \tilde{f}_{t,n}^{(N)}$. But the former follows easily by noting that the absolute value of the first summand in (3.4.35) when $\tilde{g}_{t,n} = \tilde{f}_{t,n}^{(N)}$ is bounded above by $\gamma h_n^{-1}|K|(H_n/h_n)$ for any $t \in \mathbb{R}$, and this quantity is of order $O\big((h_n/H_n)^{\eta-1}H_n^{-1}\big) = o\big(n^{-1/2}\big)$ for some $\eta > 2$ by the decay of $|K|$, h_n and H_n^{-1}. To analyse (3.4.34) when $g_{t,n} = \tilde{f}_{t,n}^{(N)}$ we note that for n large enough $\tilde{f}_{t,n}^{(N)} = \mathbb{1}_{(-\infty,\min\{u,-\varepsilon_n\}]} + \mathbb{1}_{[\min\{u,\varepsilon_n\},u]}$, where

$$u = u(t,n) := \begin{cases} t & \text{if} \quad |t - J_j| > \varepsilon_n & \text{for all } j \in \mathbb{Z}, \\ J_j - \varepsilon_n & \text{if} \quad J_j - \varepsilon_n \le t < J_j & \text{for some } j \in \mathbb{Z}, \\ J_j + \varepsilon_n & \text{if} \quad J_j \le t \le J_j + \varepsilon_n & \text{for some } j \in \mathbb{Z}\,. \end{cases} \qquad (3.4.37)$$

Inputting this into (3.4.34) gives

$$\begin{aligned}&-\gamma\Big(K_{h_n}\big(\min\{u,-\varepsilon_n\}\big) + K_{h_n}(u) - K_{h_n}\big(\min\{u,\varepsilon_n\}\big)\Big) - \lambda\int_{(-\varepsilon_n,\varepsilon_n)^C} \mathbb{1}_{(-\infty,u]}(x) K_{h_n}(x)\,dx \\ &+\int_{-\infty}^{u} \nu * K_{h_n}(x)\,dx - \sum_{j\in\mathbb{Z}\setminus\{0\}} q_j \int_{\min\{u,-\varepsilon_n\}}^{\min\{u,\varepsilon_n\}} K_{h_n}(x-\varepsilon j)\,dx + \int_{\mathbb{R}} K(y) \int_{\min\{u,-\varepsilon_n\}-h_n y}^{\min\{u,\varepsilon_n\}-h_n y} \nu_{ac}(x)\,dx\,dy.\end{aligned}$$

We now bound the absolute value of each of these terms but the first in the second line by quantities independent of t and of order $o\big(n^{-1/2}\big)$. By the symmetry of K and the decay of $|K|$ in (3.2.1), the absolute value of the first summand is bounded by $\gamma h_n^{-1}|K|(\varepsilon_n/h_n)| = O(h_n^{\eta-1}\varepsilon_n^{-\eta})$ for some $\eta > 2$ and, because $h_n \sim \exp(-n^{\vartheta_h})$, $\varepsilon_n \sim \exp(-n^{\vartheta_\varepsilon})$ and $\vartheta_h > 2 = \vartheta_\varepsilon$, it is of the required order. These arguments and $\sup_n \sup_{t\in\mathbb{R}} \sup_{x\in\mathbb{R}} |\tilde{f}_{t,n}^{(N)}(x)| \le 1$ justify that the absolute value of second summand is

bounded by $2\lambda \int_{\varepsilon_n/h_n}^{\infty} |K| = O\big((h_n/\varepsilon_n)^{\eta-1}\big) = o\big(n^{-1/2}\big)$. Furthermore, for any $t \in \mathbb{R}$ the absolute value of the second summand in the second line is bounded by

$$\sum_{j\in\mathbb{Z}\setminus\{0\}} q_j \int_{(-\varepsilon_n-\varepsilon j)/h_n}^{(\varepsilon_n-\varepsilon j)/h_n} |K|(x)\,dx \lesssim h_n^{\eta-1} \sum_{j\in\mathbb{Z}\setminus\{0\}} q_j j^{1-\eta} = o\big(n^{-1/2}\big).$$

Due to Assumption 1b and the integrability of K, for any $t \in \mathbb{R}$ the absolute value of the last summand is bounded by

$$\int_{\mathbb{R}} K(y) \int_{-\varepsilon_n-h_n y}^{\varepsilon_n-h_n y} \nu_{ac}(x)\,dx\,dy \lesssim \varepsilon_n^{\alpha} = o\big(n^{-1/2}\big),$$

where the last equality follows because $\alpha > 0$ and by the exponential decay of ε_n. We are therefore left with proving that the supremum over $t \in \mathbb{R}$ of the absolute value of

$$\begin{aligned}\int_{-\infty}^{u} \nu * K_{h_n} - N(t) = &\sum_{j\in\mathbb{Z}\setminus\{0\}} q_j \bigg(\int_{-\infty}^{u} K_{h_n}(x-\varepsilon j)\,dx - \mathbb{1}_{(-\infty,t]}(\varepsilon j)\bigg) \\ &+ \int_{\mathbb{R}} K(y) \bigg(\int_{-\infty}^{u-h_n y} \nu_{ac}(x)\,dx - \int_{-\infty}^{t} \nu_{ac}(x)\,dx \bigg) dy \end{aligned} \tag{3.4.38}$$

is of the desired order, where to arrive to the last summand we have used that $\int_{\mathbb{R}} K = 1$. Note that the quantity in brackets in the infinite sum is

$$\int_{-\infty}^{(u-\varepsilon j)/h_n} K(x)\,dx - \mathbb{1}_{(-\infty,t]}(\varepsilon j) = \begin{cases} \int_{(u-\varepsilon j)/h_n}^{\infty} K & \text{if } t \geq \varepsilon j, \\ \\ \int_{-\infty}^{(u-\varepsilon j)/h_n} K & \text{if } t < \varepsilon j. \end{cases}$$

For any $t \in \mathbb{R}$ and in view of (3.4.37) we then have that the absolute value of the first summand in (3.4.38) is bounded by

$$\sum_{j\in\mathbb{Z}\setminus\{0\}} q_j \bigg(\int_{-\infty}^{-\varepsilon_n/h_n} |K|(x)\,dx + \int_{\varepsilon_n/h_n}^{\infty} |K|(x)\,dx \bigg) = O\big((h_n/\varepsilon_n)^{\eta-1}\big) = o\big(n^{-1/2}\big).$$

Using Assumption 1b, the integrability of K, the decay of $|K|$, expression (3.4.37) and the fact that $\int_{\mathbb{R}} \nu_{ac} \leq \lambda$, we conclude that the absolute value of the second summand in (3.4.38) is bounded, up to constants independent of t and n, by

$$\int_{-\varepsilon_n/h_n}^{\varepsilon_n/h_n} |K|(y)|u-h_n y-t|^{\alpha} dy + \lambda \int_{(-\varepsilon_n/h_n,\varepsilon_n/h_n)^C} |K|(y)\,dy \lesssim \varepsilon_n^{\alpha} + (h_n/\varepsilon_n)^{\eta-1} = o\big(n^{-1/2}\big),$$

where the constant hidden in the notation $\lesssim$ is clearly independent of t because $|u-t| \leq \varepsilon_n$ for all $t \in \mathbb{R}$. We have then just shown that $\sup_{t\in\mathbb{R}} |B_{t,n}^{(N)}| = o\big(n^{-1/2}\big)$.

For the last statement of the lemma regarding the linear combination B_n note that

$$\sum_{j\in\mathbb{Z}\setminus\{0\}} C_{(j)}\big((B_n^{(q_j)}+q_j)\,\mathbb{1}_{|j|\leq \widetilde{H}_n/\varepsilon}-q_j\big) = \sum_{\substack{|j|\leq \widetilde{H}_n/\varepsilon\\ j\neq 0}} C_{(j)}B_n^{(q_j)} - \sum_{|j|>\widetilde{H}_n/\varepsilon} C_{(j)}q_j.$$

Due to $\widetilde{H}_n\to\infty$, when finitely many coefficients $C_{(j)}$ are not zero B_n has the required order because the individual quantities featuring in its expression do. For the same reason, under the alternative assumptions we only have to analyse the last display to find the order of B_n. In view of the j-independent upper bound for $|B_n^{(q_j)}|$ in (3.4.36), the absolute value of the first sum is bounded, up to constants independent of n, by

$$\sup_{j\in\mathbb{Z}\setminus\{0\}}|C_{(j)}|\;\widetilde{H}_n\left(h_n^{\eta-1}+(h_n/\varepsilon_n)^{\eta-1}+\varepsilon_n^{\alpha}\right)=o\big(n^{-1/2}\big),$$

where the last equality is justified by the polynomial growth of $\widetilde{H}_n$ and the exponential decay of h_n and ε_n. The absolute value of the second sum can be bounded, up to constants independent of n, by

$$\sup_{j\in\mathbb{Z}\setminus\{0\}}|C_{(j)}|\int_{[-\widetilde{H}_n/\varepsilon,\widetilde{H}_n/\varepsilon]^C}\nu(dx)\lesssim \widetilde{H}_n^{-\beta}\int_{[-\widetilde{H}_n/\varepsilon,\widetilde{H}_n/\varepsilon]^C}|x|^{\beta}\nu(dx),$$

and this has order $o\big(n^{-1/2}\big)$ because $\widetilde{H}_n\sim n^{\vartheta_{\widetilde{H}}}$ for some $\vartheta_{\widetilde{H}}\geq 1/(2\beta)$ by assumption. Therefore, the infinite sum in B_n is of order $o\big(n^{-1/2}\big)$ and so is B under the second set of assumptions because we analysed (3.4.35) when these hold.

□

Finally, the following result guarantees joint convergence of finitely and infinitely one-dimensional parameters and it provides an explicit representation of the covariance of the asymptotic distributions. It follows immediately from the preceding results, and Propositions 3.3.1 and 3.3.3 are corollaries of it.

3.4.9 Theorem. *Suppose K satisfies* (3.2.1) *and take $h_n\sim\exp(-n^{\vartheta_h})$, $\varepsilon_n\sim\exp(-n^{\vartheta_\varepsilon})$ and $H_n\sim\exp(n^{\vartheta_H})$ such that $0<\vartheta_\varepsilon<\vartheta_h<1/4$ and $0<\vartheta_H<\vartheta_h$. Adopt the setting and notation of Sections 3.1 and 3.2, and assume the finite measure ν satisfies Assumption 1. For any finite set $T\subset\mathbb{R}$, any $C_{(\lambda)},C_{(\gamma)},C_{(j)},C_t\in\mathbb{R}$, $j\in\mathbb{Z}\setminus\{0\}$ and $t\in T$, and any $\widetilde{H}_n\to\infty$ we define the linear combination*

$$\widehat{\Upsilon}_n:=\sqrt{n}\Big(C_{(\lambda)}\big(\hat{\lambda}_n-\lambda\big)+C_{(\gamma)}h_n^{-1}\big(\hat{\gamma}_n-\gamma\big)+\sum_{j\in\mathbb{Z}\setminus\{0\}}C_{(j)}\big(\hat{q}_{j,n}\,\mathbb{1}_{|j|\leq\widetilde{H}_n/\varepsilon}-q_j\big)+\sum_{t\in T}C_t\big(\widehat{N}_n(t)-N(t)\big)\Big).$$

Assume that either

(a) finitely many coefficients $C_{(j)}$ are not zero, Assumption 2 holds for some $\beta>2$ and $\vartheta_H\geq 1/(2\beta)$;

(b) or $\sup_{j\in\mathbb{Z}\setminus\{0\}}|C_{(j)}|<\infty$, $\int_{\mathbb{R}}|x|^{\beta}\nu(dx)<\infty$ *for some* $\beta>1$, $\widetilde{H}_n\sim n^{\vartheta_{\widetilde{H}}}$ *for some* $\vartheta_{\widetilde{H}}\in[1/(2\beta),1/2)$ *and* $\vartheta_h<(1-2\vartheta_{\widetilde{H}})/4$.

Then we have that

$$\widehat{\Upsilon}_n\to^{d}N\big(0,\sigma^2\big)\qquad \text{as } n\to\infty \tag{3.4.39}$$

where σ^2 *is finite and satisfies*

$$\Delta^2\sigma^2=\int_{\mathbb{R}}\Big(l*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\Big)^2P(dx),$$

where

$$l:=-C_{(\lambda)}f^{(\lambda)}+\sum_{j\in\mathbb{Z}\setminus\{0\}}C_{(j)}f^{(q_j)}+\sum_{t\in T}C_tf_t^{(N)}.$$

Proof. Notice that the sum of distinguished logarithms equals the distinguished logarithm of the product. Then, in view of the expressions of the estimators $\hat{\lambda}_n$, $\hat{\gamma}_n$, $\hat{q}_{j,n}$ and $\widehat{N}_n(t)$ in Section 3.2 and the expressions for $B_n^{(\lambda)}, B_n^{(\gamma)}, B_n^{(q_j)}$ and $B_{t,n}^{(N)}$ in Lemma (3.4.8) we can write

$$\widehat{\Upsilon}_n=\sqrt{n}\,\frac{1}{\Delta}\int_{\mathbb{R}}f_{t,n}(x)\,\mathcal{F}^{-1}\left[\mathrm{Log}\,\frac{\varphi_n}{\varphi}\mathcal{F}K_h\right](x)\,dx+\sqrt{n}\,B_n, \tag{3.4.40}$$

where $f_{t,n}$ and B_n are defined in (3.4.27) and (3.4.33), respectively. Note that all the assumptions of Lemma 3.4.8 are trivially satisfied by those assumed here as the latter include the former. Therefore the second summand in the last display is negligible as $n\to\infty$. To find the asymptotic distribution of the first summand we use Theorem 3.4.5 applied to $f_{t,n}$ as defined in (3.4.27). We can do this in view of Lemma 3.4.7 because its assumptions are trivially satisfied by those assumed here. Then, using the last conclusion of the lemma, it follows that

$$\widehat{\Upsilon}_n\to^{d}N(0,\tilde{\sigma}^2),$$

where $\tilde{\sigma}^2$ satisfies the same expression as σ^2 after (3.4.39) for

$$\tilde{l}:=C_{(\lambda)}f^{(\lambda)}+\sum_{j\in\mathbb{Z}\setminus\{0\}}C_{(j)}f^{(q_j)}+\sum_{t\in T}C_t\Big(f_t^{(N)}-\frac{1}{2}\,\mathbb{1}_{\{t\}}\,\mathbb{1}_{\mathbb{R}\setminus\varepsilon\times\mathbb{Z}}(t)\Big)$$

in place of l. Since $\tilde{l}(0)=0$ the variance $\tilde{\sigma}^2$ has the same expression as (3.4.13) when $\tilde{g}_t^{1)}=\tilde{g}_t^{2)}=\Delta^{-1}\tilde{l}$. Furthermore, the third summand in $\tilde{l}$ agrees with the third summand in l up to a zero Lebesgue-measure set disjoint from $\varepsilon\times\mathbb{Z}$. Hence, the last claim in Lemma 3.4.4 guarantees $\tilde{\sigma}^2=\sigma^2$. The finiteness of this quantity follows by the conclusions of Theorem 3.4.5 or simply by the boundedness of g_t.

□

3.4.2 Proof of Propositions 3.3.1 and 3.3.3

Proposition 3.3.1 follows immediately from Theorem 3.4.9 by taking $T = \emptyset$, $C_{(j)} = 0$ for all $j \in \mathbb{Z}\backslash\{0\}$ and $C_{(\lambda)} = 1, C_{(\gamma)} = 0$ or $C_{(\lambda)} = 0, C_{(\gamma)} = 1$ when estimating λ or γ, respectively. The conclusion of Proposition 3.3.3 on estimating q_j, $j \in \mathbb{Z}\backslash\{0\}$, follows analogously. To show the conclusion for p_j we write

$$\sqrt{n}\,(\hat{p}_{j,n} - p_j) = \hat{\lambda}_n^{-1}\,\sqrt{n}\,\Big((\hat{q}_{j,n} - q_j) - p_j(\lambda - \hat{\lambda}_n)\Big).$$

In view of Theorem 3.4.9 with $C_{(\lambda)} = -p_j, C_{(\gamma)} = C_l = 0$ for all $l \in \mathbb{Z}\backslash\{0, j\}$, $T = \emptyset$ and $C_{(j)} = 1$, the quantity by which $\hat{\lambda}_n^{-1}$ is multiplied on the right hand side converges to $N\big(0, \lambda^2\sigma_{p_j}^2\big)$. Since $\hat{\lambda}_n$ converges to λ (constant) by Theorem 3.4.9, the conclusion follows by Slutsky's lemma.

3.4.3 Proof of Theorem 3.3.4

3.4.3.1 Estimation of N

Note that the assumptions in the first part of Lemma 3.4.7 are included in those assumed here. Therefore the assumptions of Theorem 3.4.5 are satisfied and, in view of the expression of the estimator $\widehat{N}_n(t)$, $t \in \mathbb{R}$, in Section 3.2, the expression for $B_{t,n}^{(N)}$ in Lemma 3.4.8 and the properties of the distinguished logarithm, we can write

$$\begin{aligned}
\sqrt{n}\,\big(\widehat{N}_n(t) - N(t)\big) &= \sqrt{n}\,\frac{1}{\Delta}\int_{\mathbb{R}} f_{t,n}^{(N)}(x)\,\mathcal{F}^{-1}\left[\mathrm{Log}\,\frac{\varphi_n}{\varphi}\mathcal{F}K_h\right](x)\,dx + \sqrt{n}\,B_{t,n}^{(N)} \\
&=^{Pr} \sqrt{n}\,\frac{1}{\Delta}\int_{\mathbb{R}} f_{t,n}^{(N)} * K_{h_n} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\big(P_n - P\big)(dx) \\
&\quad + \sqrt{n}\,\frac{1}{\Delta}\int_{\mathbb{R}} \mathcal{F}f_{t,n}^{(N)}(-u)R_n(u)\mathcal{F}K_{h_n}(u)du + \sqrt{n}\,B_{t,n}^{(N)},
\end{aligned}$$

where $f_{t,n}^{(N)}$ is defined in (3.2.6) and R_n is as in Theorem 3.4.2. We are showing a central limit theorem under the uniform norm, so we now argue that the supremum over $t \in \mathbb{R}$ of the last line vanishes as $n \to \infty$ in the same sets of Pr-probability approaching 1 in which the last equality holds. By Lemma 3.4.8 we have $\sup_{t\in\mathbb{R}} |B_{t,n}^{(N)}| = o\big(n^{-1/2}\big)$ so the last summand vanishes as $n \to \infty$. Due to $\mathrm{supp}(\mathcal{F}K_h) \subseteq [-h^{-1}, h^{-1}]$, the supremum of the first summand in the last line is bounded by

$$\|K\|_1\,\Delta^{-1}\,\sqrt{n}\sup_{|u|\le h_n^{-1}}|R_n(u)|\sup_{t\in\mathbb{R}}\int_{-h_n^{-1}}^{h_n^{-1}}|\mathcal{F}f_{t,n}^{(N)}(u)|du = o_{Pr}(1),$$

where the equality follows from Theorem 3.4.2 and Lemma 3.4.7. Therefore we have to show the corresponding functional central limit theorem for the linear term

$$\sqrt{n}\,(P_n - P)\psi_{t,n} := \sqrt{n}\int_{\mathbb{R}} \psi_{t,n}(x)(P_n - P)(dx),$$

where

$$\psi_{t,n} := \Delta^{-1} f_{t,n}^{(N)} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right].$$

This type of result follows by showing a central limit theorem (for triangular arrays in this case) for the finite-dimensional distributions and tightness of the limiting process. Conveniently, Theorem 2.11.23 in van der Vaart and Wellner [1996] explicitly gives assumptions under which these follow and using it adds clarity to the proofs. Thus, we recall it adapted to our needs and refer the reader to van der Vaart and Wellner [1996] for the concepts in it such as envelope functions, outer measures, bracketing numbers and entropies, etc.

3.4.10 Theorem. *For each n, let $\Psi_n := \{\psi_{t,n} : t \in \mathbb{R}\}$ be a class of measurable functions indexed by a totally bounded semimetric space $(\mathbb{R}, \rho)$. Given envelope functions Ψ_n assume that*

$$P^*\Psi_n^2 = O(1) \quad \text{and} \quad P^*\Psi_n^2 \mathbb{1}_{\{\Psi_n > \kappa\sqrt{n}\}} \to 0$$

for every $\kappa > 0$, and, for every $\delta_n \downarrow 0$,

$$\sup_{\rho(s,t)<\delta_n} P(\psi_{s,n} - \psi_{t,n})^2 \to 0 \quad \text{and} \quad \int_0^{\delta_n} \sqrt{\log N_{[\,]}(\epsilon\|\Psi_n\|_{P,2}, \Psi_n, L_2(P))}\, d\epsilon \to 0,$$

where $\|\psi\|_{P,2} = \|\psi\|_{L_2(P)} = (\int_{\mathbb{R}} |\psi|^2 P)^{1/2}$. Then the sequence of stochastic processes

$$\left\{\sqrt{n}\,(P_n - P)\,\psi_{t,n} : t \in \mathbb{R}\right\}$$

is asymptotically tight in $l^\infty(\mathbb{R})$ and converges in distribution to a tight Gaussian process provided the sequence of covariance functions $P\psi_{s,n}\psi_{t,n} - P\psi_{s,n}P\psi_{t,n}$ converges pointwise on $\mathbb{R} \times \mathbb{R}$.

We first compute the pointwise limit of the covariance functions. Due to the limiting distribution being a tight (centred) Gaussian process, the former limit uniquely identifies the process. This allows us to identify the so-called intrinsic covariance semimetric of the Gaussian process and, as it is customary, we take ρ equal to this semimetric (see the second half of Chapter 1.5 and Chapter 2.1.2 in van der Vaart and Wellner [1996] for an illustrative discussion on this and other choices). We then check the remaining assumptions of the theorem in order of appearance.

Convergence of the covariance functions. Note that the assumptions here include those of Lemma 3.4.4 and of the first part of Lemma 3.4.7. For any $s,t\in\mathbb{R}$ fixed $P\psi_{s,n}\psi_{t,n}$ has the form of (3.4.13) when $g^{1)}=\Delta^{-1}f_{s,n}^{(N)}$ and $g^{2)}=\Delta^{-1}f_{t,n}^{(N)}$ and therefore we can use the lemmata to conclude that

$$\lim_{n\to\infty}P\psi_{s,n}\psi_{t,n}=\frac{1}{\Delta^2}\int_{\mathbb{R}}l_s*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x)\,l_t*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x)\,P(dx)$$

and that this is finite, where l_t is defined in (3.4.26). Furthermore, this function agrees with $f_t^{(N)}$ up to a zero Lebesgue-measure set disjoint from $\varepsilon\times\mathbb{Z}$. Hence, the last claim in Lemma 3.4.4 guarantees the last display equals $\Sigma_{s,t}^N$. Conclusion (3.4.14) and the fact that $l_t(0)=0$ for any $t\in\mathbb{R}$ justify that $P\psi_{t,n}=0$ for all $t\in\mathbb{R}$ and therefore the sequence of covariance functions $P\psi_{s,n}\psi_{t,n}-P\psi_{s,n}P\psi_{t,n}$ converges pointwise to $\Sigma_{s,t}^N$.

Total boundedness of $\mathbb{R}$ under the internal covariance semimetric ρ. In view of the limiting covariance we just computed we take

$$\rho(s,t)=\Delta^{-1}\left(P\Big(f_s^{(N)}*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]-f_t^{(N)}*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\Big)^2\right)^{1/2},\qquad s,t\in\mathbb{R}\,.$$

To show that $\mathbb{R}$ is totally bounded under this semimetric we bound this expression by another semimetric under which $\mathbb{R}$ is totally bounded. Due to $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ being a finite measure and $\sup_{t\in\mathbb{R}}\sup_{x\in\mathbb{R}}|f_t^{(N)}(x)|\leq 1$, Minkowski's inequality for integrals guarantees that

$$\rho(s,t)^2\lesssim\Delta^{-2}P\left|\Big(f_s^{(N)}-f_t^{(N)}\Big)*\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\right|\leq\Delta^{-2}P\left|f_s^{(N)}-f_t^{(N)}\right|*\left|\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\right|,$$

where the last inequality follows from Jordan's decomposition of finite measures and $\big|\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,\big|$ is the positive measure given by the sum of the positive and negative part of $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ in this decomposition. Finally note that

$$\begin{aligned}P\left|f_s^{(N)}-f_t^{(N)}\right|*\left|\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\right|&=\left|f_s^{(N)}-f_t^{(N)}\right|*\left|\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\right|*\bar{P}\,(0)\\&=\bar{\mu}_0\big((\min\{s,t\},\max\{s,t\}]\big),\end{aligned}$$

where $\bar{P}(A)=P(-A)$ and $\bar{\mu}(A)=\mu(-A)$ for any Borel $A\subseteq\mathbb{R}$, $\mu:=\big|\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\big|*\bar{P}$ and $\mu_0:=\bar{\mu}-\bar{\mu}(\{0\})\delta_0$. The conclusion then follows because $\bar{\mu}_0$ is a finite measure on $\mathbb{R}$.

Conditions on the envelope functions of $\Psi_n:=\{\psi_{t,n}:t\in\mathbb{R}\}$. Note that $\sup_n\sup_{t\in\mathbb{R}}\sup_{x\in\mathbb{R}}|f_{t,n}^{(N)}(x)|\leq 1$, $\|K_h\|_1=\|K\|_1<\infty$ and $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ is a finite measure. Using Minkowski's inequality for integrals we have that $\sup_n\sup_{t\in\mathbb{R}}\sup_{x\in\mathbb{R}}|\psi_{t,n}(x)|\leq\Psi$ for some $\Psi\in(0,\infty)$ and we can take $\Psi_n=\Psi$

for all n. The two conditions on the envelope functions then follow immediately.

Control of $P(\psi_{s,n}-\psi_{t,n})^2$. In the following we repeatedly use the fact that if f, g are bounded functions and μ is a finite positive measure then

$$\mu(f+g)^2 := \int_{\mathbb{R}} (f(x)+g(x))^2 \mu(dx) \leq 2\left(\mu f^2 + \mu g^2\right), \tag{3.4.41}$$

and hence to control the left hand side we can control μf^2 and μg^2 separately. Therefore, writing

$$\begin{aligned}\psi_{s,n}-\psi_{t,n} = &\left(\psi_{s,n}-\Delta^{-1}f_s^{(N)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]\right) - \left(\psi_{t,n}-\Delta^{-1}f_t^{(N)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]\right)\\ &+\Delta^{-1}(f_s^{(N)}-f_t^{(N)}) * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right],\end{aligned}$$

to control $P(\psi_{s,n}-\psi_{t,n})^2$ we only need to consider the quantities

$$P\left(\psi_{t,n}-\Delta^{-1}f_t^{(N)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]\right)^2 \qquad \text{and} \qquad \rho(s,t).$$

The behaviour of the second when $\rho(s,t) < \delta_n \downarrow 0$ is trivial so, if we show that the supremum over $t \in \mathbb{R}$ of the first vanishes as $n \to \infty$, then the first quantity in the second display of Theorem 3.4.10 also vanishes in the limit. Note that in view of the expressions for $f_{t,n}^{(N)}$ and $\tilde{f}_{t,n}^{(N)}$ in (3.2.6) and (3.4.30), respectively, we can write

$$\begin{aligned}\psi_{t,n}-\Delta^{-1}f_t^{(N)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right] = &\Delta^{-1}(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]\\ &- \Delta^{-1}\,\mathbb{1}_{[-H_n,H_n]^C}\,\tilde{f}_{t,n}^{(N)} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right].\end{aligned}$$

Then, using (3.4.41) and (3.4.28), together with the finiteness of $\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$ and P, we only need to analyse the supremum over $t \in \mathbb{R}$ of

$$\begin{aligned}&P\Big((\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]\Big)^2\\ &\qquad = P_d\Big((\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_d + (\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_{ac}\Big)^2\\ &\qquad\quad + P_{ac}\Big((\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_d + (\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_{ac}\Big)^2,\end{aligned} \tag{3.4.42}$$

where P_d and Φ_d are purely discrete finite measures and P_{ac} and Φ_{ac} are absolutely continuous with respect to Lebesgue's measure and, by the decomposition of ν in (3.1.2) and Lemma 27.1 in Sato [1999],

$$P_d = \sum_{k=0}^{\infty} \nu_d^{*k}\frac{\Delta^k}{k!}, \quad P_{ac} = P - P_d, \quad \Phi_d = \sum_{k=0}^{\infty} \bar{\nu}_d^{*k}\frac{(-\Delta)^k}{k!} \quad \text{and} \quad \Phi_{ac} = \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right] - \Phi_d.$$

By (3.4.41) we then have to control the four individual terms arising from (3.4.42). In view of Assumption 1a we notice that P_d may have atoms only at $\varepsilon \times \mathbb{Z}$ and hence for the first term we need to analyse $(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_d(\varepsilon j)$ for any $j \in \mathbb{Z}$. For the second and last quantities we analyse $(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_{ac}(x)$ for any $x \in \mathbb{R}$. For the third quantity we have that, by Fubini's theorem and Jensen's inequality,

$$\begin{aligned} P_{ac}\Big((\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_d\Big)^2 &= \int_{\mathbb{R}^2} \int_{\mathbb{R}} \big(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}\big)(x - y_1) \\ &\qquad \times \big(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}\big)(x - y_2) \\ &\qquad \times P_{ac}(dx)\, \Phi_d(dy_1)\, \Phi_{ac}(dy_2) \\ &\leq \Big(\bar{\Phi}_d\big((\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)})^2 * \bar{P}_{ac}\big)^{1/2}\Big)^2, \end{aligned}$$

where as usual $\bar{\Phi}_d(A) = \Phi_d(-A)$ and $\bar{P}_{ac}(A) = P_{ac}(-A)$ for any Borel set $A \subseteq \mathbb{R}$. Since Φ_d may have atoms only at $\varepsilon \times \mathbb{Z}$ by Assumption 1a, to control this term we therefore require to analyse $(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)})^2 * \bar{P}_{ac}(\varepsilon j)$ for all $j \in \mathbb{Z}$. However, noting that $\bar{P}_{ac} = \bar{\nu}_{ac} * \mu_1$ and $\Phi_{ac} = \bar{\nu}_{ac} * \mu_2$ for some finite measures μ_1 and μ_2, to control all the terms in (3.4.42) but the first we only have to analyse $(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)})^k * \bar{\nu}_{ac}(x)$ for all $x \in \mathbb{R}$ and $k = 1, 2$. The rest of the section is then devoted to showing that for some $\eta > 2$ and $\alpha \in (0, 1]$

$$\sup_{t \in \mathbb{R}} \sup_{j \in \mathbb{Z}} \Big| (\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_d(\varepsilon j) \Big| \lesssim \Big(\frac{h_n}{\varepsilon_n}\Big)^{\eta - 1}$$

and

$$\sup_{t \in \mathbb{R}} \sup_{x \in \mathbb{R}} \Big| (\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)})^k * \bar{\nu}_{ac}(x) \Big| \lesssim \varepsilon_n^{\alpha} + h_n^{\alpha},$$

from which it easily follows that the supremum over $t \in \mathbb{R}$ of (3.4.42) vanishes as $n \to \infty$ because $h_n, \varepsilon_n \to 0$ with $h_n = o(\varepsilon_n)$. To bound the first quantity note that using the symmetry of K we have that for any $j \in \mathbb{Z}$ and $t \in \mathbb{R}$

$$(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_d(\varepsilon j) = \sum_{l \in \mathbb{Z}} \bar{\Phi}_d(\{\varepsilon(l - j)\}) \left(\int_{\mathbb{R}} \tilde{f}_{t,n}^{(N)}(\varepsilon l + x) K_{h_n}(x)\, dx - f_t^{(N)}(\varepsilon l) \right).$$

Without loss of generality assume $\varepsilon_n < \varepsilon/2$ and recall the definition of $u(t, n)$ in (3.4.37)

$$u = u(t, n) := \begin{cases} t & \text{if} \quad |t - \varepsilon j| > \varepsilon_n \qquad \text{for all } j \in \mathbb{Z}, \\ \varepsilon j - \varepsilon_n & \text{if} \quad \varepsilon j - \varepsilon_n \leq t < \varepsilon j \quad \text{for some } j \in \mathbb{Z}, \\ \varepsilon j + \varepsilon_n & \text{if} \quad \varepsilon j \leq t \leq \varepsilon j + \varepsilon_n \quad \text{for some } j \in \mathbb{Z}. \end{cases}$$

Using that $\int_{\mathbb{R}} K = 1$ the quantity in brackets on the right hand side of the second to last

display can then be written as

$$\begin{aligned}
&\int_{\mathbb{R}} \left(\mathbb{1}_{(-\infty,u]} - \mathbb{1}_{(-\varepsilon_n,\varepsilon_n)}\,\mathbb{1}_{[0,\infty)}(t)\right)(\varepsilon l + x)K_{h_n}(x)\,dx - \mathbb{1}_{(-\infty,t]}\,\mathbb{1}_{\mathbb{R}\setminus\{0\}}(\varepsilon l) \\
&\quad = \mathbb{1}_{(-\infty,t]}\,\mathbb{1}_{\mathbb{R}\setminus\{0\}}(\varepsilon l)\int_{(u-\varepsilon l)/h_n}^{\infty} K(x)\,dx + \left(1 - \mathbb{1}_{(-\infty,t]}\,\mathbb{1}_{\mathbb{R}\setminus\{0\}}(\varepsilon l)\right)\int_{-\infty}^{(u-\varepsilon l)/h_n} K(x)\,dx \\
&\qquad - \mathbb{1}_{[0,\infty)}(t)\int_{(-\varepsilon_n-\varepsilon l)/h_n}^{(\varepsilon_n-\varepsilon l)/h_n} K(x)\,dx. \qquad (3.4.43)
\end{aligned}$$

If $t < 0$ or $t \geq 0$ we have that $u(t,n) \leq -\varepsilon_n$ or $u(t,n) \geq \varepsilon_n$, respectively, so when $l = 0$ the absolute value of this display is bounded above by

$$\int_{-\infty}^{-\varepsilon_n/h_n} |K|(x)\,dx + \int_{\varepsilon_n/h_n}^{\infty} |K|(x)\,dx \lesssim \left(\frac{h_n}{\varepsilon_n}\right)^{\eta-1}$$

for some $\eta > 2$ using the decay of $|K|$ in (3.2.1). When $l \neq 0$ the absolute value of the third summand in (3.4.43) is also bounded by the last display. The other two summands are also bounded by the last display when $l \neq 0$ because if $\varepsilon l \leq t$ then $u - \varepsilon l \geq \varepsilon_n$ and if $\varepsilon l > t$ then $u - \varepsilon l \leq \varepsilon_n$. Due to the last display not depending on t, j or l and the fact that $\bar{\Phi}_d$ is a finite measure we therefore conclude that

$$\sup_{t\in\mathbb{R}}\sup_{j\in\mathbb{Z}} |(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)}) * \Phi_d(\varepsilon j)| \lesssim \left(\frac{h_n}{\varepsilon_n}\right)^{\eta-1}.$$

To bound the other quantity we note that, using $\int_{\mathbb{R}} K = 1$, the symmetry of K and the positivity of ν_{ac}, we have that for $k = 1, 2$ and any $t, x \in \mathbb{R}$

$$\begin{aligned}
\left|(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)})^k * \bar{\nu}_{ac}(x)\right| &\leq \int_{\mathbb{R}} \left|\int_{\mathbb{R}} \left(\tilde{f}_{t,n}^{(N)}(x + y + h_n z) - f_t^{(N)}(x+y)\right)K(z)dz\right|^k \nu_{ac}(y)dy \\
&\leq \int_{\mathbb{R}}\int_{\mathbb{R}} \left|\tilde{f}_{t,n}^{(N)}(x + y + h_n z) - f_t^{(N)}(x+y)\right| \nu_{ac}(y)dy\,|K|(z)dz,
\end{aligned}$$

where in the last inequality we have used Jensen's inequality when $k = 2$, Fubini's theorem and the fact that $\tilde{f}_{t,n}^{(N)}$ and $f_t^{(N)}$ only take values $0, \pm 1$. Note that, because ν_{ac} is absolutely continuous with respect to Lebesgue's measure, the truncation of $f_t^{(N)}$ at the origin can be ignored and, similarly to above,

$$\begin{aligned}
\tilde{f}_{t,n}^{(N)}(x + y + h_n z) - \mathbb{1}_{(-\infty,t]}(x+y) = &\mathbb{1}_{(-\infty,u-x-h_n z]}(y) - \mathbb{1}_{(-\infty,t-x]}(y) \\
&- \mathbb{1}_{(-\varepsilon_n-x-h_n z,\varepsilon_n-x-h_n z)}(y)\,\mathbb{1}_{[0,\infty)}(t).
\end{aligned}$$

Therefore, if Assumption 1b is satisfied for some $\alpha \in (0,1]$, we have that for $k = 1, 2$ and

any $t, x \in \mathbb{R}$

$$\left|(\tilde{f}_{t,n}^{(N)} * K_{h_n} - f_t^{(N)})^k * \bar{\nu}_{ac}(x)\right| \lesssim \int_{\mathbb{R}} |u - t - h_n z|^\alpha |K|(z)dz + \varepsilon_n^\alpha \lesssim \varepsilon_n^\alpha + h_n^\alpha,$$

where in the last inequality we have used that $|u - t| \leq \varepsilon_n$ and that $\cdot^\alpha |K| \in L^1(\mathbb{R})$ for any $\alpha \in (0, 1]$ in view of the decay of $|K|$ assumed in (3.2.1). We conclude that the supremum over $x \in \mathbb{R}$ and $t \in \mathbb{R}$ of the left hand side is also bounded by the right hand side by noting that the constants hidden in the notation $\lesssim$ are independent of them in view of the independence of the notation $\lesssim$ in Assumption 1b.

Checking the bracketing entropy condition. To check the remaining condition of Theorem 3.4.10 we first recall that $\Psi_n = \Psi$ is independent of n. Second, we claim that the classes Ψ_n are all contained in a single ball in the space of bounded variation functions. Assuming this, the bracketing entropy in the theorem is bounded above by the bracketing entropy of this ball and, by Corollary 3.7.51 in Giné and Nickl [2016], the latter is bounded above by $(\epsilon\Psi)^{-1}$. Therefore the bracketing entropy condition follows if we prove the claim.

In view of (3.2.6), the definition of $u(t, n)$ in (3.4.37) that we recalled above and by properties of the convolution, the weak derivative of $\psi_{t,n}$ is given by

$$\Delta^{-1}\left((\delta_{-H_n} - \delta_{-\varepsilon_n} + \delta_{\varepsilon_n} - \delta_{H_n})\mathbb{1}_{(-\infty,u]} - \mathbb{1}_{[-H_n,H_n]\setminus(-\varepsilon_n,\varepsilon_n)}\,\delta_u\right) * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right].$$

Thus, using Minkowski's inequality for integrals and that $\|K_h\|_1 = \|K\|_1$, we have that

$$\|\psi'_{t,n}\|_{TV} \leq 5\Delta^{-1}\|K\|_1\|\mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]\|_{TV} < \infty$$

and the claim follows by noting that the upper bound does not depend on t or n.

□

3.4.3.2 Estimation of F

In view of the expressions for $\widehat{F}_n(t)$ and $F(t)$ in Section 3.2, for any $t \in \mathbb{R}$ we can write

$$\sqrt{n}\left(\widehat{F}_n(t) - F(t)\right) = \hat{\lambda}_n^{-1}\sqrt{n}\,\widehat{G}_n(t),$$

where

$$\widehat{G}_n(t) := \widehat{N}_n(t) - F(t)\hat{\lambda}_n = \left(\widehat{N}_n(t) - N(t)\right) + F(t)\left(\lambda - \hat{\lambda}_n\right).$$

Therefore, if we show that $\sqrt{n}\,\widehat{G}_n$ converges in distribution in $\ell^\infty(\mathbb{R})$ to $\lambda\mathbb{G}^F$ and that this limit is tight, the result for F follows by Slutsky's lemma (cf. Example 1.4.7 in van der Vaart and Wellner [1996]) because $\hat{\lambda}_n \to \lambda$ (constant) in distribution in $\mathbb{R}$ in view of Proposition 3.3.1. As we pointed out in the estimation of N, the assumptions in the first

part of Lemma 3.4.7 are included in those assumed here. Therefore the assumptions of Theorem 3.4.5 are satisfied and, in view of the expressions for $B_{t,n}^{(N)}$ and $B_n^{(\lambda)}$ in Lemma 3.4.8 and the properties of the distinguished logarithm, we can write

$$\begin{aligned}\sqrt{n}\,\widehat{G}_n(t) &= \sqrt{n}\,\frac{1}{\Delta}\int_{\mathbb{R}} f_{t,n}^{(F)}(x)\,\mathcal{F}^{-1}\left[\mathrm{Log}\,\frac{\varphi_n}{\varphi}\mathcal{F}K_h\right](x)\,dx + \sqrt{n}\left(B_{t,n}^{(N)} - F(t)B_n^{(\lambda)}\right)\\ &=^{Pr} \sqrt{n}\,\frac{1}{\Delta}\int_{\mathbb{R}} f_{t,n}^{(F)} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\big(P_n - P\big)(dx)\\ &\quad + \sqrt{n}\,\frac{1}{\Delta}\int_{\mathbb{R}} \mathcal{F}f_{t,n}^{(F)}(-u)R_n(u)\mathcal{F}K_{h_n}(u)du + \sqrt{n}\left(B_{t,n}^{(N)} - F(t)B_n^{(\lambda)}\right),\end{aligned}$$

where $f_{t,n}^{(F)} := f_{t,n}^{(N)} - F(t)f_n^{(\lambda)}$, with $f_{t,n}^{(N)}$ and $f_n^{(\lambda)}$ defined in Section 3.2, and R_n is as in Theorem 3.4.2. Because we are showing a central limit theorem under the uniform norm we now argue that the supremum over $t \in \mathbb{R}$ of the last line vanishes as $n \to \infty$ in the same sets of Pr-probability approaching 1 in which the last equality holds. By Lemma 3.4.8 we have $\sup_{t\in\mathbb{R}} |B_{t,n}^{(N)}| = o\big(n^{-1/2}\big)$ and $|B_n^{(\lambda)}| = o\big(n^{-1/2}\big)$, so the last summand vanishes as $n \to \infty$ in view of $\sup_{t\in\mathbb{R}} F(t) \leq 1$. Due to $\mathrm{supp}(\mathcal{F}K_h) \subseteq [-h^{-1}, h^{-1}]$, the supremum of the first summand in the last line is bounded by

$$\|K\|_1\,\Delta^{-1}\sqrt{n}\sup_{|u|\leq h_n^{-1}}|R_n(u)|\sup_{t\in\mathbb{R}}\int_{-h_n^{-1}}^{h_n^{-1}}\big(|\mathcal{F}f_{t,n}^{(N)}(u)| + F(t)|\mathcal{F}f_n^{(\lambda)}(u)|\big)du = o_{Pr}\big(1\big),$$

where the equality follows from Theorem 3.4.2 and Lemma 3.4.7. Consequently we have to show the corresponding functional central limit theorem for the linear term $\sqrt{n}\,(P_n - P)\widetilde{\psi}_{t,n}$, where

$$\widetilde{\psi}_{t,n} := \Delta^{-1}f_{t,n}^{(F)} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right] = \widetilde{\widetilde{\psi}}_{t,n} - \Delta^{-1}F(t)\,\mathbb{1}_{[-H_n,H_n]} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right]$$

and

$$\widetilde{\widetilde{\psi}}_{t,n} := \psi_{t,n} + \Delta^{-1}F(t)\;\mathbb{1}_{(-\varepsilon_n,\varepsilon_n)} * K_{h_n} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right].$$

Hence we apply Theorem 3.4.10 to $\widetilde{\Psi}_n := \{\widetilde{\psi}_{t,n} : t \in \mathbb{R}\}$ in place of Ψ_n. When dealing with the former set of measurable functions we equip the indexing set $\mathbb{R}$ with the intrinsic covariance semimetric $\tilde{\rho}$ provided by $\lambda\mathbb{G}^F$. We now check the conditions of Theorem 3.4.10 following the same order as in the estimation of N and drawing on conclusions therein.

Convergence of the covariance functions. Note that the assumptions here include those of Lemma 3.4.4 and of the first part of Lemma 3.4.7. From the second representation of

$\widetilde{\psi}_{t,n}$ above we have that for any $s,t\in\mathbb{R}$

$$P\widetilde{\psi}_{s,n}\widetilde{\psi}_{t,n} = P\widetilde{\widetilde{\psi}}_{s,n}\widetilde{\widetilde{\psi}}_{t,n} - P\Big(\big(F(s)\widetilde{\widetilde{\psi}}_{t,n} + F(t)\widetilde{\widetilde{\psi}}_{s,n}\big)\,\mathbb{1}_{[-H_n,H_n]} * K_{h_n} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\Big) + F(s)F(t)P\big(\,\mathbb{1}_{[-H_n,H_n]} * K_{h_n} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,\big)^2.$$

Therefore, each of the summands has the form of (3.4.13) and we can use the lemmata together with (3.4.28) to conclude that

$$\lim_{n\to\infty} P\widetilde{\psi}_{s,n}\widetilde{\psi}_{t,n} = \frac{1}{\Delta^2}\bigg(\int_{\mathbb{R}}\big(l_s + F(s)\,\mathbb{1}_{\{0\}}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x) \times \big(l_t + F(t)\,\mathbb{1}_{\{0\}}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,(x)\,P(dx) - F(s)F(t)\bigg)$$

and that this is finite, where l_t is defined in (3.4.26). Furthermore, this function agrees with $f_t^{(N)}$ up to a zero Lebesgue-measure set disjoint from $\varepsilon\times\mathbb{Z}$. Hence, the last claim in Lemma 3.4.4 guarantees the last display equals $\lambda^2\Sigma^F_{s,t}$. Conclusion (3.4.14) and the fact that $(l_t - F(t)f^{(\lambda)})(0) = 0$ for any $t\in\mathbb{R}$ justify that $P\psi_{t,n} = 0$ for all $t\in\mathbb{R}$ and therefore the sequence of covariance functions $P\psi_{s,n}\psi_{t,n} - P\psi_{s,n}P\psi_{t,n}$ converges pointwise to $\lambda^2\Sigma^F_{s,t}$.

Using the first representation of $\widetilde{\psi}_{t,n}$ and using the same arguments we also have that for any $s,t\in\mathbb{R}$

$$\lim_{n\to\infty} P\widetilde{\psi}_{s,n}\widetilde{\psi}_{t,n} - P\widetilde{\psi}_{s,n}P\widetilde{\psi}_{t,n} = \frac{1}{\Delta^2}\int_{\mathbb{R}}\big(f_s^{(N)} - F(s)f^{(\lambda)}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\;(x) \times \big(f_t^{(N)} - F(t)f^{(\lambda)}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\;(x)\,P(dx),$$

which provides a more useful representation of the limiting covariance in what follows.

Total boundedness of $\mathbb{R}$ *under the internal covariance semimetric* $\tilde{\rho}$. In view of the last expression for the limiting covariance we just computed we take for any $s,t\in\mathbb{R}$

$$\tilde{\rho}(s,t) = \frac{1}{\Delta}\bigg(P\Big(\big(f_s^{(N)} - F(s)f^{(\lambda)}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big] - \big(f_t^{(N)} - F(t)f^{(\lambda)}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\Big)^2\bigg)^{1/2}.$$

To show that $\mathbb{R}$ is totally bounded under this semimetric we bound this expression by another semimetric under which $\mathbb{R}$ is totally bounded. Due to $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ being a finite measure and $\sup_{t\in\mathbb{R}}\sup_{x\in\mathbb{R}}|f_t^{(N)}(x) - F(t)f^{(\lambda)}(x)| \le 2$, Minkowski's inequality for integrals guarantees we can use (3.4.41) and therefore

$$\tilde{\rho}(s,t)^2 \lesssim \rho(s,t)^2 + \Delta^{-2}\big(F(s) - F(t)\big)^2 P\big(f^{(\lambda)} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\big)^2 \lesssim \tilde{\mu}\big((\min\{s,t\},\max\{s,t\}]\big),$$

where ρ is the semimetric used when estimating N and $\tilde{\mu} := \bar{\mu}_0 + \nu$ with $\bar{\mu}_0$ as defined in the corresponding section when estimating N. The conclusion then follows because $\tilde{\mu}$ is

a finite measure on $\mathbb{R}$.

Conditions on the envelope functions $\widetilde{\Psi}_n$ of $\widetilde{\Psi}_n := \{\widetilde{\psi}_{t,n} : t \in \mathbb{R}\}$. Note that $\sup_n \sup_{t\in\mathbb{R}} \sup_{x\in\mathbb{R}} |f_{t,n}^{(N)}(x) - F(t) f_n^{(\lambda)}(x)| \leq 2$, $\|K_h\|_1 = \|K\|_1 < \infty$ and $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ is a finite measure. Using Minkowski's inequality for integrals we have that $\sup_n \sup_{t\in\mathbb{R}} \sup_{x\in\mathbb{R}} |\widetilde{\psi}_{t,n}(x)| \leq \widetilde{\Psi}$ for some $\widetilde{\Psi} \in (0,\infty)$ and we can take $\widetilde{\Psi}_n = \widetilde{\Psi}$ for all n. The two conditions on the envelope functions then follow immediately.

Control of $P(\widetilde{\psi}_{s,n} - \widetilde{\psi}_{t,n})^2$. In view of the first representation of $\widetilde{\psi}_{s,n}$ and the expression for $g_{t,n}$ we write

$$\begin{aligned}\widetilde{\psi}_{s,n} - \widetilde{\psi}_{t,n} = \big(\psi_{s,n} - \Delta^{-1} f_s^{(N)} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,\big) - \big(\psi_{t,n} - \Delta^{-1} f_t^{(N)} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,\big)\\ - \Delta^{-1}\big(F(s) - F(t)\big)\big(f_n^{(\lambda)} * K_{h_n} - f^{(\lambda)}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\\ + \Delta^{-1}\big((f_s^{(N)} - F(s) f^{(\lambda)}) - (f_t^{(N)} - F(t) f^{(\lambda)})\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,.\end{aligned}$$

Noting that $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ is a finite measure and $\|K_h\|_1 = \|K\|_1 < \infty$, Minkowski's inequality for integrals guarantees all the summands are bounded functions. Therefore we can use (3.4.41) to argue that in order to control $P(\widetilde{\psi}_{s,n} - \widetilde{\psi}_{t,n})^2$ we can analyse each of the four individual terms arising from this display separately. According to the conclusions of the corresponding section when estimating N, the supremum over $s,t \in \mathbb{R}$ of the first two vanishes as $n \to \infty$ and hence they also vanish when taking the supremum over $\tilde{\rho}(s,t) < \delta_n$ for any $\delta_n \downarrow 0$. The fourth quantity arising from the last display equals $\tilde{\rho}(s,t)^2$ and therefore its behaviour is trivial. Consequently we are left we analysing the third term, which satisfies

$$\begin{aligned}&\sup_{s,t\in\mathbb{R}} \Delta^{-2}\big(F(s) - F(t)\big)^2 P\Big(\big(f_n^{(\lambda)} * K_{h_n} - f^{(\lambda)}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,\Big)^2\\ &\lesssim P\Big(\big(f_n^{(\lambda)} * K_{h_n} - f^{(\lambda)}\big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\,\Big)^2.\end{aligned}$$

By Lemma 3.4.7 the function $f_n^{(\lambda)} * K_{h_n} - f^{(\lambda)}$ converges to 0 pointwise as $n \to \infty$. By the boundedness of this function, the finiteness of $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ and Minkowski's inequality for integrals we can argue as when proving Lemma 3.4.4 and dominated convergence guarantees the limit of the right hand side equals 0. As required, we conclude that for any $\delta_n \downarrow 0$

$$\sup_{\tilde{\rho}(s,t)<\delta_n} P(\widetilde{\psi}_{s,n} - \widetilde{\psi}_{t,n})^2 \to 0 \qquad \text{as } n \to \infty.$$

Checking the bracketing entropy condition. To check the remaining condition of Theorem 3.4.10 we first recall that $\widetilde{\Psi}_n = \widetilde{\Psi}$ is independent of n. Second, we claim that the

classes $\widetilde{\Psi}_n$ are all contained in a single ball in the space of bounded variation functions. Assuming this, the bracketing entropy in the theorem is bounded above by the bracketing entropy of this ball and, by Corollary 3.7.51 in Giné and Nickl [2016], the latter is bounded above by $(\epsilon\widetilde{\Psi})^{-1}$. Therefore the bracketing entropy condition follows if we prove the claim.

In view of the first expression for $\widetilde{\psi}_{t,n}$ and by properties of the convolution, its weak derivative is given by

$$\widetilde{\psi}'_{t,n} = \psi'_{t,n} - \Delta^{-1}F(t)\big(\delta_{-H_n} - \delta_{-\varepsilon_n} + \delta_{\varepsilon_n} - \delta_{H_n}\big) * K_{h_n} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big],$$

where $\psi'_{t,n}$ was computed in the corresponding section when estimating N. Thus, using Minkowski's inequality for integrals and that $\|K_h\|_1 = \|K\|_1$, we have

$$\begin{aligned}\|\widetilde{\psi}'_{t,n}\|_{TV} &\leq \|\psi'_{t,n}\|_{TV} + 4\Delta^{-1}\|K\|_1\|\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\|_{TV} \\ &\leq 9\Delta^{-1}\|K\|_1\|\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]\|_{TV} < \infty,\end{aligned}$$

where the last inequality follows from the conclusions when estimating N. The claim then follows by noting that the upper bound does not depend on t or n.

□

3.4.4 Proof of Theorem 3.3.5

Prior to dealing with joint convergence of the infinite dimensional vectors in the statement of the theorem we justify that under the assumptions of part (b) and as $n \to \infty$

$$\sqrt{n}\left(\hat{q}_n - q\right) \to^d N\big(0, \sigma_q^2\big) \quad \text{and} \quad \sqrt{n}\left(\hat{p}_n - p\right) \to^d N\big(0, \sigma_p^2\big),$$

where, recalling that $f^{(q)} := \sum_{j\in\mathbb{Z}\setminus\{0\}} f^{(q_j)}$ and $f^{(p)} := \sum_{j\in\mathbb{Z}\setminus\{0\}} f^{(p_j)} = \lambda^{-1}(f^{(q)} - pf^{(\lambda)})$,

$$\sigma_q^2 := \frac{1}{\Delta^2}\int_{\mathbb{R}}\Big(f^{(q)} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\Big)^2 P(dx) \quad \text{and} \quad \sigma_p^2 := \frac{1}{\Delta^2}\int_{\mathbb{R}}\Big(f^{(p)} * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\Big)^2 P(dx).$$

The second set of assumptions of Theorem 3.4.9 are satisfied here and, in view of the expression for $\hat{q}_n$ in Section 3.2, the first result thus follows by taking $T = \emptyset$, $C_{(\lambda)} = C_{(\gamma)} = 0$ and $C_{(j)} = 1$ for all $j \in \mathbb{Z}\setminus\{0\}$ in Theorem 3.4.9. For the second result we write

$$\sqrt{n}\left(\hat{p}_n - p\right) = \hat{\lambda}_n^{-1}\sqrt{n}\left((q_n - q) + p\left(\lambda - \hat{\lambda}_n\right)\right).$$

Taking $T = \emptyset$, $C_{(\lambda)} = -p$, $C_{(\gamma)} = 0$ and $C_{(j)} = 1$ for all $j \in \mathbb{Z}\setminus\{0\}$ in Theorem 3.4.9, the quantity by which $\hat{\lambda}_n^{-1}$ is multiplied on the right hand side converges to $N\big(0, \lambda^2\sigma_p^2\big)$. Since $\hat{\lambda}_n$ converges to λ (constant) by the same theorem, the conclusion follows by Slutsky's lemma.

We now prove parts (a) and (b) together under the respective assumptions. From

Lemma 1.4.8 in van der Vaart and Wellner [1996] we have that joint convergence of the infinite dimensional vectors in parts (a) and (b) follows if we show joint convergence of all their finite dimensional projections. To show the latter let $\delta^{i,j}$ denote the Kronecker delta, i.e. the mapping from $\mathbb{Z}\times\mathbb{Z}$ to $\{0,1\}$ that equals 1 only if $i=j$. Define $\underline{\delta} := \big(\delta^{1,m_\lambda}, \delta^{1,m_\gamma}, \delta^{1,m_q}, \delta^{1,m_p}, \underline{\delta}^{(q)}, \underline{\delta}^{(p)}, \delta^{1,m_N}, \delta^{1,m_F}\big)$, where, for any $j\in\mathbb{Z}\setminus\{0\}$, $\underline{\delta}^{(q)}_j = \delta^{1,m_{q_j}}, \underline{\delta}^{(p)}_j = \delta^{1,m_{p_j}}$ and $m_\lambda, m_\gamma, m_q, m_p, m_{q_j}, m_{p_j}, m_N, m_F \in \{0,1\}$ are such that

$$m_\lambda+m_\gamma+m_q+m_p+M_q+M_p+m_N+m_F \in \mathbb{N}, \quad \text{with} \quad M_q := \sum_{j\in\mathbb{Z}\setminus\{0\}} m_{q_j} \quad \text{and} \quad M_p := \sum_{j\in\mathbb{Z}\setminus\{0\}} m_{p_j}.$$

Then, writing $\cdot$ for the coordinate-wise product of two infinite vectors, we denote joint convergence of a finite dimensional projection as having that as $n\to\infty$

$$\sqrt{n}\,\underline{\delta}\cdot\Big(\hat\lambda_n-\lambda,\, h_n^{-1}(\hat\gamma_n-\gamma),\, \hat q_n - q,\, \hat p_n - p,\, \underline{\hat q}_n - \underline{q},\, \underline{\hat p}_n - \underline{p},\, \widehat N_n - N,\, \widehat F_n - F\Big) \to^{\mathcal{D}^{\times,\delta}} \underline{\delta}\cdot\mathbb{L},$$

where $\to^{\mathcal{D}^{\times,\delta}}$ means convergence in distribution in the corresponding product space, which we denote by

$$\mathbb{D} = \mathbb{D}\big(m_\lambda, m_\gamma, m_q, m_p, (m_{q_j})_{j\in\mathbb{Z}\setminus\{0\}}, (m_{p_j})_{j\in\mathbb{Z}\setminus\{0\}}, m_N, m_F\big).$$

Throughout we fix all these binary parameters. To show the joint convergence displayed above we note that, under the assumptions of each of the two parts of the theorem, marginal convergence of each of the non-zero coordinates holds by Propositions 3.3.1 and 3.3.3, Theorem 3.3.4 and the calculations regarding q and p at the beginning of the proof. Therefore, the sequence given by each non-zero projection is asymptotically tight and asymptotically measurable in the respective space. Then, by Lemmas 1.4.3 and 1.4.4 in van der Vaart and Wellner [1996], the sequence of random variables given by the finite dimensional vector above taking values in $\mathbb{D}$ is asymptotically tight and asymptotically measurable. By Prokhorov's theorem it is relatively compact, i.e. every subsequence has a further weakly convergent subsequence, so to finish the proof it suffices to show that all limits are the same. Denote by $\mathcal{H}$ the linear span of the functions $H : \mathbb{R}^{\mathbb{N}}\times\ell^\infty(\mathbb{R})^2 \to \mathbb{R}$ of the form

$$\begin{aligned} H(L) =& \delta^{1,m_\lambda} h^{(\lambda)}(L_\lambda) + \delta^{1,m_\gamma} h^{(\gamma)}(L_\gamma) + \delta^{1,m_q} h^{(q)}(L_q) + \delta^{1,m_p} h^{(p)}(L_p) \\ &+ \sum_{j\in\mathbb{Z}\setminus\{0\}} \delta^{1,m_{q_j}} h^{(q_j)}(L_{q_j}) + \sum_{j\in\mathbb{Z}\setminus\{0\}} \delta^{1,m_{p_j}} h^{(p_j)}(L_{p_j}) \\ &+ \delta^{1,m_N} \sum_{i=1}^{M_N} h^{(i)}\big(L_N(t_i)\big) + \delta^{1,m_F} \sum_{i=M_N+1}^{M_N+M_F} h^{(i)}\big(L_F(t_i)\big) \end{aligned}$$

for any $M_N, M_F \in \mathbb{N}$, $t_i\in\mathbb{R}$ and $h^{(\cdot)} \in C_b(\mathbb{R})$ fixed throughout, where if a sum is empty it

equals 0 by convention. Then, for any $m_\lambda, m_\gamma, m_q, m_p, (m_{q_j})_{j\in\mathbb{Z}\setminus\{0\}}, (m_{p_j})_{j\in\mathbb{Z}\setminus\{0\}}, m_N, m_F$ fixed, $\mathcal{H} \subset C_b(\mathbb{D})$ is a vector lattice containing the constant functions and separating points of $\mathbb{D}$ (see footnote ♭ of page 25 in van der Vaart and Wellner [1996] for the definition of these terms). We claim that, for any of the parameters here introduced and under the corresponding assumptions depending on whether $m_q + m_p = 0$ or not, as $n \to \infty$

$$\sqrt{n}\begin{pmatrix} \underline{\delta}_{-N,-F}\cdot\begin{pmatrix} \hat{\lambda}_n - \lambda \\ h_n^{-1}(\hat{\gamma}_n - \gamma) \\ \hat{q}_n - q \\ \hat{p}_n - p \\ \underline{\hat{q}}_n - \underline{q} \\ \underline{\hat{p}}_n - \underline{p} \end{pmatrix} \\ \delta^{1,m_N}\big(\underline{\widehat{N}}_n - \underline{N}\big) \\ \delta^{1,m_F}\big(\underline{\widehat{F}}_n - \underline{F}\big) \end{pmatrix} \to \begin{pmatrix} \underline{\delta}_{-N,-F}\cdot\mathbb{L}_{-N,-F} \\ \delta^{1,m_N}\begin{pmatrix} \mathbb{L}_N(t_1) \\ \vdots \\ \mathbb{L}_N(t_{M_N}) \end{pmatrix} \\ \delta^{1,m_F}\begin{pmatrix} \mathbb{L}_F(t_{M_N+1}) \\ \vdots \\ \mathbb{L}_F(t_{M_N+M_F}) \end{pmatrix} \end{pmatrix}$$

in distribution in $\mathbb{R}^m$, $m := m_\lambda + m_\gamma + m_q + m_p + M_p + M_q + \delta^{1,m_N}M_N + \delta^{1,m_F}M_F$, where subscript $\cdot_{-N,-F}$ denotes vector $\cdot$ without its last two coordinates, $\underline{N} = \big(N(t_1),\ldots,N(t_{M_N})\big), \underline{F} = \big(F(t_{M_N+1}),\ldots,F(t_{M_N+M_F})\big)$ and $\underline{\widehat{N}}_n, \underline{\widehat{F}}_n$ are the respective coordinate-wise estimators. Then the continuous mapping theorem together with Lemma 1.3.12 (ii) in van der Vaart and Wellner [1996] justify the joint convergence we are seeking. To show the claim we note that by the Cramér-Wold device it is sufficient to check that any linear combination of the coordinates on the left hand side converges to the same combination of the coordinates on the right hand side. For any $c_{(\lambda)}, c_{(\gamma)}, c_{(q)}, c_{(p)} \in \mathbb{R}$ any row vectors $\underline{c}_{(q)}, \underline{c}_{(q)} : \mathbb{Z}\setminus\{0\} \to \mathbb{R}$ with respective j-th entries $c_{(q_j)}$ and $c_{(p_j)}$, any $\underline{C}_N = (c_1,\ldots,c_{M_N})$ and $\underline{C}_F = (c_{M_N+1},\ldots,c_{M_N+M_F})$, the obvious linear combination of the left hand side arising from these parameters can be written as

$$\sqrt{n}\big(c_{(\lambda)}, c_{(\gamma)}, c_{(q)}, \underline{c}_{(q)}, \underline{C}_N\big)\begin{pmatrix} \delta^{1,m_\lambda}(\hat{\lambda}_n - \lambda) \\ \delta^{1,m_\gamma}(h_n^{-1}(\hat{\gamma}_n - \gamma)) \\ \delta^{1,m_q}(\hat{q}_n - q) \\ \underline{\delta}^{(q)}\cdot(\underline{\hat{q}}_n - \underline{q}) \\ \delta^{1,m_N}(\underline{\widehat{N}}_n - \underline{N}) \end{pmatrix} + \hat{\lambda}_n^{-1}\sqrt{n}\big(-\tilde{c}_\lambda, c_{(p)}, \underline{c}_{(p)}, \underline{C}_F\big)\begin{pmatrix} (\hat{\lambda}_n - \lambda) \\ \delta^{1,m_p}(\hat{q}_n - q) \\ \underline{\delta}^{(p)}\cdot(\underline{\hat{q}}_n - \underline{q}) \\ \delta^{1,m_F}(\underline{\widehat{\underline{N}}}_n - \underline{\underline{N}}) \end{pmatrix},$$

where

$$\tilde{c}_\lambda := \delta^{1,m_p}c_{(p)}p + \sum_{j\in\mathbb{Z}\setminus\{0\}}\delta^{1,m_{p_j}}c_{(p_j)}p_j + \delta^{1,m_F}\sum_{i=M_N+1}^{M_N+M_F}c_iF(t_i),$$

$\underline{\underline{N}} = \big(N(t_{M_N+1}),\ldots,N(t_{M_N+M_F})\big)$ and $\underline{\widehat{N}}_n$ is its coordinate-wise estimator. To justify that the penultimate display converges to the correct linear combination of limiting distributions we first note that by Theorem 3.4.9 the finitely many non-zero coordinates in the column vectors of the second to last display converge jointly to the vector comprising

their respective limits. The conclusion then follows by the continuous mapping theorem and in view of the explicit representation of the variance of the limiting random variable given by Theorem 3.4.9 when $T = \{t_1, \ldots, t_{M_N+M_F}\}$, $C_{(\lambda)} = \delta^{1,m_\lambda} c_{(\lambda)} - \lambda^{-1}\tilde{c}_\lambda$, $C_{(\gamma)} = c_{(\gamma)}$, $C_{(j)} = \delta^{1,m_q} c_{(q)} + \delta^{1,m_{q_j}} c_{(q_j)} + \lambda^{-1}\big(\delta^{1,m_p} c_{(p)} + \delta^{1,m_{p_j}} c_{(p_j)}\big)$, $C_t = \delta^{1,m_N} c_t$ if $t = t_1, \ldots, t_{M_N}$ and $C_t = \lambda^{-1}\delta^{1,m_F} c_t$ if $t = t_{M_N+1}, \ldots, t_{M_N+M_F}$.

3.4.5 Proof of Lemma 3.3.2

This lemma follows immediately from the expressions for $\mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big]$ and P given in (2.6.1) and (2.6.2). In view of these and of the observation after (2.6.3), we have that for any f and g bounded in $\mathbb{R}$

$$
\begin{aligned}
&\int_{\mathbb{R}} \Big(f * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x) \Big) \Big(g * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x) \Big) P(dx) \\
&\quad = e^{\lambda\Delta} \left(f(0) + f * \sum_{k=1}^{\infty} \bar{\nu}^{*k} \frac{(-\Delta)^k}{k!}(0) \right) \left(g(0) + g * \sum_{k=1}^{\infty} \bar{\nu}^{*k} \frac{(-\Delta)^k}{k!}(0) \right) \\
&\qquad + e^{\lambda\Delta} \int_{\mathbb{R}} \left(f(x) + f * \sum_{k=1}^{\infty} \bar{\nu}^{*k} \frac{(-\Delta)^k}{k!}(x) \right) \left(g(x) + g * \sum_{k=1}^{\infty} \bar{\nu}^{*k} \frac{(-\Delta)^k}{k!}(x) \right) \sum_{k=1}^{\infty} \nu^{*k} \frac{\Delta^k}{k!}(dx) \\
&\quad = f(0)\, g(0) + \Delta \left(\int_{\mathbb{R}} f(x) g(x)\, \nu(dx) - f(0)\, g * \bar{\nu}\,(0) - g(0)\, f * \bar{\nu}\,(0) + \lambda\, f(0) g(0) \right) \\
&\qquad + O\big((\lambda\Delta)^2\big).
\end{aligned}
$$

Chapter 4

Applications

The purpose of this chapter is to illustrate the applications, implementation and practical performance of the results from Chapter 3. In Section 4.1 we briefly review existing literature on the practical use of nonparametric estimators of discretely observed compound Poisson processes and other related stochastic processes. In Section 4.2 we discuss numerous applications of our results from Chapter 3 and construct the corresponding statistical procedures. Section 4.3 contains the practical implementation of these and, lastly, in Section 4.4 we illustrate their practical performance through simulations in several settings. In doing so we compare them to the respective estimators and procedures from Buchmann and Grübel [2003].

4.1 Introduction

As mentioned in Section 1.3, limit theorems of the type of those developed in Chapter 3 are of great importance for statistics because several type of procedures can be derived from them. The obvious ones are estimators and confidence regions for the parameters. In addition, goodness-of-fit and two-sample tests for each or all of the parameters can be derived too. As discussed in Section 1.4, prior to Coca [2015] only Buchmann and Grübel [2003] and Buchmann and Grübel [2004] showed similar limit theorems for the strict setting of discretely observed compound Poisson processes —van Es et al. [2007] showed pointwise asymptotic normality of their estimator of the density, so global confidence regions for the jump measure cannot be constructed from it and are not comparable. They also discussed applications of their results, including estimation, confidence regions and goodness-of-fit tests. However, they only implemented those for the mass function; as we clarified in Section 2.6, these are different from their estimator of the distribution function both for finite samples and asymptotically. Furthermore, in the implementation of the former they only considered very few and simple examples. In particular, when illustrating the testing applications they only briefly considered the case of $dF = \delta_1$ corresponding to a Poisson

process.

Surprisingly, we have not found other pieces of work in which this practical performance is analysed further. Therefore, in Section 4.4 we include the first illustration of the estimators of a general distribution function of Buchmann and Grübel [2003] and of the coverage of the confidence regions arising from them, and we compare their performance to that of ours. We do this for a number of different settings depending on the value of λ, which drives how hard the estimation problem is in practice. Consequently, our experiments add value to the existing literature. It would also be interesting to explore other topics such as robustness to the underlying model assumptions and quantile estimation, as well as applying the abovementioned statistical procedures to real-life data. However, all this is beyond the scope of this thesis and may appear in future work. We note in passing that Buchmann and Grübel [2004] carried out a simulation study of how well their estimators estimate the support of the underlying jump distribution. Recall that they constructed estimators of the mass function and each of these stayed within the interval $[0, 1]$. Therefore, this is a natural study for their estimators but not for ours, as we do not apply their monotone transformation.

Some limited work on the illustration of confidence regions can be found in the more general setting of estimating discretely observed Lévy processes. At the end of Section 2.2.2 we mentioned the work of Cont and Tankov [2003], Belomestny and Reiß [2006], Söhl [2014] and Söhl and Trabs [2014]. Recall that they assumed the logarithm of the price of a financial asset is a compound Poisson process with drift and with a diffusion component. Under this assumption, they developed confidence sets for the jump density using prices of financial options and illustrated them through simulations. In a closer spirit to that taken here, although in the high-frequency regime, Nickl et al. [2016] constructed confidence bands for functionals of the (not-necessarily-finite) Lévy measure from their functional central limit theorems and illustrated them through simulations. Regarding testing, Reiß [2013] studied this problem for the Lévy triplet and for the Blumenthal-Getoor index of the Lévy measure under different observation regimes. He focused on understanding the complexity of the problem, which is reflected in the asymptotic separation rates. He constructed tests for all these parameters, although he did not perform any simulations. Therefore, our results in this chapter also contribute towards the more practical side of this area.

4.2 Statistical procedures and their construction

All of the procedures following from the results in Chapter 3 make direct use of the estimators and of the limiting random quantities therein. The implementation of the former follows directly from their expressions and is given in Section 4.3. However, the latter depend on the underlying parameters of the unknown process so, prior to discussing their

implementation, we first need to address how we circumvent this issue. We do this using the idea of 'Studentisation' and in Section 4.2.1 we justify why we take this approach and give details of how we adapt it to our setting. In the remaining sections we show the construction of procedures following from it. These include confidence regions in Section 4.2.2, and goodness-of-fit and two-sample tests in Sections 4.2.3 and 4.2.4, respectively. Lastly, in Section 4.2.5, we discuss approximations to all of these when $\lambda\Delta$ is small, including 'process-free' procedures. Throughout we work with the same notation introduced in Chapter 3.

4.2.1 Approximating the limiting processes

The random limiting quantities in the central limit theorems of Chapter 3 can be split into finite- and infinite-dimensional, depending on the dimension of the metric space in which they take values. In either case, to develop applications from them the interest concentrates on the quantiles of the real-valued nonnegative random variables corresponding to their norms in the respective spaces. We begin by discussing how we proceed with the finite-dimensional case since it motivates the approach we take for the infinite-dimensional case. Furthermore, we focus on the one-dimensional quantities as the extension to other finite-dimensional ones is immediate.

From now on, we largely exclude γ from the discussions due to the null-limit in Proposition 3.3.1, and we also exclude q and $\{q_j : j \in \mathbb{Z}\setminus\{0\}\}$ because their counterparts p and $\{p_j : j \in \mathbb{Z}\setminus\{0\}\}$ are generally of more interest in practice. The construction of statistical procedures for the former follows easily from the same arguments we use to construct them for the latter. Throughout let $\theta \in \{\lambda, p\} \cup \{p_j : j \in \mathbb{Z}\setminus\{0\}\}$. In the corresponding one-dimensional specialisations of Theorem 3.3.5, all the limiting quantities can be written as $N(0,\sigma_\theta^2) =^d \sigma_\theta N(0,1)$, for some $\sigma_\theta > 0$. The simple 'Studentisation' idea is to approximate this general distribution by $\hat{\sigma}_{\theta,n}N(0,1)$, where $\hat{\sigma}_{\theta,n}$ is a consistent estimator of σ_θ. Then, by the continuous mapping theorem, or in particular Slutsky's lemma, this approximation converges in distribution to $N(0,\sigma_\theta^2)$. Denoting the upper $\alpha \in (0,1)$ quantile of the distribution $|N(0,1)|$ by $Q^N(\alpha)$, the continuous mapping theorem again justifies that $\hat{\sigma}_{\theta,n}Q^N(\alpha)$ is a consistent estimator of the upper α quantile of the absolute value of the limiting distribution. The question now is how to find an appropriate estimator $\hat{\sigma}_{\theta,n}$. Recall that, for all $\theta \in \{\lambda, p\} \cup \{p_j : j \in \mathbb{Z}\setminus\{0\}\}$,

$$\sigma_\theta^2 := \frac{1}{\Delta^2}\int_{\mathbb{R}} \left(f^{(\theta)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\right)^2 P(dx),$$

where

$$f^{(\lambda)} := \mathbb{1}_{\mathbb{R}\setminus\{0\}}, \quad f^{(p_j)} := \lambda^{-1}\left(\mathbb{1}_{\{J_j\}} - p_j f^{(\lambda)}\right), \quad f^{(p)} := \sum_{j\in\mathbb{Z}\setminus\{0\}} f^{(p_j)} \quad \text{and} \quad \varphi^{-1} = 1/\mathcal{F}P.$$

Therefore, given the i.i.d. observations $Z_1, \dots, Z_n \sim P$ we used to construct the estimator of θ, and taking h_n, ε_n, H_n and $\tilde{H}_n$ as in Chapter 3, we propose the natural estimator

$$\hat{\sigma}^2_{\theta,n} := \frac{1}{\Delta^2}\frac{1}{n}\sum_{k=1}^{n}\left(f_n^{(\theta)} * \mathcal{F}^{-1}\left[\varphi_n^{-1}(-\cdot)\mathcal{F}K_{h_n}\right](Z_k)\right)^2, \tag{4.2.1}$$

where K has the same properties as in (3.2.1),

$$f_n^{(\lambda)} := \mathbb{1}_{[-H_n,H_n]\setminus(-\varepsilon_n,\varepsilon_n)}, \quad f_n^{(p_j)} := \hat{\lambda}_n^{-1}\left(\mathbb{1}_{(J_j-\varepsilon_n,J_j+\varepsilon_n)} - \hat{p}_{j,n} f_n^{(\lambda)}\right),$$

$$f_n^{(p)} := \sum_{\substack{|j|\le \widetilde{H}_n/\varepsilon \\ j\neq 0}} f_n^{(p_j)} \quad \text{and} \quad \varphi_n^{-1} = \left(\frac{1}{n}\sum_{k=1}^{n} e^{i\cdot Z_k}\right)^{-1}.$$

Using the same arguments used to prove the central limit theorems of Chapter 3 it follows that, as $n \to \infty$, $\hat{\sigma}^2_{\theta,n}$ is well-defined in sets of Pr-probability approaching 1, and

$$\hat{\sigma}^2_{\theta,n} \to^{Pr} \sigma^2_\theta \qquad \text{for all } \theta \in \{\lambda, p\} \cup \{p_j : j \in \mathbb{Z}\setminus\{0\}\}.$$

We thus conclude we can use the idea above to consistently estimate the desired quantiles.

For the infinite-dimensional limiting quantities we use a similar idea and look for a (uniformly) consistent estimator of their covariance functions. In particular, notice that the centred Gaussian process $\mathbb{G}^F$ in Theorem 3.3.4 has a covariance function with a very similar form to $\sigma^2_{(\theta)}$ above. Namely,

$$\Sigma^F_{s,t} := \frac{1}{\Delta^2}\int_{\mathbb{R}}\left(f_s^{(F)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\right)\left(f_t^{(F)} * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\right)P(dx), \quad s,t \in \mathbb{R},$$

where

$$f_t^{(F)} := \lambda^{-1}\big(f_t^{(N)} - F(t) f^{(\lambda)}\big) \quad \text{with} \quad f_t^{(N)} := \mathbb{1}_{(-\infty,t]}\, f^{(\lambda)}. \tag{4.2.2}$$

Then, in analogy to (4.2.1), for all $s,t \in \mathbb{R}$ we can approximate $\Sigma^F_{s,t}$ by

$$\widehat{\Sigma}^F_{s,t,n} := \frac{1}{\Delta^2}\frac{1}{n}\sum_{k=1}^{n}\left(f_{s,n}^{(F)} * \mathcal{F}^{-1}\left[\varphi_n^{-1}(-\cdot)\mathcal{F}K_{h_n}\right](Z_k)\right)\left(f_{t,n}^{(F)} * \mathcal{F}^{-1}\left[\varphi_n^{-1}(-\cdot)\mathcal{F}K_{h_n}\right](Z_k)\right)$$

where

$$f_{t,n}^{(F)} := \hat{\lambda}_n^{-1}\Big(f_{t,n}^{(N)} - \widehat{F}_n(t) f_n^{(\lambda)}\Big) \quad \text{with} \quad f_{t,n}^{(N)} := \mathbb{1}_{(-\infty,t]}\, f_n^{(\lambda)}.$$

This is well-defined in sets of Pr-probability approaching 1 as $n \to \infty$ and, crucially, the uniform convergence of the estimators shown in the proofs of Chapter 3 also holds for this

estimator so we have that

$$\sup_{s,t\in\mathbb{R}} \left|\widehat{\Sigma}^F_{s,t,n} - \Sigma^F_{s,t}\right| \to^{Pr} 0 \qquad \text{as } n \to \infty.$$

We define $\widehat{\mathbb{G}}^F_n$ to be the centred Gaussian process with covariance structure $\widehat{\Sigma}^F_{s,t,n}$. Then, by similar arguments as those used to prove Theorem 3.3.4, $\widehat{\mathbb{G}}^F_n \to^{\mathcal{D}} \mathbb{G}^F$ in Pr-probability as $n \to \infty$ and, by the continuous mapping theorem, the quantiles of $\|\widehat{\mathbb{G}}^F_n\|_\infty$ converge to those of $\|\mathbb{G}^F\|_\infty$. This justifies the following procedures because by the continuous mapping theorem again $\sqrt{n}\|\widehat{F}_n - F\|_\infty \to^d \|\mathbb{G}^F\|_\infty$ in $\mathbb{R}$. Throughout we denote the upper $\alpha \in (0,1)$ quantile of $\|\widehat{\mathbb{G}}^F_n\|_\infty$ by $\widehat{Q}^F_n(\alpha)$, which can be easily approximated by Monte-Carlo techniques by simulating numerous paths of $\widehat{\mathbb{G}}^F_n$ and computing the empirical quantile. In the context of estimating discretely observed compound Poisson processes, Buchmann and Grübel [2004] already suggested to use this approach for their estimators in Buchmann and Grübel [2003], although they did not give further details of it. They justified it because, from the expressions of their covariance functions, which in Section 2.6 we argued are the same as ours, it is clear they continuously depend on the defining parameters of the process so the continuous mapping theorem immediately guarantees that consistent estimators of the parameters can be plugged-in.

A second approach may be taken: Bøgsted and Pitts [2010] used the bootstrap method, which is based on approximating the left hand side in the central limit theorem $\sqrt{n}\left(\widehat{F}_n - F\right) \to^{\mathcal{D}} \mathbb{G}^F$ instead of the right hand side. They justified its use in the convoluted dependence of the covariance functions on the parameters observed in the expression right after (2.1.12), which was never rewritten in the easier-to-implement alternative form we give in Section 2.6 and that we exploit below. The bootstrap method is well-known to be optimal under mild conditions that are satisfied here (cf. Giné and Zinn [1990] and van der Vaart and Wellner [1996]) and we refer the reader to their work for more details. We simply mention that, asymptotically, there should be no difference between the two approaches. However, for finite samples there may be: bootstrap may return more accurate coverage values than our Studentisation approach since it is not based on estimating the asymptotically correct approximation on the right side of the central limit theorem but rather on estimating the finite-sample left side. Nevertheless, it has a disadvantage that stops us from using it: in the same way that, for each sample of P, in our approach we have to simulate hundreds of paths of the approximate limiting Gaussian process $\widehat{\mathbb{G}}^F_n$, bootstrap requires to compute our estimator just as many times. The former can be quickly computed as soon as we compute $\widehat{\Sigma}^F_{s,t,n}$, but the latter requires to implement (a slight variation of) our estimator $\widehat{F}_n$ and that of Buchmann and Grübel [2003] for every simulation. For each sample of P, the average computational time to compute each of these two estimators of F varies from a few seconds to a few minutes. Therefore, computing them hundreds of times and then repeating the operation for hundreds of dif-

ferent samples becomes impractical if using a modest machine like we do. Bøgsted and Pitts [2010] were able to use this method because, just like Buchmann and Grübel [2003] did, they discretised their data and used the discrete estimators following from Panjer recursions which are much faster to compute. As already discussed in detail in Section 2.1 and as it becomes clearer from the simulations in Section 4.4, the drawback of the latter estimators is that the estimation deteriorates heavily away from the origin.

In the following sections we also address joint estimation and testing of all the parameters of the process. With this in mind, let $\mathbb{L}^{F,\lambda}$ be the limiting process of $\sqrt{n}\big(\widehat{F}_n - F + \hat{\lambda}_n - \lambda\big)$. By Theorem 3.3.5, its covariance function is given by

$$\Sigma^{F,\lambda}_{s,t} := \frac{1}{\Delta^2}\int_{\mathbb{R}}\Big(\Big(f^{(F)}_s + f^{(\lambda)}\Big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\Big)\Big(\Big(f^{(F)}_t + f^{(\lambda)}\Big) * \mathcal{F}^{-1}\big[\varphi^{-1}(-\cdot)\big](x)\Big)P(dx),$$

and we recall that the covariance of $\mathbb{B}^N$ from Theorem 3.3.4 is

$$\Sigma^{N}_{s,t} := \frac{1}{\Delta^2}\int_{\mathbb{R}}\Big(f^{(N)}_s * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)\Big(f^{(N)}_t * \mathcal{F}^{-1}\left[\varphi^{-1}(-\cdot)\right](x)\Big)P(dx).$$

In line with the above, these can be consistently and uniformly estimated by

$$\begin{aligned}\widehat{\Sigma}^{F,\lambda}_{s,t,n} := \frac{1}{\Delta^2}\frac{1}{n}\sum_{k=1}^{n}&\Big(\Big(f^{(F)}_{s,n} + f^{(\lambda)}_n\Big) * \mathcal{F}^{-1}\left[\varphi_n^{-1}(-\cdot)\mathcal{F}K_{h_n}\right](Z_k)\Big)\\ &\times\Big(\Big(f^{(F)}_{t,n} + f^{(\lambda)}_n\Big) * \mathcal{F}^{-1}\left[\varphi_n^{-1}(-\cdot)\mathcal{F}K_{h_n}\right](Z_k)\Big),\end{aligned}$$

and

$$\widehat{\Sigma}^{N}_{s,t,n} := \frac{1}{\Delta^2}\frac{1}{n}\sum_{k=1}^{n}\Big(f^{(N)}_{s,n} * \mathcal{F}^{-1}\left[\varphi_n^{-1}(-\cdot)\mathcal{F}K_{h_n}\right](Z_k)\Big)\Big(f^{(N)}_{t,n} * \mathcal{F}^{-1}\left[\varphi_n^{-1}(-\cdot)\mathcal{F}K_{h_n}\right](Z_k)\Big)$$

respectively. Then, the same arguments justify the use in what follows of $\widehat{Q}^N_n(\alpha)$ and $\widehat{Q}^{F,\lambda}_n(\alpha)$, which denote consistent estimators of the upper α-quantile of the supremum norm of centred Gaussian processes with the covariances of the last two displays.

4.2.2 Confidence regions

Given the estimators we just constructed, we propose

$$\widehat{C}^F_n(\alpha) := \left\{G :\, G(t) \in \left[\widehat{F}_n(t) - \frac{\widehat{Q}^F_n(\alpha)}{\sqrt{n}}, \widehat{F}_n(t) + \frac{\widehat{Q}^F_n(\alpha)}{\sqrt{n}}\right] \text{ for all } t \in \mathbb{R}\right\}$$

as a confidence band of level $\alpha \in (0,1)$ for F, and

$$\widehat{C}^\theta_n(\alpha) := \left[\hat{\theta}_n - \hat{\sigma}_{\theta,n}\frac{Q^N(\alpha)}{\sqrt{n}}, \hat{\theta}_n + \hat{\sigma}_{\theta,n}\frac{Q^N(\alpha)}{\sqrt{n}}\right]$$

as a confidence interval of level $\alpha \in (0,1)$ for $\theta \in \{\lambda, p\} \cup \{p_j : j \in \mathbb{Z}\backslash\{0\}\}$. By the arguments given in the previous section and under the corresponding assumptions of Theorem 3.3.5, the global confidence band for F and the confidence interval for any θ indeed have asymptotic coverage

$$\lim_{n\to\infty} Pr\Big(F \in \widehat{C}_n^F(\alpha)\Big) = \lim_{n\to\infty} Pr\Big(\theta \in \widehat{C}_n^\theta(\alpha)\Big) = 1-\alpha.$$

4.2.3 Goodness-of-fit tests

The confidence regions we just constructed allow us to derive goodness-of-fit tests with asymptotic level α as follows. Let H_0^θ represent the null-hypothesis that $\theta \in \{F, \lambda, p\} \cup \{p_j : j \in \mathbb{Z}\backslash\{0\}\}$ belongs to a certain set $\mathcal{S}^\theta$. Then we propose to accept H_0^θ if $T_n^\theta := \mathbb{1}\{\mathcal{S}^\theta \cap \widehat{C}_n^\theta(\alpha) = \emptyset\}$ is zero and reject it otherwise. If $\theta \in \mathcal{S}^\theta$ and the corresponding assumptions of Theorem 3.3.5 are satisfied

$$\lim_{n\to\infty} Pr\Big(T_n^\theta = 1\Big) \leq \lim_{n\to\infty} Pr\Big(\theta \notin \widehat{C}_n^\theta(\alpha)\Big) = \alpha.$$

To test whether the jump distribution is purely discrete, purely absolutely continuous or a mixture of the two we propose $T_n^d := \mathbb{1}\{1 \notin \widehat{C}_n^p(\alpha)\}$, $T_n^{ac} := \mathbb{1}\{0 \notin \widehat{C}_n^p(\alpha)\}$ and $T_n^{d+ac} := 1 - T_n^d\, T_n^{ac}$. If the assumptions of Theorem 3.3.5b are satisfied these tests have asymptotic level α under their respective null-hypotheses and the third test further satisfies that whenever $p \in (0,1)$

$$\lim_{n\to\infty} Pr\big(T_n^{d+ac} = 1\big) \leq \lim_{n\to\infty} Pr\big(T_n^d = 0\big) + \lim_{n\to\infty} Pr\big(T_n^{ac} = 0\big) = 0.$$

Combinations of these tests provide tests for several parameters at once due to Theorem 3.3.5. For instance, to check whether a data set comes from a compound Poisson process whose jump distribution and intensity belong to some sets $\mathcal{S}^F$ and $\mathcal{S}^\lambda$, the test $T_n^{F,\lambda} := 1 - (1 - T_n^F)(1 - T_n^\lambda)$ can be used.

We now propose two ways to test whether the defining parameters of the compound Poisson process are equal to some values $\widetilde{F}$, $\tilde{\lambda}$ and $\tilde{\gamma}$. Define

$$\begin{aligned} D_n^{1)} :=& \sqrt{n}\Big\|\big(\widehat{F}_n - \widetilde{F}\big) + \big(\hat{\lambda}_n - \tilde{\lambda}\big) + \big(\hat{\gamma}_n - \tilde{\gamma}\big)\Big\|_\infty \\ =& \Big\|\sqrt{n}\Big(\big(\widehat{F}_n - F\big) + \big(F - \widetilde{F}\big) + \big(\hat{\lambda}_n - \lambda\big) + \big(\lambda - \tilde{\lambda}\big) + \big(\hat{\gamma}_n - \gamma\big) + \big(\gamma - \tilde{\gamma}\big)\Big)\Big\|_\infty \end{aligned}$$

and

$$\begin{aligned} D_n^{2)} :=& \sqrt{n}\Big\|\big(\widehat{N}_n - \tilde{\lambda}\widetilde{F}\big) + \big(\hat{\gamma}_n - \tilde{\gamma}\big)\Big\|_\infty \\ =& \Big\|\sqrt{n}\Big(\big(\widehat{N}_n - \lambda F\big) + \big(\lambda F - \tilde{\lambda}\widetilde{F}\big) + \big(\hat{\gamma}_n - \gamma\big) + \big(\gamma - \tilde{\gamma}\big)\Big)\Big\|_\infty. \end{aligned}$$

Then, if the assumptions of Theorem 3.3.5a are satisfied and under the null-hypothesis $H_0 : \|F - \widetilde{F}\|_\infty + |\lambda - \tilde{\lambda}| + |\gamma - \tilde{\gamma}| = 0$, we have that $D_n^{1)} \to \|\mathbb{L}^{F,\lambda}\|_\infty$ and $D_n^{2)} \to \|\mathbb{B}^N\|_\infty$ in law as $n \to \infty$. In view of the arguments included at the beginning of the section, the tests $T_n^{1)} := \mathbb{1}\{D_n^{1)} > \widehat{Q}_n^{F,\lambda}(\alpha)\}$ and $T_n^{2)} := \mathbb{1}\{D_n^{2)} > \widehat{Q}_n^{N}(\alpha)\}$ have asymptotic level α. In practice, we should choose the test whose underlying limiting process has smallest covariance structure. This strongly depends on the value of λ so we have to rely on its estimator to decide. It should also be possible to develop tests that improve upon any of these two by allowing for more general linear combinations of the quantities in $D_n^{1)}$ and $D_n^{2)}$, and varying the general coefficients. When generalising $D_n^{1)}$ the resulting tests have the unlikely drawback that under an alternative to H_0 they may have asymptotic level α. This is because in the expressions for $D_n^{1)}$ disagreements between the underlying and the proposed one-dimensional parameters may cancel out. Varying the coefficients in the linear combinations solves this limitation.

4.2.4 Two-sample tests

Throughout let $Z^{1)} := \big(Z_1^{1)}, \dots, Z_m^{1)}\big)$ and $Z^{2)} := \big(Z_1^{2)}, \dots, Z_n^{2)}\big)$ be two sets of observed increments from two independent compound Poisson processes. We denote their respective parameter $\theta \in \{\lambda, \gamma\} \cup \{p_j : j \in \mathbb{Z}\setminus\{0\}\}$ or $\Theta \in \{F, N\}$ and its estimator $\hat{\theta}$ or $\hat{\Theta}$ with the corresponding superscripts 1) and 2). Define

$$D_{m,n}^{\theta} := \sqrt{\frac{mn}{m+n}}\ \Big|\hat{\theta}_m^{1)} - \hat{\theta}_n^{2)}\Big| \qquad \text{and} \qquad D_{m,n}^{\Theta} := \sqrt{\frac{mn}{m+n}}\ \Big\|\widehat{\Theta}_m^{1)} - \widehat{\Theta}_n^{2)}\Big\|_\infty,$$

and, writing $\rho_{m,n} := m/(m+n)$, notice that

$$\sqrt{\frac{mn}{m+n}}\,\big(\hat{\theta}_m^{1)} - \hat{\theta}_n^{2)}\big) = \sqrt{1-\rho_{m,n}}\,\sqrt{m}\,\big(\hat{\theta}_m^{1)} - \theta^{1)}\big) - \sqrt{\rho_{m,n}}\,\sqrt{n}\,\big(\hat{\theta}_n^{2)} - \theta^{2)}\big) + \sqrt{\frac{mn}{m+n}}\,\big(\theta^{1)} - \theta^{2)}\big)$$

and that the same relationship holds for any Θ. Then, under the null-hypothesis $H_0 : \theta^{1)} = \theta^{2)} = \theta$ or $\Theta^{1)} = \Theta^{2)} = \Theta$ and the corresponding assumptions of Theorem 3.3.5, we have that $D_{m,n}^{\theta} \to^d \sigma_\theta |N(0,1)|$ and $D_{m,n}^{\Theta} \to^d \|\mathbb{L}^{\Theta}\|_\infty$ as $m, n \to \infty$, where $\mathbb{L}^{\Theta}$ is $\mathbb{G}^F$ if $\Theta = F$ and $\mathbb{B}^N$ if $\Theta = N$. Define $\hat{\sigma}^2_{\theta,m,n} := \big(1-\rho_{m,n}\big)\,\hat{\sigma}^2_{\theta^{1)},m} + \rho_{m,n}\,\hat{\sigma}^2_{\theta^{2)},n}$ and, in line with the definition of $\widehat{Q}_n^{\Theta}(\alpha)$, let $\widehat{Q}_{m,n}^{\Theta}(\alpha)$ be a consistent estimator of the upper α-quantile of the supremum norm of the centred Gaussian process with covariance structure

$$\Sigma^{\Theta}_{s,t,m,n} := \big(1-\rho_{m,n}\big)\,\Sigma^{\Theta^{1)}}_{s,t,m} + \rho_{m,n}\,\Sigma^{\Theta^{2)}}_{t,s,n}, \qquad s,t \in \mathbb{R}\,.$$

Then, we accept the respective null-hypothesis if the test

$$T_{m,n}^{\theta} := \mathbb{1}\{D_{m,n}^{\theta} > \hat{\sigma}_{\theta,m,n} Q^N(\alpha)\} \quad \text{or} \quad T_{m,n}^{\Theta} := \mathbb{1}\{D_{m,n}^{\Theta} > \widehat{Q}_{m,n}^{\Theta}(\alpha)\}$$

is zero and reject it otherwise. These tests have asymptotic level α and, under an alternative to H_0, $Pr(T^{\theta}_{m,n}=1), Pr(T^{\Theta}_{m,n}=1)\to 1$ as $m,n\to\infty$ due to $D^{\theta}_{m,n}, D^{\Theta}_{m,n}\to\infty$ as $m,n\to\infty$.

Two-sample tests for several parameters and hence for the process as a whole can be derived using the same arguments. For the latter we propose the following two. Define

$$D^{1)}_{m,n} := \sqrt{\frac{mn}{m+n}}\left\|\left(\widehat{F}^{1)}_m - \widehat{F}^{2)}_n\right) + \left(\hat{\lambda}^{1)}_m - \hat{\lambda}^{2)}_n\right) + \left(\hat{\gamma}^{1)}_m - \hat{\gamma}^{2)}_n\right)\right\|_{\infty}$$

and

$$D^{2)}_{m,n} := \sqrt{\frac{mn}{m+n}}\left\|\left(\widehat{N}^{1)}_m - \widehat{N}^{2)}_n\right) + \left(\hat{\gamma}^{1)}_m - \hat{\gamma}^{2)}_n\right)\right\|_{\infty}.$$

Then, under the null-hypothesis $H_0 : \varphi^{1)} = \varphi^{2)} = \varphi$ and the assumptions of Theorem 3.3.5a, we have that $D^{1)}_{m,n} \to^d \|\mathbb{L}^{F,\lambda}\|_{\infty}$ and $D^{2)}_{m,n} \to^d \|\mathbb{B}^{N}\|_{\infty}$ as $m,n\to\infty$. Furthermore, let $\widehat{Q}^{F,\lambda}_{m,n}(\alpha)$ be a consistent estimator of the upper α-quantile of the supremum norm of the centred Gaussian process with covariance structure

$$\Sigma^{F,\lambda}_{s,t,m,n} := \left(1-\rho_{m,n}\right)\Sigma^{F^{1)},\lambda^{1)}}_{s,t,m} + \rho_{m,n}\,\Sigma^{F^{2)},\lambda^{2)}}_{s,t,n}, \qquad s,t\in\mathbb{R}\,.$$

Then we propose the tests with asymptotic level α

$$T^{1)}_{m,n} := \mathbb{1}\{D^{1)}_{m,n} > \widehat{Q}^{F,\lambda}_{m,n}(\alpha)\} \quad \text{and} \quad T^{2)}_{m,n} := \mathbb{1}\{D^{2)}_{m,n} > \widehat{Q}^{N}_{m,n}(\alpha)\}.$$

Finally, we remark that under an alternative to H_0, $Pr(T^{2)}_{m,n}=1)\to 1$ as $m,n\to\infty$ due to $D^{2)}_{m,n}\to\infty$ as $m,n\to\infty$. The same asymptotic behaviour is obtained for $T^{1)}_{m,n}$ except when $\lambda^{1)}+\gamma^{1)} = \lambda^{2)}+\gamma^{2)}$, in which case it can also be attained if we allow for more general linear combinations in the expression for $D^{1)}_{m,n}$. A simple trick would be compute the test again with $(\hat{\lambda}^{2)}_n - \hat{\lambda}^{1)}_m)$ in place of $(\hat{\lambda}^{1)}_m - \hat{\lambda}^{2)}_n)$.

4.2.5 Approximations for $\lambda\Delta$ sufficiently small

The variance $\sigma^2_{\mathbb{L}}$ in Theorem 3.3.5 and its particular cases are not very intuitive at first sight as a consequence of the inverse nature of the problem under the low-frequency observation regime. However, if $\Delta=\Delta_n\to 0$ with $n\Delta_n\to\infty$ as $n\to\infty$ the inverse nature vanishes because the jumps are observed directly and its expression simplifies. Lemma 3.3.2 quantified the simplification and it allowed us to give more insight into our estimators. Since all the statistical applications discussed so far depend heavily on $\sigma^2_{\mathbb{L}}$, here we explore their simplifications when Δ, or rather $\lambda\Delta$, is sufficiently small. If the jump distribution has no discrete component some of these provide *process-free* procedures.

In view of Lemma 3.3.2 and the expression for $\sigma^2_{\mathbb{L}}$ in (3.3.3), it follows that

$$\frac{\Delta}{\lambda}\,\sigma^2_{\mathbb{L}} = \int_{\mathbb{R}} f^{(\mathbb{L})}(x)^2\,dF(x) + O\big(\lambda\Delta\big), \qquad s,t\in\mathbb{R}\,. \tag{4.2.3}$$

Simple algebra shows this expression provides the approximations to the variances we introduced in the discussions of Section 3.3. We refer the reader to that section for more details. The approximations of the difference of a one-dimensional parameter and its estimator rescaled by $\sqrt{n}$ therein are then justified because convergence of the variances guarantees convergence in distribution to centred normal random variables. In the case of N and F, the approximations of their differences follow because the terms in the big O notation of the lemma can be bounded by others depending on the supremum-norm of $f^{(\mathbb{L})}$ and moreover $\sup_{t\in\mathbb{R}}\|f_t\|_\infty \vee \|F(t)\|_\infty < \infty$. Furthermore note that the Δ/λ term on the left hand side of the last display indicates these differences should be further rescaled by $\sqrt{\Delta/\lambda}$. When dealing with F, p_j or p, their covariance and variances carry a λ^{-2} extra term, so the rescaling for these quantities should be $\sqrt{\lambda\Delta}$ instead. We justify this as follows. The quantity $\lambda\Delta n$ is the expected number of jumps the compound Poisson process has given in the Δn observation window. Therefore, when Δ is small enough, this product is approximately the total number of non-zero increments we observe. Due to our estimators being efficient we expect those for F, p_j and p to disregard the remaining increments and this justifies the rescaling in their differences with their respective estimators. For the rest of parameters except for γ, which is of no interest in this remark, a λ factor is implicitly included in their expressions and their estimators, thus justifying the rescaling by $\sqrt{\Delta/\lambda}$ instead.

When the jump distribution of the compound Poisson process is absolutely continuous, $p = 0$, the approximations to the limits can be further simplified as follows. In particular, expression (4.2.3) implies that $\sqrt{\lambda\Delta n}\big(\widehat{F}_n - F\big)$ converges in law to a Gaussian process with covariance structure approximately equal to $F(\min\{s,t\}) - F(s)F(t)$. This agrees with classical Donsker's theorem and, consequently, if $p = 0$ we propose to substitute the quantiles $\widehat{Q}^F_n$ and $\widehat{Q}^F_{m,n}$ by that of $(\lambda\Delta)^{-1/2}\sup_{r\in[0,1]}|\mathbb{G}(r)|$, where $\mathbb{G}$ is a standard Brownian bridge. Similarly and under the respective null-hypotheses, $\sqrt{\lambda\Delta}\,D^{1)}_n$, $\sqrt{\lambda\Delta}\,D^{1)}_{m,n}$ and $\sqrt{\Delta/\lambda}\,D^{2)}_n$, $\sqrt{\Delta/\lambda}\,D^{2)}_{m,n}$ converge in law to the supremum norm of two Gaussian processes whose covariances can be approximated by $F(\min\{s,t\})-F(s)F(t)+\lambda^2$ and $F(\min\{s,t\})$, respectively. As a result, if $p = 0$ we propose to substitute the quantiles $\widehat{Q}^{F,\lambda}_n, \widehat{Q}^{F,\lambda}_{m,n}$ and $\widehat{Q}^N_n, \widehat{Q}^N_{m,n}$ by those of $(\lambda\Delta)^{-1/2}\sup_{r\in[0,1]}|\mathbb{G}(r) + \mathbb{B}(\sigma^2)|$ and $\sqrt{\lambda/\Delta}\sup_{r\in[0,1]}|\mathbb{B}(r)|$, respectively, where $\mathbb{B}$ is a standard Brownian motion. When using these approximations for the applications in Sections 4.2.2 and 4.2.3, λ can be estimated either by $\hat{\lambda}_n$ or simply by the proportion of non-zero increments we observe divided by Δ. Similarly, when using them for the applications in Section 4.2.4, λ can be estimated either by $\hat{\lambda}_{m,n} := (1-\rho_{m,n})\hat{\lambda}^{1)}_n + \rho_{m,n}\hat{\lambda}^{2)}_m$ or simply by the $\rho_{m,n}$-weighted average of the proportions

of non-zero increments observed in each independent compound Poisson process divided by Δ.

4.3 Implementation

In this section we include details of the implementation of the estimators of the compound Poisson process parameters in Coca [2015] and Buchmann and Grübel [2003], and of limiting variances and covariances therein. Note that all the statistical procedures constructed in Sections 4.2.2, 4.2.3, 4.2.4 and 4.2.5 make use only of these quantities so their implementation follows as mentioned in them. From now on we take $\Delta = 1$ without loss of generality as otherwise it can be absorbed into the value of λ.

4.3.1 Spectral approach-based estimators

In order to implement the estimators introduced in Chapter 3 and in Section 4.2.1, a kernel function K satisfying the assumptions in (3.2.1) has to be chosen. Throughout we take it to be such that $\mathcal{F}K(u) = (1-u^2)^2 \mathbb{1}_{[-1,1]}(u)$, $u \in \mathbb{R}$. Its explicit expression is not used in the implementation so we do not include it here. It is related to the Bessel functions of the first kind and their properties guarantee K satisfies the desired assumptions. A similar kernel function is considered in van Es et al. [2007], and in Wand [1998] in the related deconvolution problem, who also emphasise that the choice of K is not the most important factor in the performance of the estimator.

Recall the empirical distribution function $\varphi_n(u) := \frac{1}{n}\sum_{k=1}^n e^{iuZ_k}$, $u \in \mathbb{R}$ and note that any estimator of Section 3.2 is of the form $\int_{\mathbb{R}} f_n(x)\mathcal{F}^{-1}[\mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}](x)dx$. All of these can be more generally written as

$$T_n = T_n(f_n, g_n) := \int_{\mathbb{R}} f_n(x)\mathcal{F}^{-1}g_n(x)\,dx = \int_{\mathbb{R}} g_n(u)\mathcal{F}^{-1}f_n(u)\,du, \tag{4.3.1}$$

where f_n is compactly supported and bounded, and g_n satisfies $\mathrm{supp}(g_n) \subseteq [-h_n^{-1}, h_n^{-1}]$, $g_n(-\cdot) = \overline{g_n}$, where $\bar{\cdot}$ denotes the complex conjugate, and it is bounded in sets of probability approaching 1 as $n \to \infty$. Hence, $f_n, g_n \in L^2(\mathbb{R})$ in such sets and T_n is well-defined there. This general estimator can be approximated numerically by discretising the last integral: for any $\eta > 0$ and $M \in \mathbb{N}$ such that $\eta M = h_n^{-1}$ to accommodate the support of g_n, define

$$\widehat{T}_n = \widehat{T}_n(f_n, g_n) := \sum_{j=-M}^{M} \omega_j g_n(j\eta)\mathcal{F}^{-1}f_n(j\eta),$$

where $(\omega_j)_{j=-M}^M$ are the discretisation weights. We choose the weights according to com-

posite Simpson's rule, i.e.

$$\omega_j = \frac{\eta}{3} \times \begin{cases} 1 & \text{if } j = -M, M, \\ 2 & \text{if } j = -M+1, -M+3, \ldots, M-1, \\ 4 & \text{if } j = -M+2, -M+4, \ldots, M-2. \end{cases}$$

Note that any f_n from Section 3.2 except for $f_n^{(\gamma)}$ can be written as a linear combination of indicators of open or closed intervals with lower and upper limits a_n and b_n, respectively, satisfying that $\operatorname{sign}(a_n) = \operatorname{sign}(b_n)$. If f_n is such an indicator

$$\begin{aligned} \widehat{T}_n &= \frac{b_n - a_n}{2\pi}\omega_0 g_n(0) + \frac{i}{2\pi\eta}\sum_{0<|j|\le M} \omega_j g_n(j\eta)\frac{e^{-ib_n j\eta} - e^{-ia_n j\eta}}{j} \\ &= \frac{b_n - a_n}{2\pi}\omega_0 g_n(0) + \frac{1}{\pi\eta}\sum_{j=1}^{M} \omega_j \operatorname{Im}\left(g_n(j\eta)\frac{e^{-ia_n j\eta} - e^{-ib_n j\eta}}{j}\right), \end{aligned} \tag{4.3.2}$$

where the last equality is justified because $g_n(-\cdot)\frac{e^{ix\cdot}}{-\cdot} = -\overline{g_n(\cdot)\frac{e^{-ix\cdot}}{\cdot}}$. The last expression avoids some computations and round-off errors, it returns strictly real values and can be implemented directly. However, to further speed up the computations we approximate its value with the aid of the fast Fourier transform, which was popularised by Cooley and Tukey [1965] and, in its simplest form, was already described by C. F. Gauss in 1805. For this we introduce the following notation and impose some additional assumptions on M, η, ε_n and H_n: for any $v := (v_j)_{j=0}^M \in \mathbb{R}^{M+1}$ and $l = 0, \ldots, M$, define

$$\mathcal{DF}[v](l) := \sum_{j=0}^{M} v_j e^{-i\frac{2\pi}{M+1}jl} \quad \text{and} \quad \mathcal{DF}^{-1}[v](l) := \frac{1}{M+1}\sum_{j=0}^{M} v_j e^{i\frac{2\pi}{M+1}jl}.$$

Let v be such that $v_0 = 0$ and $v_j = \omega_j \frac{g_n(j\eta)}{j}$ for $j = 1, \ldots, M$, and, to equate the exponents of the last display to those in (4.3.2), let a_n and b_n belong to a symmetric grid of $2M+1$ elements that we now specify. Note that for the indicators we are considering $\varepsilon_n \le |a_n|, |b_n| \le H_n$ and recall the $\pm\varepsilon_n$-intervals around each jump in the definition of $f_{\cdot,n}^{(N)}$ in (3.2.6). Then we take the grid to be ε_n-equispaced and assume without loss of generality that $H_n = \varepsilon_n M$. Writing $l_x := [x/\varepsilon_n]$ for any $x \in \mathbb{R}$, where $[\cdot]$ denotes the closest integer function, and assuming $\eta\varepsilon_n = \frac{2\pi}{M+1}$, we have $\widehat{T}_n(f_n, g_n) \approx \widehat{\widehat{T}}_n(f_n, g_n)$, where for the f_n we are considering

$$\widehat{\widehat{T}}_n := \frac{b_n - a_n}{2\pi}\omega_0 g_n(0) + \frac{1}{\pi\eta} \times \begin{cases} (M+1)\operatorname{Im}\left(\mathcal{DF}^{-1}[v](-l_{a_n}) - \mathcal{DF}^{-1}[v](-l_{b_n})\right) & \text{if } a_n \le b_n < 0, \\ \operatorname{Im}\left(\mathcal{DF}[v](l_{a_n}) - \mathcal{DF}[v](l_{b_n})\right) & \text{if } 0 < a_n \le b_n. \end{cases}$$

The transformations $\mathcal{DF}$ and $\mathcal{DF}^{-1}$ can be computed in $O(M\log(M))$ operations by a

fast Fourier transform type algorithm. In particular, the well-known radix-2 Cooley–Tukey algorithm can be used if we assume $M+1$ is a power of 2. Hence, for a given bandwidth h_n we need M, η, ε_n and H_n to satisfy that

$$\eta M = h_n^{-1}, \quad H_n = \varepsilon_n M \quad \eta\varepsilon_n = \frac{2\pi}{M+1} \quad \text{and} \quad \log_2(M+1) \in \mathbb{N}\,. \tag{4.3.3}$$

At the end of the section we discuss the choices of these parameters but prior to that we give the implementation of all the estimators.

If $g_n = \mathrm{Log}(\varphi_n)\mathcal{F}K_{h_n}$, each can be approximated by

$$\begin{aligned}
\hat{\hat{\lambda}}_n &:= \hat{\hat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]}, g_n\big) + \hat{\hat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]}, g_n\big),\\
\hat{\hat{q}}_{j,n} &:= \hat{\hat{T}}_n\big(\mathbb{1}_{\left(\varepsilon_n(l_{J_j}-1),\varepsilon_n(l_{J_j}+1)\right)}, g_n\big), \quad j \in \mathbb{Z}\setminus\{0\},\\
\hat{\hat{q}}_n &:= \sum_{\substack{|j|\le H_n/\varepsilon\\ j\neq 0}} \hat{\hat{q}}_{j,n},\\
\hat{\hat{N}}_n(t) &:= \mathbb{1}_{t\ge -H_n}\hat{\hat{T}}_n\big(\mathbb{1}_{[-H_n,\varepsilon_n\min\{l_t,-1\}]}, g_n\big) + \mathbb{1}_{t\ge\varepsilon_n}\hat{\hat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,\varepsilon_n\min\{l_t,M\}]}, g_n\big),\\
\hat{\hat{F}}_n &:= \hat{\hat{\lambda}}_n^{-1}\hat{\hat{N}}_n, \qquad \hat{\hat{p}}_{j,n} := \hat{\hat{\lambda}}_n^{-1}\hat{\hat{q}}_{j,n}, \;\; j \in \mathbb{Z}\setminus\{0\}, \quad \text{and} \quad \hat{\hat{p}}_n := \hat{\hat{\lambda}}_n^{-1}\hat{\hat{q}}_n.
\end{aligned}$$

By similar arguments to those employed above and noting that for $f_n = c^{-1}\cdot\mathbb{1}_{(-h_n,h_n)}$

$$\mathcal{F}^{-1}f_n(u) = \frac{i}{cu\pi}\left(h_n\cos(uh_n) - \frac{sin(uh_n)}{u}\right),$$

$\hat{\hat{\gamma}}_n$ can be approximated by

$$\begin{aligned}
\hat{\hat{\gamma}}_n &:= -2\sum_{j=1}^{M}\omega_j\,\mathrm{Im}(g_n(j\eta))\,\mathrm{Im}(\mathcal{F}^{-1}f_n(j\eta))\\
&= \frac{2}{c\pi\eta}\sum_{j=1}^{M}\omega_j\arctan\frac{\mathrm{Im}(\varphi_n(j\eta))}{\mathrm{Re}(\varphi_n(j\eta))}\mathcal{F}K_{h_n}(j\eta)\frac{1}{j}\left(\frac{\sin(j\eta h_n)}{j\eta} - h_n\cos(j\eta h_n)\right),
\end{aligned}$$

where $c := 2(\int_0^1 K - K(1)) \approx 0.015637923740278$ for the choice of kernel mentioned above. Note that in the implementation of $\hat{\hat{\gamma}}_n$ we have chosen not to compute a fast Fourier transform. The reason is that, effectively, we are only interested in one value of the discrete Fourier transform and it is therefore more costly to compute the whole vector than to directly compute the desired element of it.

Some remarks about these approximations to the estimators are due. Note that the expression for $\hat{\hat{N}}_n$ is based on (3.2.3) instead of on (3.2.5). This is justified by the fact that we are defining $\hat{\hat{N}}_n$ on a grid of $[-H_n,-\varepsilon_n]\cup[\varepsilon_n,H_n]$ and in practice this acts as the discretisation implied by $f_{t,n}^N$ defined in (3.2.6). Similarly, in the definition of $\hat{\hat{q}}_{j,n}$,

and consequently in those of $\hat{\hat{q}}_n, \hat{\hat{p}}_{j,n}$ and $\hat{\hat{p}}_n$ too, centring the integration interval around the desired jump is not necessary and therefore we do not impose this. In the following sections we discuss in more detail the choice of the grid when there is a non-zero discrete jump component. Lastly, notice that in the definition of $\hat{\hat{q}}_n$ we have summed up to H_n instead of up to $\widetilde{H}_n$. This is not important given that in practice we are concerned with the finite sample properties rather than with the asymptotic ones.

We are missing to discuss the implementation of the variance and covariance estimators in Section 4.2.1. Any of these involves quantities of the form

$$f_n * \mathcal{F}^{-1}\big[\varphi_n^{-1}(-\cdot)\mathcal{F}K_{h_n}\big](Z) = \int_{\mathbb{R}} f_n(x)\, \mathcal{F}^{-1}\big[e^{iZ\cdot}\varphi_n^{-1}\mathcal{F}K_{h_n}\big](x)\, dx,$$

where the equality is justified by the symmetry of K. The functions f_n and $e^{iZ\cdot}\varphi_n^{-1}\mathcal{F}K_{h_n}$ enjoy the same properties as f_n and g_n in the definition of T_n in (4.3.1). Furthermore, all the f_n under consideration are linear combinations of the same type of indicators as above. Therefore, if $g_n(Z) = g_n(\cdot, Z) = e^{iZ\cdot}\varphi_n^{-1}\mathcal{F}K_{h_n}$ and using the same arguments as before, we approximate the estimators of Section 4.2.1 by

$$\begin{aligned}
\hat{\hat{\sigma}}^2_{\lambda,n} &:= \frac{1}{n}\sum_{k=1}^{n}\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]}, g_n(Z_k)\big) + \widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]}, g_n(Z_k)\big)\Big)^2,\\
\hat{\hat{\sigma}}^2_{p_j,n} &:= \frac{1}{\hat{\hat{\lambda}}^2_n}\frac{1}{n}\sum_{k=1}^{n}\Bigg(\widehat{\widehat{T}}_n\big(\mathbb{1}_{\left(\varepsilon_n(l_{J_j}-1),\varepsilon_n(l_{J_j}+1)\right)}, g_n(Z_k)\big)\\
&\qquad - \hat{\hat{p}}_{j,n}\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]}, g_n(Z_k)\big) + \widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]}, g_n(Z_k)\big)\Big)\Bigg)^2,\\
\hat{\hat{\sigma}}^2_{p,n} &:= \frac{1}{\hat{\hat{\lambda}}^2_n}\frac{1}{n}\sum_{k=1}^{n}\Bigg(\sum_{\substack{|j|\le H_n/\varepsilon\\ j\neq 0}}\widehat{\widehat{T}}_n\big(\mathbb{1}_{\left(\varepsilon_n(l_{J_j}-1),\varepsilon_n(l_{J_j}+1)\right)}, g_n(Z_k)\big)\\
&\qquad - \hat{\hat{p}}_n\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]}, g_n(Z_k)\big) + \widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]}, g_n(Z_k)\big)\Big)\Bigg)^2,
\end{aligned}$$

and, for all $s,t\in\mathbb{R}$,

$$\begin{aligned}
\widehat{\widehat{\Sigma}}^F_{s,t,n} := \frac{1}{\hat{\hat{\lambda}}^2_n}\frac{1}{n}\sum_{k=1}^{n}\Bigg(&\mathbb{1}_{s\ge -H_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,\varepsilon_n\min\{l_s,-1\}]}, g_n(Z_k)\big) + \mathbb{1}_{s\ge\varepsilon_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,\varepsilon_n\min\{l_s,M\}]}, g_n(Z_k)\big)\\
&- \widehat{\widehat{F}}_n(s)\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]}, g_n(Z_k)\big) + \widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]}, g_n(Z_k)\big)\Big)\Bigg)\\
\times\Bigg(&\mathbb{1}_{t\ge -H_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,\varepsilon_n\min\{l_t,-1\}]}, g_n(Z_k)\big) + \mathbb{1}_{t\ge\varepsilon_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,\varepsilon_n\min\{l_t,M\}]}, g_n(Z_k)\big)\\
&- \widehat{\widehat{F}}_n(t)\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]}, g_n(Z_k)\big) + \widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]}, g_n(Z_k)\big)\Big)\Bigg),
\end{aligned}$$

$$\widehat{\widehat{\Sigma}}^{N}_{s,t,n} := \frac{1}{n}\sum_{k=1}^{n}\Big(\mathbb{1}_{s\geq -H_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,\varepsilon_n\min\{l_s,-1\}]},g_n(Z_k)\big)+\mathbb{1}_{s\geq\varepsilon_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,\varepsilon_n\min\{l_s,M\}]},g_n(Z_k)\big)\Big)$$
$$\times\Big(\mathbb{1}_{t\geq -H_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,\varepsilon_n\min\{l_t,-1\}]},g_n(Z_k)\big)+\mathbb{1}_{t\geq\varepsilon_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,\varepsilon_n\min\{l_t,M\}]},g_n(Z_k)\big)\Big)$$

and

$$\widehat{\widehat{\Sigma}}^{F,\lambda}_{s,t,n} := \frac{1}{\hat{\lambda}_n^2}\frac{1}{n}\sum_{k=1}^{n}\Big(\mathbb{1}_{s\geq -H_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,\varepsilon_n\min\{l_s,-1\}]},g_n(Z_k)\big)+\mathbb{1}_{s\geq\varepsilon_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,\varepsilon_n\min\{l_s,M\}]},g_n(Z_k)\big)$$
$$+\big(\hat{\lambda}_n^2-\widehat{\widehat{F}}_n(s)\big)\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]},g_n(Z_k)\big)+\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]},g_n(Z_k)\big)\Big)\Big)$$
$$\times\Big(\mathbb{1}_{t\geq -H_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,\varepsilon_n\min\{l_t,-1\}]},g_n(Z_k)\big)+\mathbb{1}_{t\geq\varepsilon_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,\varepsilon_n\min\{l_t,M\}]},g_n(Z_k)\big)$$
$$+\big(\hat{\lambda}_n^2-\widehat{\widehat{F}}_n(t)\big)\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]},g_n(Z_k)\big)+\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]},g_n(Z_k)\big)\Big)\Big).$$

Despite the seemingly complicated implementation of these estimators, it can be written in a few lines of code. Notice that, instead of applying the discrete Fourier transform to a vector v determined by $g_n(Z)=e^{iZ\cdot}\varphi_n^{-1}\mathcal{F}K_{h_n}$, we can apply it to a vector v with $g_n(a_n,b_n)=\big(e^{ib_n\cdot}-e^{ia_n\cdot}\big)\,\varphi_n^{-1}\mathcal{F}K_{h_n}$. This results in notationally simpler expressions and avoids restricting the computation of the quantities above to an ε_n-equally spaced grid of $[-H_n,-\varepsilon_n]\cup[\varepsilon_n,H_n]$. In turn, it requires discretising the values of the observed increments Z_k, $k=1,\dots,n$, and hence it changes the fundamental quantities driving the final estimates. We opt not to do this and choose the approximate estimators above instead.

The use of these estimators is to approximate the quantiles of the absolute value of the limiting normal distributions of Section 3.3 and of the supremum of the absolute value of the limiting Gaussian processes of the same section. For $\alpha\in(0,1)$, typically $\alpha=0.05$, and writing $\hat{\sigma}^2_{\theta,n}$ for $\theta\in\{\lambda,p\}\cup\{p_j:j\in\mathbb{Z}\backslash\{0\}\}$, in the following we approximate the former quantiles by sampling 5000 independent standard normal distributions, ordering their absolute values increasingly, selecting the element in the $[5000(1-\alpha)]$-th position and multiplying it by $\hat{\sigma}_{\theta,n}$. For each of the infinite-dimensional quantities we simulate 5000 paths of the respective approximate Gaussian process, order the maximums of the absolute values of the paths increasingly and select the element in the $[5000(1-\alpha)]$-th position. To simulate each (discrete) path, we first compute the covariance matrix of the discretised Gaussian process using the respective quantity $\widehat{\widehat{\Sigma}}^{\Theta}_{s,t,n}$ from above, where $\Theta\in\{F,N,(F,\lambda)\}$. Then, in the case of $\Theta=N$, we perform its (lower) Cholesky factorisation and multiply the resulting matrix by a column vector of independent standard normal distributions. This is justified because, as mentioned in Section 2.6, the limiting process is of Brownian motion type. In the other cases we not only have to force the initial point of the path to be zero but also the final point. Thus, we proceed similarly to the standard interpolation method to sample approximate paths of a Brownian bridge (cf. p. 82-86 of Glasserman [2004]): set

the values at $-H_n$ and H_n to zero and sample two zero-mean normal distributions with variance $\widehat{\widehat{\Sigma}}^{\Theta}_{-H_n+\varepsilon_n,-H_n+\varepsilon_n,n}$ and $\widehat{\widehat{\Sigma}}^{\Theta}_{H_n-\varepsilon_n,H_n-\varepsilon_n,n}$. Take these to be the values of the path at $-H_n+\varepsilon_n$ and $H_n-\varepsilon_n$ and obtain the value at the (approximate) mid-point by sampling from the resulting conditional distribution. The mid-point now becomes an end-point and we can do this iteratively with the subsequent mid-points until the path is computed for every point of the grid. Note that, for a particular mid-point s and interpolating end-points r and t, $-H_n+\varepsilon_n \leq r < s < t \leq H_n-\varepsilon_n$, the conditional distribution is simply given by

$$\widehat{\widehat{\Theta}}_n(s) \,|_{\widehat{\widehat{\Theta}}_n(r),\widehat{\widehat{\Theta}}_n(t)} \sim N\left(\left(\widehat{\widehat{\Sigma}}^{\Theta}_{s,r,n},\widehat{\widehat{\Sigma}}^{\Theta}_{s,t,n}\right)\left(\widehat{\widehat{\Sigma}}^{\Theta}_{-s,n}\right)^{-1}\begin{pmatrix}\widehat{\widehat{\Theta}}_n(r)\\ \widehat{\widehat{\Theta}}_n(t)\end{pmatrix},\right.$$
$$\left.\widehat{\widehat{\Sigma}}^{\Theta}_{s,s,n}-\left(\widehat{\widehat{\Sigma}}^{\Theta}_{s,r,n},\widehat{\widehat{\Sigma}}^{\Theta}_{s,t,n}\right)\left(\widehat{\widehat{\Sigma}}^{\Theta}_{-s,n}\right)^{-1}\begin{pmatrix}\widehat{\widehat{\Sigma}}^{\Theta}_{r,s,n}\\ \widehat{\widehat{\Sigma}}^{\Theta}_{t,s,n}\end{pmatrix}\right),$$

where, with some abuse of notation,

$$\widehat{\widehat{\Sigma}}^{\Theta}_{-s,n} := \begin{pmatrix}\widehat{\widehat{\Sigma}}^{\Theta}_{r,r,n} & \widehat{\widehat{\Sigma}}^{\Theta}_{r,t,n}\\ \widehat{\widehat{\Sigma}}^{\Theta}_{t,r,n} & \widehat{\widehat{\Sigma}}^{\Theta}_{t,t,n}\end{pmatrix}.$$

All these implementations apply to individual $\theta \in \{\lambda, p\} \cup \{p_j : j \in \mathbb{Z}\backslash\{0\}\}$ or $\Theta \in \{F, N, (F,\lambda)\}$. However, if we want to implement several of them together we should introduce the dependence structure given in Theorem 3.3.5. For instance, when performing simulations to construct procedures such as confidence regions for F and λ simultaneously, we first construct them for the former as just mentioned but for the latter we need to simulate its limiting normal distribution conditional on all the simulated values of F. The penultimate display readily generalises to conditional distributions with more than two conditioning values (cf. p. 65 in Glasserman [2004]) and this generalisation is what we use to construct the confidence interval for λ. In it we have to input the limiting covariance between the estimator of λ and that of $F(t)$, with t in the grid considered above. This is simply given by

$$\frac{1}{\hat{\lambda}_n}\frac{1}{n}\sum_{k=1}^{n}\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]},g_n(Z_k)\big)+\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]},g_n(Z_k)\big)\Big)$$
$$\times\Big(\mathbb{1}_{t\geq -H_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,\varepsilon_n\min\{l_t,-1\}]},g_n(Z_k)\big)+\mathbb{1}_{t\geq\varepsilon_n}\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,\varepsilon_n\min\{l_t,M\}]},g_n(Z_k)\big)$$
$$-\widehat{\widehat{F}}_n(t)\Big(\widehat{\widehat{T}}_n\big(\mathbb{1}_{[-H_n,-\varepsilon_n]},g_n(Z_k)\big)+\widehat{\widehat{T}}_n\big(\mathbb{1}_{[\varepsilon_n,H_n]},g_n(Z_k)\big)\Big)\Big).$$

In the same way, one combines the brackets in the expressions for $\hat{\sigma}^2_{\theta,n}$ and $\widehat{\widehat{\Sigma}}^{\Theta}_{s,t,n}$ above

to make joint inference on other parameters, including the mass at some of the potential atoms of the jump measure.

Lastly, we discuss the choices of the (hyper)parameters in (4.3.3), which apply to the implementation of all the estimators presented so far. After numerous simulations we have found that $\eta = 5^3/2^{10} \approx 0.1221$, $M = (\eta h_n)^{-1}$, $\varepsilon_n = 2\pi h_n M/(M+1) \approx 2\pi h_n$ and $H_n = \varepsilon_n M$, where $h_n \in [0.0005, 0.001]$, generally provide reasonable results in terms of accuracy and computational times. We note in passing that η should be fixed at first so that it does not change with other parameters, potentially resulting in considerably wrong or slow estimates. When taking h_n beyond the lower or upper limit in $[0.0005, 0.001]$, the implementations become too slow or start losing accuracy, respectively. The choice of the bandwidth returning best approximation and coverage values depends on the sample size n and on the true underlying parameters F and λ. Thus, this is discussed further in Section 4.3, where we propose a purely data-driven objective function to choose it. For now we mention that in general we take $h_n = 0.0005$ except for when n and λ are relatively small, in which case we take $h_n = 0.001$. Despite these being the preferred choices, we remark that the latter does not generally lead to a large deterioration of the results and therefore can be chosen in most cases for the sake of speed of computations. When computing the estimates above, we generally do not take the intervals to go down to $-H_n = -\varepsilon_n M$ or up to $H_n = \varepsilon_n M$ but truncate them to smaller values: H_n tends to be too large in comparison with the support of most of the mass of the objective function being estimated and worsens the estimations. Under the positivity assumption on the support of dF that Buchmann and Grübel [2003] make, they propose to take the left limit to be zero and the right limit to be $\max_k Z_k$, i.e. equal to the largest observation. In their setting this is a sensible choice because, given the fully nonparametric approach they take, the data cannot not inform us of the distribution any farther than that and in such cases we make the same choice. However, in the general setting for which our estimators are developed, it is not sensible because even the largest observation may arise from sums of jumps of different signs. When F has support on both sides of $\mathbb{R}$ we take the right and left limits to be $\max_k Z_k + 0.2(\max_k Z_k - \min_k Z_k)$ and $\min_k Z_k - 0.2(\max_k Z_k - \min_k Z_k)$, respectively. We have found this choice to perform well and we propose it as a rule of thumb when the support of F is not known a priori. We emphasise that H_n does not appear in the relationships arising from the fast Fourier transform so this rather arbitrary choice of it does not affect the radix-2 Cooley–Tukey algorithm.

4.3.2 Direct approach-based estimators

In this section we include the implementation of the estimators introduced in Buchmann and Grübel [2003]; those in Buchmann and Grübel [2004] follow directly from it. Recall that they assume $\gamma = 0$ and $\operatorname{supp}(F) \subseteq \mathbb{R}^+$, and they propose two estimators: the first makes inference on the probability mass function of a discrete jump distribution with

atoms on $\mathbb{N}$ and the second is developed for general distributions. We introduce their implementation in reverse order as we only claim originality for the second.

Recall from Section 2.1.2 that this second estimator in Buchmann and Grübel [2003] is

$$\widetilde{F}_n(t) := \Lambda(\widetilde{G}_{0,n})(t) := \sum_{i=1}^{\infty}(-1)^{i+1}\frac{e^{i\lambda}}{\lambda i}\widetilde{G}_{0,n}^{*i}(t), \quad t \in \mathbb{R}^+,$$

where

$$\widetilde{G}_{0,n}(t) := \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{[0,t]}(Z_k) - \tilde{q}_{0,n} := \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{(0,t]}(Z_k), \quad t \in \mathbb{R}^+ .$$

This is much simpler to implement than our estimator: note that $\widetilde{G}_{0,n}^{*i}$ is $(1-q_{0,n})^i$ times the empirical distribution of the sum of the elements of every i-tuple of the non-zero observations. Therefore we sum up all such combinations and compute the resulting function on the grid we mentioned at the end of the previous section, i.e. $[0, \max_k Z_k]$ split into intervals of length ε_n. We choose this increment as above to ensure fair comparison between the estimators and remark that this choice is not as important when implementing $\widetilde{F}_n$ as it is when doing so for $\widehat{F}_n$, as no integration is required here. In addition, we need to truncate the infinite sum in its definition and we take it to go up to $I = 101$. This is quite a large value but, as we show in Section 4.4, the fluctuations of $\widetilde{F}_n$ away from the origin may be very large and the larger I is the better they are controlled. We emphasise that it should be chosen to be an odd number as otherwise the leading term in the sum is negative and so is the implementation of $\widetilde{F}_n$. Lastly, we recall that Buchmann and Grübel [2003] assumed knowledge of λ and so do we when implementing it. We tried inputting the naive estimator $\check{\lambda}_n$, but when it is larger than the true λ it makes the sum fluctuate even further and, in practice, the estimation may deteriorate considerably.

To construct the statistical procedures of Section 4.2, we use the 'Studentisation' approach introduced in Section 4.2.1 for the same reasons mentioned therein. Therefore, to find an estimator of the covariance function of the limiting processes in Buchmann and Grübel [2003] we define

$$\widetilde{H}_n(t) := \frac{1}{\lambda}\sum_{i=1}^{I}(-1)^{i+1}e^{i\lambda}G_{0,n}^{*(i-1)}(t), \quad t \in \mathbb{R}^+,$$

where $\widetilde{G}_{0,n}^{*(i-1)}$ and I are as mentioned above. Additionally, define $\widetilde{H}_{0,n} := \widetilde{H}_n - \frac{e^{\lambda}}{\lambda}$. Then, the estimator we propose is

$$\widetilde{\Sigma}_{s,t,n} := \frac{1}{n}\sum_{k=1}^{n}\widetilde{H}_n(s-Z_k)\widetilde{H}_n(t-Z_k)\,\mathbb{1}_{(0,s\wedge t]}(Z_k) - e^{2\lambda}\widetilde{H}_{0,n}(s)\widetilde{H}_{0,n}(t), \quad s,t \in \mathbb{R}^+ .$$

Hence, we can approximate the limiting Gaussian process $\mathbb{B}^F$ in (2.1.12) by the centred Gaussian process with this covariance structure. Recall that, as discussed in Section 2.6,

$\mathbb{B}^F$ is of Brownian motion type. Consequently, we can simulate paths of the Gaussian process with covariance $\widetilde{\Sigma}_{s,t,n}$ with the simple Cholesky factorisation strategy proposed in the previous section. We remark that not all the procedures from Section 4.2 can be developed from the results in Buchmann and Grübel [2003] because they assume knowledge of λ, so the only uncertainty is in the parameter F. Furthermore, the norm $\|\cdot\|_\infty$ used in Section 4.2.1 should be substituted by the norm $\|\cdot\|_{\infty,\tau}$ introduced in Section 2.1.2. Hence, we are missing to choose a value of τ. As suggested by Bøgsted and Pitts [2010] and in line with expression (2.1.13) in our discussions in Section 2.1.2, we take

$$\tilde{\tau}_n := \min\left\{\tau : 10\,\tau \in \mathbb{N} \text{ and } \int_{\mathbb{R}^+} e^{-\tau x} d\widetilde{F}_n(x) < \frac{\log 2}{\lambda}\right\}.$$

Note that the results in Buchmann and Grübel [2003] hold for any τ satisfying the non-empirical counterpart of the second condition in this display. Therefore, the choice of the grid in the first condition is not very important as long as it results in a large enough value of $\tilde{\tau}_n$. Moreover, as we show in Section 4.4, in all experiments $\tilde{\tau}_n$ overestimates the minimum value for τ and no issues arise from this calculation.

In Buchmann and Grübel [2003] and Buchmann and Grübel [2004] they included details of the implementation of their recursive estimators and of the covariances of the limiting processes. We next include the implementation of that of Buchmann and Grübel [2003], and the implementation of the latter follows by the same arguments. We therefore claim no originality in what follows. Let $w_{j,n} := \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\{j\}}(Z_k)$ be the mass the empirical distribution gives to $j \in \mathbb{N}$ and assume $w_{0,n} > 0$. Then, they estimate the parameters λ and p_j by

$$\tilde{\lambda}_n := -\log w_{0,n} \quad \text{and} \quad \tilde{p}_{j,n} := \frac{1}{w_{0,n}}\left(\frac{w_{j,n}}{\tilde{\lambda}_n} - \frac{1}{j}\sum_{l=1}^{j-1} l\tilde{p}_{l,n} w_{j-l,n}\right), \quad j \in \mathbb{N}.$$

As mentioned at the end of Section 2.1.2, these expressions are obtained by inverting the distribution function of Z_1, given in terms of so-called Panjer recursion, and plugging in the empirical values of the mass function of Z_1. They are recursive in nature and therefore Buchmann and Grübel [2003] stop their computation at the largest observation. If $w_{0,n} = 0$ they set $\tilde{\lambda}_n = 1$, $p_{1,n} = 1$ and $p_{j,n} = 0$ for $j > 1$ so that the estimators are well-defined in the space $\mathbb{R}\times\ell_1$. Due to $Pr(w_{n_0} = 0) = (1 - e^{-\lambda})^n$ and $\lambda > 0$, this rather arbitrary definition does not affect the asymptotic nature of their theoretical results. However, for finite samples and in particular in simulations this is undesirable. When simulating these estimators we simply discard samples with no zero-increments and, in practice, we suggest to use $\hat{\tilde{\lambda}}_n$ in place of $\tilde{\lambda}_n$ so that $\tilde{p}_{j,n}$ are well-defined.

Buchmann and Grübel [2003] show joint convergence in distribution of their $\sqrt{n}$-rescaled and centred estimators above to a centred random element $(\xi, (B_j)_{j\in\mathbb{N}})$ taking

values in $\mathbb{R} \times \ell_1$. Its dependence structure can be implemented as follows assuming $w_{0,n} > 0$. Let

$$r_{0,n} := \frac{1}{w_{0,n}} \quad \text{and} \quad r_{j,n} := -r_{0,n} \sum_{l=1}^{j} w_{j,n} r_{j-l}, \quad j \in \mathbb{N}\,.$$

Then $\mathbb{E}[\xi^2], \mathbb{E}[\xi B_j]$ and $\mathbb{E}[B_j B_l]$ can be respectively approximated by

$$\tilde{\sigma}^2_{\lambda,n} := r_{n,0} - 1, \qquad \tilde{\sigma}^2_{\lambda,p_j,n} := \frac{1}{\lambda}\left(\tilde{p}_{j,n} - r_{j,n} - \tilde{p}_{j,n} r_{0,n}\right), \quad j \in \mathbb{N}$$

and, for $j, l \in \mathbb{N}$ with $1 \leq j \leq l$,

$$\tilde{\sigma}^2_{p_j,p_l,n} := \frac{1}{\lambda^2}\left(\tilde{p}_{j,n} r_{l,n} + \tilde{p}_{l,n} r_{j,n} + \tilde{p}_{j,n}\tilde{p}_{l,n} r_{0,n} - \tilde{p}_{j,n}\tilde{p}_{l,n} + \sum_{m=0}^{j} r_{m,n} r_{m+l-j,n} w_{j-m,n}\right). \tag{4.3.4}$$

Note that, even though the limiting random element here does coincide with ours and is therefore of Brownian bridge type, it is not forced to be so for finite samples. Therefore we can simulate paths of it analogously to above using Cholesky's factorisation method and recalling that we stop the path at the largest observation as suggested by Buchmann and Grübel [2003]. By the continuous mapping theorem, we can then construct the procedures of Section 4.2 using either the ℓ^1 norm for all the parameters jointly or the absolute value for each of them individually.

4.4 Simulated illustrations

In this section we focus on illustrating the behaviour and precision of the estimators of F in Buchmann and Grübel [2003] and Coca [2015], together with the behaviour and coverage of the corresponding confidence regions, all in different settings. We limit the study to these two estimators and confidence regions not only because they have not been implemented in the literature yet but also because the contain all the features we have mentioned so far in this thesis: among others, consequences of the efficiency and of the lack of it, and divergence of the series defining the estimators of the direct approach. In Section 4.4.1 we motivate the choice of the different inverse settings we consider, we provide some practical recommendations in the implementation of the estimators and, last, we explain the format in which we present the data. In Sections 4.4.2, 4.4.3 and 4.4.4 we analyse the results of the simulations in different settings. In Section 4.4.5 we make some final conclusions and remarks, and, based on these, in Section 4.4.6 we propose several topics that seem worth investigating further.

4.4.1 Preliminary remarks

Throughout the section we simulate the $\Delta = 1$ increments of a zero-drift compound Poisson process with intensity $\lambda > 0$ and jump distribution

$$F(x) = \frac{4}{5}\left(1 - e^{-10x}\right) \mathbb{1}_{[0,\infty)}(x) + \frac{1}{5}\mathbb{1}_{[\frac{1}{2},\infty)}(x), \qquad x \in \mathbb{R}\,. \tag{4.4.1}$$

I.e., F is the weighted sum of an exponential distribution with parameter 10 and an atomic distribution at $\frac{1}{2}$ with weights $\frac{4}{5}$ and $\frac{1}{5}$, respectively. This distribution is supported in $\mathbb{R}^+$, it contains both an absolutely continuous component and a discrete component, and the decay of the tail of its associated measure is exponential. Hence, it allows us to illustrate several aspects of the estimators of Buchmann and Grübel [2003] and Coca [2015], including the estimation they provide for each component and the need of an exponentially downweighted norm $\|\cdot\|_{\infty,\tau}, \tau > 0$, for the convergence of the former. We do not include more examples as, instead, we choose to analyse its properties in detail, thus giving deeper insight into the practical behaviour of the estimators.

4.4.1.1 Division of Section 4.4

Typically, in applications where compound Poisson processes are used, the number of discrete observations does not exceed several hundreds. However, existing literature on estimation of the jump distribution and density has only illustrated their results for several if not many thousands of observations. Here, we choose to run our simulations with more realistic sample sizes in order to add real practical value.

In practice, the difficulty of the estimation problem is driven by the intensity λ, which represents the expected number of jumps per unit of time or, equivalently due to $\Delta = 1$, per observation increment. Consequently, the first step is to determine for what values the precision of the estimators is acceptable when the sample size n is of a few hundreds. The first obvious division is between $\lambda < 1$ and $\lambda \geq 1$, as in the former we expect to observe the jumps directly in many occasions whilst in the latter we do not. Therefore, when $\lambda < 1$ we are approximately in the simple setting of $[n\lambda]$ i.i.d. observations of F and we can really check the small-sample behaviour by taking $n \in \{100, 400\}$ and $\lambda \in \{0.2, 0.5\}$. When $\lambda \geq 1$ the issue of the precision is more subtle and it is easier to start discussing it for the estimation of λ. Recall that for the example presented above in (4.4.1), all the estimators of λ introduced in Chapter 2 satisfy asymptotic normality with limiting variance $e^\lambda - 1$. Thus, the relative error of any estimator $\dot{\lambda}_n$ satisfies that

$$\frac{\dot{\lambda}_n - \lambda}{\lambda} \approx \sqrt{\frac{e^\lambda - 1}{\lambda}\frac{1}{n\lambda}} N(0,1), \tag{4.4.2}$$

and, for $n \in \{100, 250, 400, 1000, 4000\}$ and $\lambda \in \{0.2, 0.5, 1, 2, 3, 4, 5\}$, with 95% probability its absolute value is less or equal than the quantities in Table 4.1. We have included the

values $\lambda \in \{0.2, 0.5\}$ and $n = 4000$ for illustrative purposes and because we refer to them later in the section.

	$\lambda = 0.2$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$
$n = 100$	46.1	31.6	25.7	24.8	28.5	35.9	47.6
$n = 250$	29.2	20.0	16.2	15.7	18.1	22.7	30.1
$n = 400$	23.1	15.8	12.8	12.4	14.3	17.9	23.8
$n = 1000$	14.6	10.0	8.1	7.8	9.0	11.3	15.1
$n = 4000$	7.3	5.0	4.1	3.9	4.5	5.7	7.5

Table 4.1: Approximate 0.05-quantiles, expressed in %, of the absolute value of the relative errors in estimating λ with a sample of size n.

The first observation we make is that the (relative) precision is best around 2 and it deteriorates as λ departs from it. The improvement for small values of λ comes from the effective number of non-zero increments increasing with λ; note that to estimate the intensity these observations are just as important, if not more, as the zero-increments as they inform us of the length of the interarrival times. The behaviour for $\lambda > 2$ is expected because the larger the intensity, the less likely it is to have at least one zero-increment and hence the harder it is to estimate it. At $\lambda = 2$ the expected ratio of zero and non-zero increments is best for estimating the intensity. To attain the different precision values in the table we do not require any decompounding, which is intuitive because no perfect cancellations of jumps can arise in this example. The situation for estimating F is quite different as the probability of having at least one observation in the sample that is the sum of more than one jump is strictly positive for any $\lambda > 0$. Therefore we always have to decompound and it may seem that, as λ increases, the estimation problem gets harder. However, in the following sections and other examples not included here we observe a behaviour in the precision for estimating F similar to that for λ in Table 4.1. For estimation of F, the precision is highest around $\lambda = 1$, which really shows the ability of the estimators of Buchmann and Grübel [2003] and of Coca [2015] to decompound: decompounding when $\lambda \in (0, 1)$ is not too hard, since we expect most increments to carry at most one jump; when λ increases within that interval, we have more such observations and therefore the precision improves. As we see in Section 4.4.4, the empirical distribution function does not have this property as it does not have the ability to decompound. When $\lambda > 1$, however, the difficulty of decompounding dominates over the higher effective number of non-zero increments and the precision deteriorates. In short, a bit, but not too much, of inverse character improves the precision of the estimates, which to our knowledge has not yet been reported in the literature. We explore this further in Section 4.4.5.

From Table 4.1, it follows that if we set a reasonable limit of 20%, we should take $n \in \{250, 1000\}$ when $\lambda \in \{1, 2, 3\}$ and $n \in \{1000, 4000\}$ when $\lambda = 5$. The case $\lambda = 4$ is somewhere in between the two and we choose to include it in the former for reasons

having to do with the precision of the estimate for F that we mention in Section 4.4.5. This analysis gives rise to the division into $\lambda < 1$, $1 \leq \lambda \lesssim 5$ and $\lambda \gtrsim 5$ of Sections 4.4.2, 4.4.3 and 4.4.4, respectively. In view of Table 4.1, we are choosing a symmetric range in the precision of the estimators of λ. As we see below, this 'symmetry' does not extend to other properties of the estimators. The splitting at 5 is only an approximate one and depends on the complexity of the underlying F. Nonetheless, for these sample sizes, it also seems to approximately hold for other simulation experiments we have performed. The choice of $n = 4000$ when $\lambda = 5$ may seem to contradict our motivation to take moderate sample sizes. However, and as illustrated in Section 4.4.3, for such intensity large sample sizes are required for the estimation errors to vanish and start achieving some reasonable precision. Existing literature takes $\lambda \leq 3$, generally focusing on $\lambda \leq 1$, so Section 4.4.3 has interest in itself by showing the limit at which our estimators begin to require large sample sizes to work. Furthermore, and in view of Table 2.1 in Chapter 2, this choice (and the others) guarantees that the likelihood of not observing a zero-increment in a sample of size n is not considerable. This is required for the naive estimator of λ to return a non-trivial value and, indeed, all our simulations in subsequent sections only include samples with at least one zero-increment. We emphasise that this is not necessary for our estimators of Chapter 3 to work.

Lastly, we remark that, within each choice of λ, we have chosen two values of n, one being 4 times larger than the other. Therefore, the errors and confidence regions should decrease by half from one to the other. Yet, this does not always happen and it allows us to identify for what regimes the bias and stochastic errors in the estimators are not negligible giving further insight into them.

4.4.1.2 Remarks on the practical implementation

We first recall a couple of remarks made in previous sections that we take into account in our implementations and would like to emphasise.

In Section 4.3.1, we mentioned that the implementation of the estimator of N, and hence of F, can be safely based on its expression for the absolutely continuous case given in Chapter 3 and not on its modification to accommodate a potential discrete or absolutely continuous component. This is due to the discretisation of the integral appearing in it and hence can also be interpreted as the implementation of the distribution function when we postulate all jumps are discrete with potential support in $\epsilon_n \times \mathbb{Z} \setminus \{0\}$. In the figures of the following sections we observe that this strategy works perfectly well to estimate the absolutely continuous component and to identify the location of the atomic component without prior knowledge of their support, unlike assumed in the theoretical results of Chapter 3.

In theory, all the estimators make use of the distinguished logarithm but, as remarked in Chapter 2, for $\lambda\Delta = \lambda < \pi$ we can safely use the principal branch of the logarithm

if there is no drift in the process. This was justified by the upper bound on the complex part of the exponent in the characteristic function of the increments $\varphi = \varphi_1$ in (1.1.5). Therefore condition $\lambda < \pi$ is by no means necessary for the correct behaviour of the estimators implemented using the principal branch of the logarithm. Indeed, in all simulations with $\lambda \leq 5$ we observe that the only noticeable difference between using one or the other logarithm comes from the computational times, with the implementation of the distinguished logarithm being considerably more expensive. Although not included here for reasons mentioned above, we remark that if $\gamma \neq 0$ and/or λ is relatively larger than 5, the distinguished logarithm should be used to avoid well-known winding issues (cf. Embrechts et al. [1993] and Grübel and Hermesmeier [1999]). This does not depend on the sample size. Unlike depicted in our theoretical results of Chapter 3, in practice our estimators do not seem to be robust to drifts $|\gamma| > 0.05$. Therefore we suggest to first estimate γ with our estimator using the distinguished logarithm (or with the naive one depending on the prior knowledge of F as argued in Section 2.4.2) and to translate the observations accordingly before implementing the rest of the estimators.

At the end of Section 4.3.1 we discussed the choice of all the hyperparameters in the implementation of our estimators. All of these depend solely on the bandwidth h_n, which we mentioned that, as a rule of thumb, we take to be 0.0005 unless λ and n are small when we take $h_n = 0.001$. Indeed, we make the latter choice only in Section 4.4.4 and remark that in other examples it also seems to be the right choice for larger values of λ up to 2, or even 3. The implementation with this choice is generally around 4 times faster than that with $h_n = 0.0005$. This apparent rule of thumb is justified by minimising a purely data-driven loss function we found to work well across different examples and sample sizes. In order to motivate its form let us denote the implementation of the estimator $\check{\lambda}_n$ of Section 2.4 by $\check{\check{\lambda}}_n$, which follows using the same arguments of Section 4.3.1. Notice that, according to the results of Section 2.4, $\tilde{\lambda}_n$ is a consistent estimator of λ for the F we work with here. Therefore, $\check{\lambda}_n - \tilde{\lambda}_n$ is a sum of bias and stochastic errors that are negligible and is explicitly given in expression (2.4.1). In addition, the difference $\hat{\hat{\lambda}}_n - \check{\check{\lambda}}_n$, given in (2.4.2) is also a sum of such errors. Crucially, both terms are the errors that the estimator of F carries resulting from the truncations of the tails and around the origin and only depend on h_n. Hence, we propose to choose the bandwidth as

$$h_n := \operatorname{argmin}_{h>0} \left\{ \left| \hat{\hat{\lambda}}_n(h) - \check{\check{\lambda}}_n(h) \right| + \left| \check{\check{\lambda}}_n(h) - \tilde{\lambda}_n \right| \right\}. \tag{4.4.3}$$

This choice seems to minimise all the estimation errors, including the uniform distance between our estimator for F and the true distribution, and to optimise coverage of all the empirical confidence regions. As observed in the following sections, the naive estimator of the intensity is superior than the spectral estimators, especially when λ is large, which we attribute to the fewer error terms it carries. Thus, one can multiply the second summand

by a constant larger to 1 prior to the optimisation, although we do not explore this further. Let us remark that, if the practitioner does not have evidence of $\tilde{\lambda}_n$ being consistent in their setting, the second summand above can be removed from the loss function and the estimators obtained by using the minimiser of the resulting loss function still seem to work well.

We make a last remark on the implementation of the estimators of subsequent sections. After numerous simulations we have concluded that using $\check{\check{\lambda}}_n$ instead of $\hat{\hat{\lambda}}_n$, both to estimate λ and F, gives more accurate estimates and coverage of the confidence regions. The use of the former and the fact that the efficiency is kept when using it was justified in Section 2.6, so below we only make use of it. Despite being superior than $\hat{\hat{\lambda}}_n$, it still consistently underestimates λ and some times does not return satisfactory coverage for λ and for (λ, F) jointly. Therefore, in the following sections we have also included joint estimation of F (constructed with $\check{\check{\lambda}}_n$) and λ using the naive estimator, which returns more satisfactory values. We can do this because, as mentioned above, all the samples used in the simulations have at least one zero-increment and therefore $\tilde{\lambda}_n$ is always well-defined. Furthermore, using the naive estimator still retains the efficiency as argued in Section 2.6. If in practice one is confronted with a sample with no zero-increments, only the spectral estimators can be used. In such a case further investigations on improving the implementation of our estimators are worth considering prior to using them. We suggest to try to optimise the choices of the relationships between the hyperparameters made at the end of Section 4.3.1 through simulations with zero-increments using (4.4.3).

4.4.1.3 A guide to read the tables

In the following sections we introduce two types of tables: one to illustrate the estimation errors, and another one to show the empirical quantiles and the resulting coverage. For both types and in all sections we run 250 independent simulations and average out the results.

The entries of the first type of table correspond to the absolute error made by the estimator of the quantity of the corresponding row, measured in the corresponding norm. In particular, for the estimator of F of Buchmann and Grübel [2003] the error is measured in the norm $\|\cdot\|_{\infty,\tilde{\tau}_n}$, with $\tilde{\tau}_n$ as defined in Section 4.3.2, and in brackets we give the error measured in the standard $\|\cdot\|_\infty = \|\cdot\|_{\infty,0}$ norm. For the quantities τ and λ we can compute the relative error and this is given in brackets in % next to their absolute errors. In Section 4.4.4 the table includes a row with the estimation error made by the empirical distribution function, which we refer to as the naive estimator of F.

The entries of the second type of table represent the empirical $\alpha = 0.05$-quantiles of the appropriate norm of the limiting quantity in the central limit theorem for the parameter in each row. In brackets we include the coverage arising from them. We choose to construct confidence regions of level 0.05 for the whole process by first doing so for F,

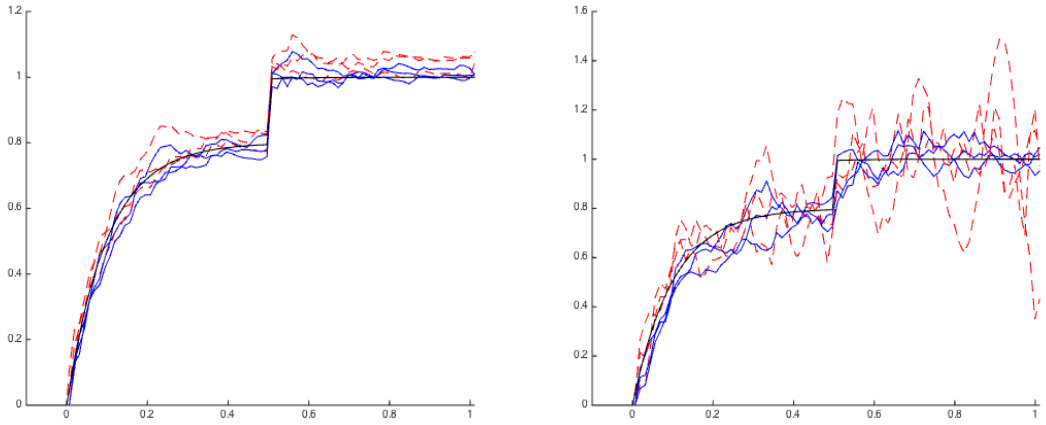
and then for λ conditioned on the value of F. We denote the latter by $\lambda|_F$ and therefore joint estimation is denoted by $F\&\lambda|_F$. The quantiles for $\lambda|_F$ generally return very low coverage as expected, since those of F already have level 0.05 by construction. The last entries in this second type of tables are classified as C+Naive and correspond to making joint estimation of F and λ when using the naive estimator for the latter as mentioned in the last section. In Section 4.4.4 we include a row for the empirical 0.05-quantiles arising from the empirical distribution function, again referred to as the naive estimator. We considered including joint estimation of it and λ but the coverage of the former is unsatisfactory and we therefore gain nothing by doing so.

The structure of the following sections is generally the same: first, we discuss the performance of the estimators by looking at the figures in order to gain intuition; we then analyse the results of the first type of tables, starting from those of F and then of λ; after this we analyse the second type of tables using the same order and then discuss joint coverage of the regions of the parameters; the sections end with a short paragraph summarising the conclusions.

4.4.2 Moderate intensity ($1 \leq \lambda \lesssim 5$)

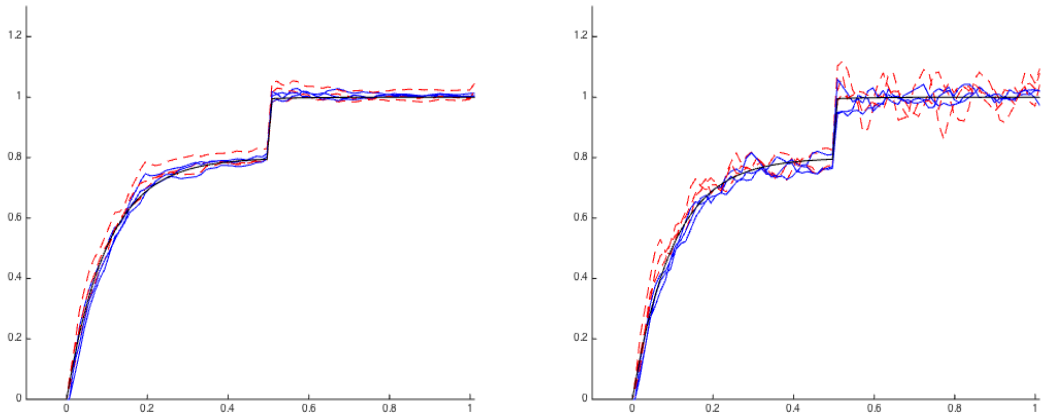
Figures 4.1 and 4.2 nicely illustrate most of the properties of the estimators. Those of Coca [2015] tend to mimic the shape of F satisfactorily, relative to the difficulty of the problem (driven by λ) and to the amount of information (driven by the sample size n). In particular, we remark that they are very satisfactory for $\lambda = 1$ and $n = 1000$ (Figure 4.2a), and they seem similar when increasing λ from 1 to 3 and compensating this additional difficulty by increasing n from 250 to 1000 (Figures 4.1a and 4.2b). For $\lambda = 3$ and $n = 250$ (Figure 4.1b) they deteriorate but still pick up the location and value of the atom in some cases. Looking at the estimators of Buchmann and Grübel [2003] we clearly observe the transition between values of λ for which the infinite series in them converges and when it does not relative to the sample sizes. In the former, the value of τ is so low that usual uniform convergence is guaranteed in the range of the abscissa shown. As a consequence, when $\lambda = 1$ the estimator also mimics F well and we clearly observe the central limit theorem around 1 on the right hand side of the graph anticipated in Section 2.6. The graphs for $\lambda = 3$ readily show the fluctuations arising from the alternating signs in the infinite sum in the estimator and the lack of convergence of it in the range shown. As expected, this already suggests the superiority of the estimators of Coca [2015] in any situation when it comes to estimation errors in the supremum norm.

Tables 4.2 and 4.3 show that the errors made by both estimators of F are halved as the sample size increases from $n = 250$ to $n = 1000$ when $\lambda \in \{1, 2\}$. Therefore, the bias and stochastic errors that they carry are already small in these regimes. As anticipated by the figures, the infinite sum in the estimator of Buchmann and Grübel [2003] starts giving



(a) $\lambda = 1$ **(b)** $\lambda = 3$

Figure 4.1: Three simulations of the estimators of Buchmann and Grübel [2003] (dashed red) and Coca [2015] (solid blue) for $n = 250$. True distribution in solid black.



(a) $\lambda = 1$ **(b)** $\lambda = 3$

Figure 4.2: Three simulations of the estimators of Buchmann and Grübel [2003] (dashed red) and Coca [2015] (solid blue) for $n = 1000$. True distribution in solid black.

large fluctuations for $\lambda = 3$, and for $\lambda = 4$ it breaks up. The errors of our estimator still approximately halve for these intensities and are not yet too large (note that under the standard supremum norm, relative and absolute errors are the same because $\|F\|_\infty = 1$ always). While for our estimator of F the errors seem to increase with λ, this is not the case for the estimator of Buchmann and Grübel [2003] when $\lambda = 2$. Note that, like for our estimator, its uniform error (in brackets) hardly increased when compared to that in $\lambda = 1$. Therefore the series seems to still be converging, and the increase in the optimal value of τ coming from the increase of the intensity makes the error in the exponentially downweighted norm smaller. For larger values of λ the series diverges and the norm $\|\cdot\|_{\infty,\tau}$

can no longer compensate for it. This parallels the behaviour of the relative errors of λ from Table 4.2 although does not have the same meaning: the relative errors of the estimator of Buchmann and Grübel [2003] correspond to the absolute errors divided by $\|F\|_{\infty,\tau} < \frac{\log 2}{\lambda}$; therefore, they actually grow as λ increases, just as in our estimator. More can be said about these relative errors but we postpone the analysis to Section 4.4.5. We remark that the optimal τ is always overestimated, which is good since the estimator of Buchmann and Grübel [2003] converges under the exponentially downweighted norm for any value larger than the optimal one.

		$n = 250$			
		$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$
B&G	$F(\ell^\infty)$	0.07 (0.10)	0.04 (0.11)	0.11 (0.35)	19.23 (6.96×10^6)
	$\tau(\%)$	1.47 (108)	2.35 (41)	4.20 (80)	2.00 (40)
C	F	0.08	0.10	0.13	0.21
	$\lambda(\%)$	0.01 (-1.09)	0.03 (-1.57)	0.24 (-7.87)	0.90 (-22.5)
Naive	$\lambda(\%)$	0.01 (1.14)	0.02 (0.97)	0.08 (2.71)	0.14 (3.47)

Table 4.2: Average empirical errors, $\|\theta_n - \theta\|_\theta \left(\frac{\theta_n - \theta}{\theta}\%\right)$, of the estimators of Buchmann and Grübel [2003] (B&G), Coca [2015] (C) and naive ones (N) after 250 simulations with at least one zero-increment.

		$n = 1000$			
		$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$
B&G	$F(\ell^\infty)$	0.03 (0.05)	0.02 (0.06)	0.04 (0.10)	1.18 (1.78)
	$\tau(\%)$	1.30 (95)	2.00 (35)	2.48 (58)	1.13 (-6.9)
C	F	0.04	0.05	0.08	0.12
	$\lambda(\%)$	0.02 (-1.89)	0.03 (-1.36)	0.08 (-2.63)	0.46 (-11.6)
Naive	$\lambda(\%)$	0.00 (0.01)	0.01 (0.39)	0.01 (0.31)	0.01 (0.25)

Table 4.3: Average empirical errors, $\|\theta_n - \theta\|_\theta \left(\frac{\theta_n - \theta}{\theta}\%\right)$, of the estimators of Buchmann and Grübel [2003] (B&G), Coca [2015] (C) and naive ones (N) after 250 simulations with at least one zero-increment.

Note that the absolute errors made by any estimator of λ increase with λ for any n and, as anticipated in the previous section, the spectral estimator underestimates and the naive one is superior especially as the intensity grows. The relative errors of the estimators of λ show different behaviours. For the naive, they agree with the behaviour predicted by Table 4.2 when $n = 250$, and for the spectral one they do when $n = 1000$. The disagreement in the other two cases can be attributed to two reasons: to computational errors in the cases when the absolute errors are very small; and, in the cases when the errors are not halved when increasing n from 250 to 1000, to the inherent randomness of the experiments. We do not attribute the disagreement of the spectral estimator to the bias and stochastic

errors mentioned in the previous section because in the hardest case when $\lambda = 4$, the errors are halved.

Moving on to Tables 4.4 and 4.5, we observe that the value of the average empirical quantiles coming from the estimators of F behave like the absolute errors made by them, which we attribute to the same reasons mentioned above. The coverage of the resulting confidence regions for $\lambda < 4$ is slightly lower than expected, with our procedures returning better values for high sample sizes, and for $\lambda = 4$ we observe an interesting difference: the coverage of our procedures deteriorates noticeably, which mainly comes from the quantiles being slightly underestimated, whilst that arising from the estimators of Buchmann and Grübel [2003] does not. The latter may have two meanings: the errors from the infinite series in the quantiles are too large and hence coverage of the truth is guaranteed in most cases; or, more interestingly, despite the lack of efficiency of the estimator of Buchmann and Grübel [2003], the confidence regions coming from it may behave better than ours in practice and be more accurate. The second would be very important for testing purposes. Our understanding, reinforced by the results of the next section, is that the first is the reason for the null coverage when $n = 250$, whilst the coverage from the estimator of Buchmann and Grübel [2003] is actually reasonable for large enough samples as observed when $n = 1000$.

		$n = 250$			
		$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$
B&G	F	0.13 (0.8)	0.09 (1.6)	1.75×10^6 (1.2)	$3.1{\times}10^{19}$ (0)
C	F	0.14 (1.2)	0.17 (0.8)	0.23 (2.4)	0.24 (18.4)
	λ	0.15 (10)	0.26 (8)	0.30 (35.6)	0.32 (100)
	$\lambda\|_F$	0.33 (1.2)	1.04 (0)	1.31 (0.0)	1.62 (10.8)
	$F \,\&\, \lambda\|_F$	(2.4)	(0.8)	(2.4)	(25.6)
Naive	λ	0.14 (10.8)	0.26 (7.2)	0.48 (7.6)	0.84 (3.2)
C+Naive	$\lambda\|_F$	0.33 (0.8)	1.01 (0)	1.12 (1.6)	1.32 (0)
	$F \,\&\, \lambda\|_F$	(1.8)	(0.8)	(2.4)	(18.4)

Table 4.4: Average empirical 0.05-quantiles (coverage in %) from the estimators of Buchmann and Grübel [2003] (B&G), Coca [2015] (C) and naive ones (N) under the respective norms and after 250 simulations with at least one zero-increment.

The average empirical quantiles coming from the estimators of λ increase with this parameter. This agrees with Table 4.1 which, rescaled by the value of the intensity, is given in Table 4.6. Indeed, the average empirical quantiles from the naive estimator included in Tables 4.4 and 4.5 are close to those in Table 4.6, although they tend to be underestimated resulting in higher coverage than expected. Surprisingly, the underestimation is more

		$n = 1000$			
		$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$
B&G	F	0.06 (1.2)	0.04 (1.6)	1.29×10^9 (2.0)	$1.24{\times}10^{16}$ (0.8)
C	F	0.07 (2.8)	0.08 (2.8)	0.12(1.2)	0.14 (15.6)
	λ	0.07 (9.6)	0.13 (18.8)	0.19 (18.4)	0.23 (99.2)
	$\lambda\|_F$	0.17 (0.4)	0.51 (0)	0.85 (0.0)	1.01 (0)
	$F \,\&\, \lambda\|_F$	(3.0)	(2.8)	(1.2)	(15.6)
Naive	λ	0.07 (8)	0.13 (15.6)	0.23 (13.2)	0.38 (12)
C+Naive	$\lambda\|_F$	0.17 (0.4)	0.50 (0.0)	0.81 (0.0)	0.83 (0.4)
	$F \,\&\, \lambda\|_F$	(3.0)	(2.8)	(1.2)	(15.6)

Table 4.5: Average empirical 0.05-quantiles (coverage in %) from the estimators of Buchmann and Grübel [2003] (B&G), Coca [2015] (C) and naive ones (N) under the respective norms and after 250 simulations with at least one zero-increment.

noticeable for $\lambda \in \{2, 3\}$ than for $\lambda = 4$. The average empirical quantiles from the spectral estimator agree with naive ones for $\lambda = 1$ and $\lambda = 2$, and then do not grow fast enough with λ resulting in terrible coverage. We attribute this to the bias and stochastic errors they carry, since average empirical quantiles do not halve with n neither for $\lambda = 3$ nor for $\lambda = 4$. Therefore, we again observe the superiority of the naive estimator, especially for large intensity and sample size.

	$\lambda = 0.2$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$
$n = 100$	0.09	0.16	0.26	0.50	0.86	1.43	2.38
$n = 250$	0.06	0.10	0.16	0.31	0.54	0.91	1.51
$n = 400$	0.05	0.08	0.13	0.25	0.43	0.72	1.19
$n = 1000$	0.03	0.05	0.08	0.16	0.27	0.45	0.75
$n = 4000$	0.01	0.02	0.04	0.08	0.14	0.23	0.38

Table 4.6: Approximate 0.05-quantiles of the absolute errors made to estimate λ with a sample of size n.

In general, we observe the intuitive and abovementioned large values for the quantiles arising from any estimator of $\lambda\,|_F$. The only exception comes for the spectral estimator when $\lambda = 4$ and $n = 250$ because they fail to grow fast enough with λ. Then, joint coverage of F and λ tends to follow that of F, and the superiority of the use of the naive estimator of the intensity is only observed when $\lambda = 4$ and $n = 250$. For $\lambda = 1$ the use of the spectral estimator returns better joint coverage, but this is not conclusive of it being necessarily better.

Summarising, we have made several conclusions: for low intensities, the direct and spectral estimators of F and λ behave similarly, so the latter should be preferred for efficiency reasons; for high intensities, our estimator of F is superior to that of Buchmann and Grübel [2003] for estimation, whilst for confidence regions the latter is preferable;

for these intensities the naive estimator of λ is far superior than the spectral one for all purposes. When λ is high, joint confidence regions of F and λ is best using our estimator for the former but it is not satisfactory enough to suggest to use it in practice.

4.4.3 High intensity ($\lambda \gtrsim 5$)

Figure 4.3 confirms the intuition built up so far: our estimator of F still performs relatively well for large sample sizes, identifying the location of the atomic component. Just as depicted by Figure 4.1b, the size of some samples is not large enough to be able decompound well and the estimate is somewhere in between the truth and the distribution of the observations. The latter must tend to a normal distribution with large variance in view of the classical central limit theorem. For large intensity and small sample size, the infinite series in the estimator of F of Buchmann and Grübel [2003] does not converge and the resulting fluctuations are even more pronounced.

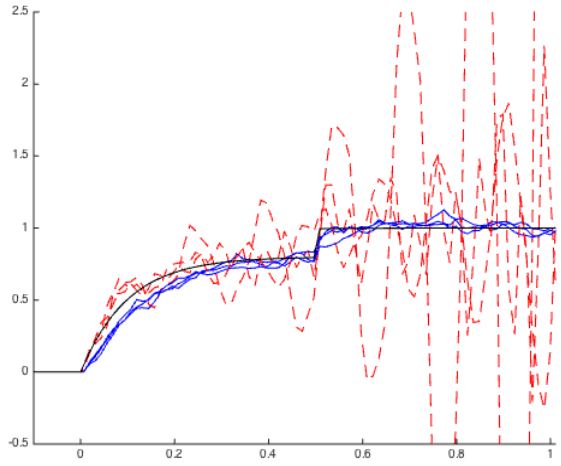


Figure 4.3: Three simulations of the estimators of Buchmann and Grübel [2003] (dashed red) and Coca [2015] (solid blue) for $\lambda = 5$ and $n = 4000$. True distribution in solid black.

The impact of the large fluctuations in the estimator of F in Buchmann and Grübel [2003] is reflected in Table 4.7 through its estimation errors in the uniform norm and through the large values of the estimates of the optimal τ. The average error made by our estimator of F becomes acceptable for $n = 4000$ and it is similar to its error when $\lambda = 4$ and $n = 1000$ (Table 4.3). This, together with the growth of the errors from the last section, allows us to conclude an exponential increase in the errors for large and increasing λ. This is also observed in the errors of the spectral estimator of λ in line with Tables 4.1 and 4.6. Instead, the naive estimator is highly robust to increasing intensity (recall that we only display the results for samples with at least one zero-increment, whose probability was included in Table 2.1), concluding the superiority of it for estimation.

		$n = 1000$	$n = 4000$
		$\lambda = 5$	$\lambda = 5$
B&G	$F(\ell^\infty)$	0.03 (2.27×10^{10})	0.02 (5.23)
	$\tau(\%)$	15.8 (114)	10.8 (87.8)
C	F	0.22	0.13
	$\lambda(\%)$	1.26 (-25.2)	0.76 (-15.2)
Naive	$\lambda(\%)$	0.08 (1.61)	0.01 (0.19)

Table 4.7: Average empirical errors, $\|\theta_n - \theta\|_\theta \left(\frac{\theta_n - \theta}{\theta}\%\right)$, of the estimators of Buchmann and Grübel [2003] (B&G), Coca [2015] (C) and naive ones (N) after 250 simulations with at least one zero-increment.

The conclusions following from Table 4.8 agree with those of the previous section: the spectral estimators of F and λ return regions that are too tight resulting in even worse coverage (this includes the quantiles of $\lambda|_F$, which should have coverage close to zero); the estimator of F of Buchmann and Grübel [2003] requires enough observations to reduce the errors and then it again returns quantiles that are slightly lower than expected but not too high or too low; the naive estimator of the intensity slightly underestimates the quantiles resulting in realistic but slightly high coverage; and the joint regions for F and λ using our estimator of F return unacceptable high coverage regardless of the estimator of the intensity due to the bad coverage of the former estimator.

		$n = 1000$	$n = 4000$
		$\lambda = 5$	$\lambda = 5$
B&G	F	2.11×10^{15} (0.0)	0.03 (1.6)
C	F	0.22 (76.0)	0.13 (63.2)
	λ	0.31 (99.6)	0.21 (99.6)
	$\lambda\|_F$	1.99 (52.0)	1.18 (16.0)
	$F \,\&\, \lambda\|_F$	(80.0)	(66.0)
Naive	λ	0.67 (7.6)	0.32 (11.2)
C+Naive	$\lambda\|_F$	1.48 (0.8)	0.95 (0.8)
	$F \,\&\, \lambda\|_F$	(76.0)	(63.4)

Table 4.8: Average empirical 0.05-quantiles (coverage in %) from the estimators of Buchmann and Grübel [2003] (B&G), Coca [2015] (C) and naive ones (N) under the respective norms and after 250 simulations with at least one zero-increment.

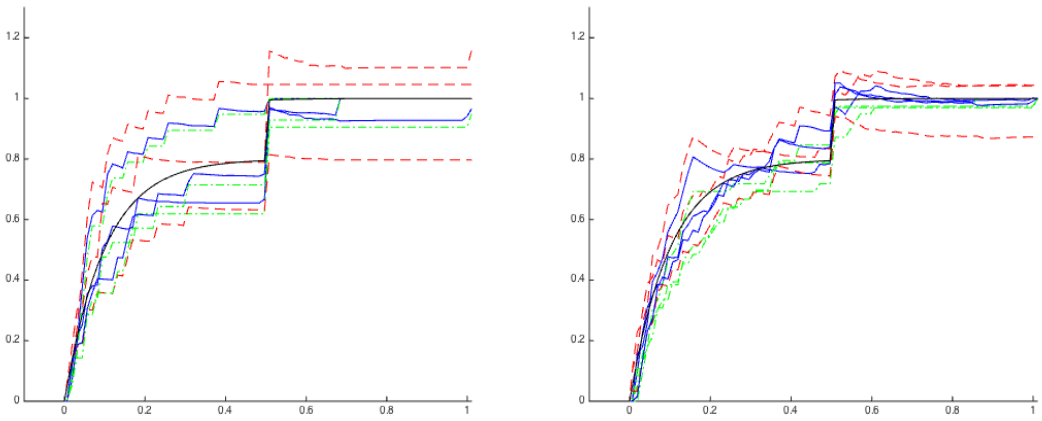
To summarise, the results here confirm the conclusions made in the previous section regarding high intensities: for estimation of F one should use the spectral estimator, which makes acceptable errors for large samples; for confidence regions it should be discarded in favour of the estimator of Buchmann and Grübel [2003] as long as the sample size is large enough, when it returns slightly low coverage; the naive estimator of λ should be used for both estimation and confidence regions, and it produces very accurate results

in the former and slightly high coverage in the latter. Joint confidence regions should be constructed in a different way and we discuss this in Section 4.4.5.

4.4.4 Low intensity ($\lambda < 1$)

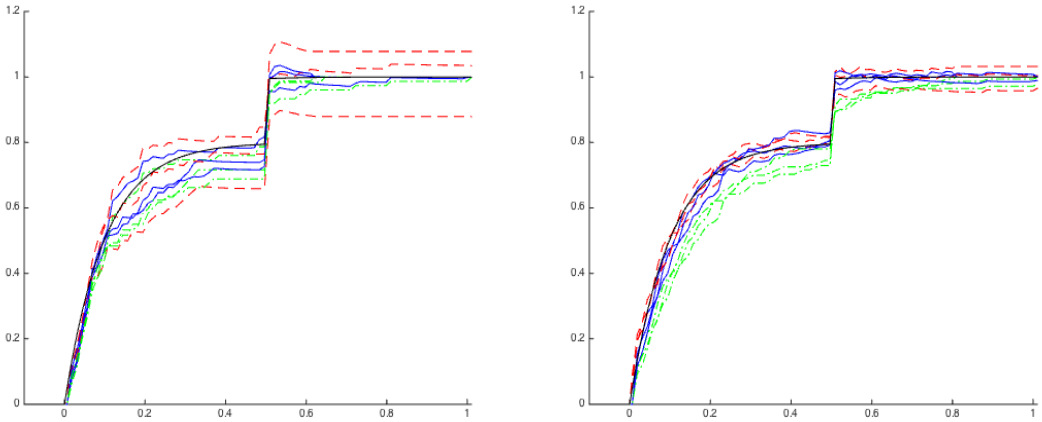
We now turn to the opposite regime than the one considered in the last section and extend the conclusions of Section 4.4.2 to low intensities. As mentioned in Section 4.4.1, here we allow the sample size to be quite small to test this relevant-in-practice regime. We include simulations of the empirical distribution function of the observations to see when it is and when it is not a good approximation.

Figures 4.4 and 4.5 clearly illustrate several features of the estimators of F for low intensities. We first note that no fluctuations in the estimator of Buchmann and Grübel [2003] are observed unlike in the previous two sections. This is because both values of the intensity considered here are smaller than $\log 2$, so the infinite series in it always converges under the standard supremum norm. This, together with the lower sample sizes considered here, means the central limit theorem of its limit value around 1 becomes clearer at this regime. Except for this feature, the three estimators show similar behaviours for $\lambda = 0.2$ (Figures 4.4a and 4.5a) because most jumps are directly observed: they pick up the atom and are almost monotone, and the estimation is not too good for $n = 100$ (Figure 4.4a) because the effective sample size (approximately $n\lambda$) is around 20. For $n = 400$ (Figure 4.5a), we more clearly see that the naive estimator performs similarly to ours although it slightly under estimates F as a consequence of the few observations that carry more than one jump and its lack of ability to decompound. When the intensity increases to $\lambda = 0.5$ (Figures 4.4b and 4.5b) all estimators accurately pick up the location of the atom. Nevertheless, here we start to see the departure of the estimators with the ability to decompound from the naive one: some decompounding is needed now, which means the terms with alternating signs in the expansion of the logarithm in-built in the decompounding estimators are no longer negligible. This is clearly reflected in Figure 4.4b, where we observe the loss of their approximate monotonicity when compared to Figure 4.4a, even though the effective number of observations in the former, 50, is a 150% larger than that in the latter, 20. In Figure 4.5b, the decompounding estimators are closer to being monotone due to the effective sample size being 200 and hence being better estimates of the true distribution. The lack of ability to decompound of the naive estimator becomes more visible now: it is estimating the distribution of the observations, which for $\lambda = 0.5$ is already visibly distinct from the underlying jump distribution. In particular, it has larger variance because the observations are (random) sums of independent variables. Consequently, in this regime we explicitly observe the decompounding power of the estimators in Buchmann and Grübel [2003] and Coca [2015], as their accuracy improves with the intensity, unlike in the previous sections. In all cases, our estimator is the one showing the best performance.



(a) $\lambda = 0.2$ **(b)** $\lambda = 0.5$

Figure 4.4: Three simulations of the estimators of Buchmann and Grübel [2003] (dashed red), Coca [2015] (solid blue) and the naive one (dashed-dotted green) for $n = 100$. True distribution in solid black.



(a) $\lambda = 0.2$ **(b)** $\lambda = 0.5$

Figure 4.5: Three simulations of the estimators of Buchmann and Grübel [2003] (dashed red), Coca [2015] (solid blue) and the naive one (dashed-dotted green) for $n = 400$. True distribution in solid black.

Table 4.9 confirms the intuition gathered in the previous paragraph: the errors made by our estimator are the smallest in all cases; the errors of the decompounding estimators of F halve with the sample size because the problem is not too complicated, so they decompound well and the bias and stochastic errors they carry are negligible; and the errors made by the naive estimator do not halve, especially when $\lambda = 0.5$, because the distribution of the observations does not coincide with that of the underlying jumps. We do remark that, when $\lambda = 0.2$, its errors are not much larger than those made by our

estimator. Hence, when F is such that the approximations of Section 4.2.5 apply, these must be accurate for smaller intensities. As expected, the estimates of the optimal τ are zero or approximately zero, which is reflected in the norm of the estimator of the distribution of Buchmann and Grübel [2003]. This also means that the errors made by their estimator of F are relative errors, just as the errors of ours. Therefore, and due to the larger effective sample size, they decrease as λ increases as anticipated by Table 4.1. This contrasts with the errors of the previous sections, which increased with the intensity; in Section 4.4.5 we discuss the value of λ at which these may be optimal. The spectral estimator of λ shows the increase (decrease) in the absolute (relative) errors with λ anticipated by Table 4.6 (Table 4.1). Yet, and also in view of that table, the absolute errors it makes are unexpectedly high. This, together with the fact that they do not halve with the increase in n and the good behaviour of the spectral estimator of F, suggests the presence of errors. It also means that we again observe the superiority of the naive estimator of λ.

		$n = 100$		$n = 400$	
		$\lambda = 0.2$	$\lambda = 0.5$	$\lambda = 0.2$	$\lambda = 0.5$
B&G	$F(\ell^\infty)$	0.235 (0.235)	0.172 (0.178)	0.123 (0.123)	0.089 (0.089)
	$\tau(\%)$	0.000	0.102	0.000	0.000
C	F	0.180	0.139	0.092	0.075
	$\lambda(\%)$	0.010 (-5.2)	0.016 (-3.24)	0.009 (-4.25)	0.018 (-3.64)
Naive	F	0.193	0.172	0.111	0.136
	$\lambda(\%)$	0.004 (2.01)	0.011 (2.23)	0.000 (-0.09)	0.002 (0.38)

Table 4.9: Average empirical errors, $\|\theta_n - \theta\|_\theta \left(\frac{\theta_n - \theta}{\theta}\%\right)$, of the estimators of Buchmann and Grübel [2003] (B&G), Coca [2015] (C) and naive ones (N) after 250 simulations.

The average empirical quantiles in Table 4.10 go in line with everything mentioned so far. Those arising from the decompounding estimators of F return slightly less coverage than expected whilst the coverage of those of the naive estimator worsens with λ and n. The latter is of course a consequence of the higher effective size of the sample, which results in tighter regions around the distribution of the observations and not around the jump distribution. We again observe the slightly higher coverage of the confidence intervals for the intensity, especially for the naive estimator. However, and despite the smaller number of effective observations compared to that of other sections, the joint coverage of the estimators seems to be best for this regime. That arising from the naive estimator of λ remains superior. We remark that, just as with the errors in the previous table, the quantiles coming from the estimators of the distribution decrease with λ and n as the coverage does not change dramatically, which is a consequence of the higher number of observations and the ability of the estimators to decompound.

We make several conclusions: as in previous sections, our estimator of the distribution is superior than that of Buchmann and Grübel [2003] for estimation and, unlike in previous sections, for confidence regions too. Crucially, these improve with λ unlike in other sections, as a consequence of the effective increase in the sample size due to their ability to decompound. The empirical distribution function of the observations should be avoided for estimation and confidence regions unless the intensity is so small that the number of observations with more than one jump is negligible. The naive estimator of the intensity is superior to the spectral one, especially for estimation and joint coverage.

		$n = 100$		$n = 400$	
		$\lambda = 0.2$	$\lambda = 0.5$	$\lambda = 0.2$	$\lambda = 0.5$
B&G	F	0.627 (0.4)	0.364 (0.8)	0.326 (0.4)	0.188 (0.4)
C	F	0.374 (0.8)	0.266 (0.8)	0.182 (0.0)	0.128 (3.2)
	λ	0.119 (0.8)	0.159 (5.2)	0.060 (2.0)	0.079 (8.4)
	$\lambda\|_F$	0.142 (0.0)	0.193 (2.8)	0.071 (0.8)	0.100 (3.2)
	F & $\lambda\|_F$	(0.8)	(3.6)	(0.8)	(6.4)
Naive	F	0.299 (8.0)	0.206 (26.4)	0.155 (15.2)	0.107 (85.6)
	λ	0.078 (7.6)	0.134 (10.8)	0.039 (7.6)	0.066 (11.2)
C+Naive	$\lambda\|_F$	0.095 (3.2)	0.047 (5.6)	0.166 (4.4)	0.092 (3.2)
	F & $\lambda\|_F$	(3.2)	(5.6)	(5.2)	(3.6)

Table 4.10: Average empirical 0.05-quantiles (coverage in %) from the estimators of Buchmann and Grübel [2003] (B&G), Coca [2015] (C) and naive ones (N) after 250 simulations.

4.4.5 Conclusions and remarks

The detailed analysis of the results in the preceding sections allows us to conclude that, in practice, inference on discretely observed compound Poisson processes is possible for moderate sample sizes. This had not been remarked in existing literature and is relevant for many applications.

In particular, for estimation purposes our estimator of the distribution function is superior than that of Buchmann and Grübel [2003] and we hence start by summarising its behaviour. For a fixed sample size n, its precision increases approximately linearly with the intensity up to $\lambda = 1$ for the following reasons: for such low intensities decompounding is not hard; the expected total number of jumps in the observation interval is $n\lambda$; therefore, an increase in λ translates into an approximately equal increase in the number of jumps used to estimate the jump distribution. For larger intensities, decompounding is harder and cannot be achieved if the sample size is not increased alongside λ, as expected from the classical central limit theorem. We observed that the precision of the estimator is best around $\lambda = 1$ and starts being low at around $\lambda = 4$; for $\lambda = 5$ we had to leave the moderate sample-size assumption for the accuracy to be acceptable. Even so, the errors in this case

were larger than those for $\lambda = 4$ after increasing the sample size by 4. In other examples in which the jump measure has a density, these conclusions seem to extend a bit further to approximately $\lambda = 6$. In any case, we see that, for fixed n, the precision decreases exponentially with the intensity. This goes in line with Table 4.1, whose entries are driven by the function $\lambda\frac{\lambda}{e^{\lambda}-1}$. One can show that a slower decrease would contradict the results of Duval [2014] and that, furthermore, it cannot decay faster than $e^{-2\lambda}$. The precision of estimating F with a sample of size n and intensity λ is therefore approximately driven by the expression $\lambda e^{-\lambda}$, which peaks at $\lambda = 1$ (we also observed this peak in other examples). As far as we know, no remark has been made in this direction in existing literature, not even on the improvement in the precision when λ grows within $(0, 1)$.

This insight has a further consequence for practical purposes: as mentioned in Chapter 1, in some applications observing the stochastic process at hand continuously in time is too expensive and instead it is observed at discrete times. In Section 4.4 we have taken $\Delta = 1$ and let the intensity λ vary in $\mathbb{R}^+$ playing the role of $\lambda\Delta$ from preceding sections. However, in these applications $\lambda > 0$ is fixed and $\Delta > 0$ can be varied. Therefore, to extract the maximum information from the discrete observations and reduce observation costs the statistician should choose $\Delta \approx \lambda^{-1}$. This can be achieved by observing the process discretely for several values of $\Delta > 0$, estimating λ as above and then fixing Δ accordingly to estimate the jump distribution function F.

The estimator of the distribution function of Buchmann and Grübel [2003] is also best around $\lambda = 1$. However, for other intensities its accuracy decreases dramatically due to its lack of efficiency (lower intensities) and to the divergence of the series in its definition (larger intensities). Nevertheless, the coverage of its confidence regions is far superior than ours, especially for large λ and n. Therefore, if used for testing purposes we suggest to input the naive estimator of λ into its expression as mentioned in Section 2.6 to decrease the covariance function and thus increase the statistical power. For low intensities, the resulting estimates are very similar to ours (recall from Chapter 3 that, in this regime, the only practical difference is regularisation of the kernel and as we see from the figures in Section 4.4.4, this is not noticeable) so in these cases it also improves its estimation accuracy. Nevertheless, for large intensities the accuracy of the modified estimator tends to deteriorate due to the overestimation of the naive estimator of λ we typically observe in simulations. From our experiments, we also conclude that this estimator of the intensity is much better than ours, both for estimation and confidence regions, and especially for large intensities. However, for completeness we recall its main two drawbacks: when there are no zero observations it cannot be used, unlike our estimator; and in more general cases it may not even be consistent (cf. Section 2.4). Neither of these are relevant in the simulations we perform because of our choice of F and because we only included samples with at least one zero-increment. We emphasise that it can be used in conjunction with our estimator of F to construct joint confidence regions. These give slightly high but still

reasonable coverage for intensities up to approximately 3 but not beyond it. This is a consequence of the unacceptably low coverage of the distribution function arising from the use of our estimator. Therefore, and due to the better behaviour of the estimator of the distribution function of Buchmann and Grübel [2003] for this purpose, for testing purposes we propose to use them in conjunction after making the appropriate modification of the latter mentioned a few lines ago. Their dependence should be well approximated by that of ours since all these estimators are asymptotically equivalent. An alternative is to use the discrete estimator in Buchmann and Grübel [2003], whose implementation we included in Section 4.3.2, and which, in practice, is equivalent to what we just proposed and computationally faster. We ignore the effects of using the transformed estimators in Buchmann and Grübel [2004] but recall that the authors could not show tightness of them. Ignoring this crucial point, we remark that for low intensities they should guarantee monotonicity of the estimates but for large intensities is not clear to us that they will work due to the large fluctuations of the untransformed estimator.

Lastly, we make two more remarks that we have omitted making in previous sections. For low and high intensities the pointwise variances of the limiting process in the functional central limit theorem of Buchmann and Grübel [2003] are considerably low. This results in covariance matrices with very small determinants and Matlab's Cholesky algorithm used to simulate paths of the limiting process fails in relatively many occasions. However, and surprisingly, it does not return errors when inverting these matrices. Hence, a way to avoid using Cholesky's algorithm is to first sample the 'end point' of the path and then use the Brownian bridge construction introduced in Section 4.3.1. The last remark we make is on computational times. The main factor driving the time to compute the estimator of F of Buchmann and Grübel [2003] and of the confidence regions arising from it is the sample size n. In contrast, the computational time of our respective procedures does not strongly depend on n but on other factors, including the choice of the bandwidth. For moderate sample size values such as those considered in Sections 4.4.2 and 4.4.4, the former procedures are faster than ours and start being slower for $n = 4000$ considered in Section 4.4.3.

4.4.6 Further investigations

The simulation experiments we carried in the preceding sections do not cover all natural applied questions following from our theoretical results. Nonetheless, they provide a deep understanding of the practical behaviour of existing estimators in realistic situations and therefore they are useful in that they provide a not-yet-existent solid basis to perform further investigations. Here we mention some potential directions that directly arise from our experiments.

In a few cases mentioned in the previous section, our estimators do not seem to return satisfactory results. In particular, the spectral estimator of the intensity consistently

underestimates the true λ. We have proposed alternatives based on the estimators in Buchmann and Grübel [2003] but, in cases such as when the distribution is not only supported in $\mathbb{R}^+$, it is not possible to use these. In these cases, the implementation of our estimators should be improved; in Section 4.4.1 we suggested to use the purely data-driven loss function proposed therein to achieve this and should be investigated further if need be.

In the simulations we included the coverage of the confidence regions arising from the existing estimators. This allows to construct tests for the defining parameters as mentioned in Section 4.2. It would be interesting to investigate whether they work best for some particular alternatives and, if so, for which and for what reasons. In this direction, it would also be interesting to investigate the behaviour of the novel test for the presence of an atomic or a continuous distribution function we also constructed in Section 4.2 using the results of Chapter 3.

More interesting topics can be explored too but are beyond the scope of our analysis. These include studying the robustness of our estimators with respect to both the interarrival times of the process and to the jumps. Hence, we can rephrase this as investigating the robustness of the estimators to other renewal and Lévy processes. Related to the latter, and due to the truncation at the origin of our estimators, one can also try to use them to develop tests of whether the Lévy process at hand has finite or infinite activity. Additionally, it would be exciting and relevant to use our estimators in a real data study for an application in which the compound Poisson model is reasonable. This includes migration studies as mentioned in Chapter 1 and many other applications such as those mentioned in Buchmann and Grübel [2004].

Bibliography

N. Antunes and V. Pipiras. Probabilistic sampling of finite renewal processes. *Bernoulli*, 17(4):1285–1326, 2011. (Cited on page 26).

S. Asmussen. *Applied Probability and Queues*. Springer Science & Business Media, 2008. (Cited on page 20).

M. Bec and C. Lacour. Adaptive pointwise estimation for pure jump Lévy processes. *Statistical Inference for Stochastic Processes*, 18(3):229–256, 2015. (Cited on page 38).

D. Belomestny and M. Reiß. Spectral calibration of exponential Lévy models. *Finance and Stochastics*, 10(4):449–474, 2006. (Cited on pages 15, 29, 38, and 108).

D. Belomestny, F. Comte, V. Genon-Catalot, H. Masuda, and M. Reiß. *Lévy Matters IV*. Springer, 2015. (Cited on pages 38 and 49).

N. Bingham and R. Kiesel. *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*. Springer Finance Textbooks. Springer, 2004. (Cited on page 4).

M. Bøgsted and S. M. Pitts. Decompounding random sums: a nonparametric approach. *Annals of the Institute of Statistical Mathematics*, 62(5):855–872, 2010. (Cited on pages 20, 21, 22, 23, 24, 26, 40, 111, 112, and 125).

B. Böttcher, R. Schilling, and J. Wang. *Lévy matters III: Lévy-type processes: construction, approximation and sample path properties*. Springer, 2014. (Cited on page 29).

B. Buchmann. Weighted empirical processes in the nonparametric inference for Lévy processes. *Mathematical Methods of Statistics*, 18(4):281–309, 2009. (Cited on page 33).

B. Buchmann and R. Grübel. Decompounding: an estimation problem for Poisson random sums. *Annals of Statistics*, 31(4):1054–1074, 2003. (Cited on pages 14, 15, 16, 17, 18, 20, 22, 23, 24, 26, 27, 33, 37, 40, 41, 42, 50, 53, 55, 56, 57, 58, 59, 107, 108, 111, 112, 117, 123, 124, 125, 126, 127, 128, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, and 145).

B. Buchmann and R. Grübel. Decompounding Poisson random sums: recursively truncated estimates in the discrete case. *Annals of the Institute of Statistical Mathematics*, 56(4):743–756, 2004. (Cited on pages 16, 20, 27, 42, 50, 59, 107, 108, 111, 123, 125, 144, and 145).

F. P. Cantelli. Sulla determinazione empirica delle leggi di probabilitá. *Giornale dell' Istituto Italiano degli Attuari*, 4:421–424, 1933. (Cited on page 8).

I. Castillo and R. Nickl. Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *The Annals of Statistics*, 41(4):1999–2028, 2013. (Cited on page 60).

I. Castillo and R. Nickl. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *The Annals of Statistics*, 42(5):1941–1969, 2014. (Cited on page 60).

S. Chen, A. Delaigle, and P. Hall. Nonparametric estimation for a class of Lévy processes. *Journal of Econometrics*, 157(2):257–271, 2010. (Cited on pages 30 and 38).

C. Chesneau, F. Comte, and F. Navarro. Fast nonparametric estimation for convolutions of densities. *Canadian Journal of Statistics*, 41(4):617–636, 2013. (Cited on page 25).

K. L. Chung. *A Course in Probability Theory*. Academic press, 2001. (Cited on pages 34 and 75).

A. J. Coca. Efficient nonparametric inference for discretely observed compound Poisson processes. *Under revision, arXiv preprint arXiv:1512.08472*, 2015. (Cited on pages 14, 15, 16, 17, 18, 20, 35, 36, 41, 42, 43, 47, 50, 51, 53, 58, 59, 61, 63, 64, 66, 107, 117, 126, 127, 128, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, and 142).

F. Comte and V. Genon-Catalot. Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and their Applications*, 119(12): 4088–4123, 2009. (Cited on page 38).

F. Comte and V. Genon-Catalot. Nonparametric adaptive estimation for pure jump Lévy processes. *Annales de l'institut Henri Poincaré (B)*, 46(3):595–617, 2010a. (Cited on page 38).

F. Comte and V. Genon-Catalot. Non-parametric estimation for pure jump irregularly sampled or noisy Lévy processes. *Statistica Neerlandica*, 64(3):290–313, 2010b. (Cited on page 38).

F. Comte and V. Genon-Catalot. Estimation for Lévy processes from high frequency data within a long time interval. *The Annals of Statistics*, 39(2):803–837, 2011. (Cited on page 38).

F. Comte, C. Duval, and V. Genon-Catalot. Nonparametric density estimation in compound Poisson processes using convolution power estimators. *Metrika*, 77(1):163–183, 2014. (Cited on pages 17, 20, 22, 25, 26, 36, 41, 42, 50, 52, and 59).

F. Comte, C. Duval, V. Genon-Catalot, and J. Kappus. Estimation of the jump size density in a mixed compound Poisson process. *Scandinavian Journal of Statistics*, 42 (4):1023–1044, 2015. (Cited on page 26).

R. Cont and P. Tankov. Calibration of jump-diffusion option-pricing models: a robust nonparametric approach. *Journal of Computational Finance*, 7:1–49, 2003. (Cited on pages 28, 38, and 108).

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90):297–301, 1965. (Cited on page 118).

S. Csörgő and V. Totik. On how long interval is the empirical characteristic function uniformly consistent. *Acta Scientiarum Mathematicarum*, 45(1-4):141–149, 1983. (Cited on page 73).

D. J. Daley and D. Vere-Jones. *An Introduction to the General Theory of Point Processes*. Springer, 1998. (Cited on page 20).

B. De Finetti. Sulle funzioni ad incremento aleatorio. *Rendiconti della Accademia Nazionale dei Lincei*, 10(3):163–168, 1929. (Cited on page 6).

A. den Boer and M. Mandjes. Convergence rates of Laplace-transform based estimators. *Bernoulli, to appear*, 2016. (Cited on page 40).

M. D. Donsker. Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, 23(2):277–281, 1952. (Cited on page 10).

J. L. Doob. Heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, 20(3):393–403, 1949. (Cited on page 10).

R. M. Dudley. Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois Journal of Mathematics*, 10(1):109–126, 1966. (Cited on page 10).

R. M. Dudley. *Uniform Central Limit Theorems*. Cambridge University Press, 1999. (Cited on pages 10 and 62).

C. Duval. Density estimation for compound Poisson processes from discrete data. *Stochastic Processes and their Applications*, 123(11):3963–3986, 2013a. (Cited on pages 17, 20, 22, 25, 26, 36, 42, 50, 52, and 59).

C. Duval. Nonparametric estimation of a renewal reward process from discrete data. *Mathematical Methods of Statistics*, 22(1):28–56, 2013b. (Cited on page 26).

C. Duval. When is it no longer possible to estimate a compound Poisson process? *Electronic Journal of Statistics*, 8(1):274–301, 2014. (Cited on pages 26 and 143).

C. Duval and M. Hoffmann. Statistical inference across time scales. *Electronic Journal of Statistics*, 5:2004–2030, 2011. (Cited on page 54).

A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642–669, 1956. (Cited on page 11).

P. Embrechts, R. Grübel, and S. M. Pitts. Some applications of the fast Fourier transform algorithm in insurance mathematics; This paper is dedicated to Professor WS Jewell on the occasion of his 60th birthday. *Statistica Neerlandica*, 47(1):59–75, 1993. (Cited on page 130).

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance*. Springer Science & Business Media, 2013. (Cited on page 20).

J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272, 1991. (Cited on page 13).

G. B. Folland. *Real analysis: modern techniques and their applications (Second Edition)*. John Wiley & Sons, 1999. (Cited on page 61).

E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, 2016. (Cited on pages 8, 10, 98, and 102).

E. Giné and J. Zinn. Bootstrapping general empirical measures. *The Annals of Probability*, 18(2):851–869, 1990. (Cited on page 111).

P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Application of Mathematics–Stochastic Modelling and Applied Probability. 2004. (Cited on pages 121 and 122).

V. L. Glivenko. Sulla determinazione empirica delle leggi di probabilitá. *Giornale dell' Istituto Italiano degli Attuari*, 4(1):92–99, 1933. (Cited on page 8).

R. Grübel and R. Hermesmeier. Computation of Compound Distributions I: Aliasing Errors and Exponential Tilting. *Astin Bulletin*, 29(02):197–214, 1999. (Cited on page 130).

S. Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *Journal of Nonparametric Statistics*, 21(3):321–343, 2009. (Cited on pages 38 and 49).

S. Gugushvili. Nonparametric inference for discretely sampled Lévy processes. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(1):282–307, 2012. (Cited on page 38).

S. Gugushvili, F. van der Meulen, and P. Spreij. Nonparametric Bayesian inference for multidimensional compound Poisson processes. *Modern Stochastics: Theory and Applications*, 2(1):1–15, 2015. (Cited on pages 43 and 60).

S. Gugushvili, F. van der Meulen, and P. Spreij. A non-parametric Bayesian approach to decompounding from high frequency data. *Statistical Inference for Stochastic Processes, online*, 2016. (Cited on pages 43 and 60).

J. Hájek. A characterization of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 14(4):323–330, 1970. (Cited on page 11).

J. Hájek. Local asymptotic minimax and admissibility in estimation. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1:175–194, 1972. (Cited on page 11).

A. Hald. *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935*. Sources and Studies in the History of Mathematics and Physical Sciences. Springer-Verlag New York, 2007. (Cited on page 7).

M. B. Hansen and S. M. Pitts. Nonparametric inference from the M/G/1 workload. *Bernoulli*, 4(2):737–759, 2006. (Cited on pages 20, 22, 23, 24, and 40).

J. Hoffmann-Jørgensen. *Stochastic processes on Polish spaces*. Number 39. Published (1991). Aarhus Universitet. Matematisk Institut, 1984. (Cited on page 10).

K. Itô. On stochastic processes. I. (Infinitely divisible laws of probability). *Japanese Journal of Mathematics*, 18:261–301, 1942. (Cited on page 6).

J. Kappus and M. Reiß. Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Statistica Neerlandica*, 64(3):314–328, 2010. (Cited on pages 32 and 49).

A. Khintchine. A new derivation of one formula by Lévy P. *Bulletin of the Moscow State University, I*, 1(1):1–5, 1937. (Cited on page 6).

J. Kiefer and J. Wolfowitz. Asymptotically minimax estimation of concave and convex distribution functions. *Probability Theory and Related Fields*, 34(1):73–85, 1976. (Cited on page 11).

A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell' Istituto Italiano degli Attuari*, 4:83–91, 1933. (Cited on page 8).

L. Le Cam. Limits of experiments. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1:245–261, 1972. (Cited on page 11).

P. Lévy. Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 3(3-4):337–366, 1934. (Cited on page 6).

E. Mariucci. Asymptotic equivalence for pure jump Lévy processes with unknown Lévy density and Gaussian white noise. *Stochastic Processes and their Applications*, 126(2): 503–541, 2016. (Cited on page 25).

M. H. Neumann and M. Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248, 2009. (Cited on pages 13, 15, 30, 31, 32, 73, and 74).

R. Nickl and M. Reiß. A Donsker theorem for Lévy measures. *Journal of Functional Analysis*, 263(10):3306–3332, 2012. (Cited on pages 13, 15, 16, 17, 30, 31, 32, 33, 34, 37, 43, 51, 54, and 60).

R. Nickl, M. Reiß, J. Söhl, and M. Trabs. High-frequency Donsker theorems for Lévy measures. *Probability Theory and Related Fields*, 164(1):61–108, 2016. (Cited on pages 32, 33, 34, 43, 52, 54, 60, 71, and 108).

D. Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics. Springer, 1984. (Cited on page 24).

R. Pyke and G. R. Shorack. Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *The Annals of Mathematical Statistics*, 39 (3):755–771, 1968. (Cited on page 10).

M. Reiß. Testing the characteristics of a Lévy process. *Stochastic Processes and their Applications*, 123(7):2808–2828, 2013. (Cited on pages 39 and 108).

K.-I. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999. (Cited on pages 5, 28, 37, 73, 78, and 95).

R. L. Schilling. Growth and Hölder conditions for the sample paths of Feller processes. *Probability Theory and Related Fields*, 112(4):565–611, 1998. (Cited on page 29).

A. V. Skorokhod. Limit theorems for stochastic processes with independent increments. *Theory of Probability and Its Applications*, 2(2):138–171, 1957. (Cited on page 10).

N. V. Smirnov. Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin Moscow University*, 2(2):3–16, 1939. (Cited on page 8).

J. Söhl. Confidence sets in nonparametric calibration of exponential Lévy models. *Finance and Stochastics*, 18(3):617–649, 2014. (Cited on pages 38 and 108).

J. Söhl and M. Trabs. A uniform central limit theorem and efficiency for deconvolution estimators. *Electronic Journal of Statistics*, 6:2486–2518, 2012. (Cited on page 15).

J. Söhl and M. Trabs. Option calibration of exponential Lévy models: Confidence intervals and empirical results. *Journal of Computational Finance*, 18(2), 2014. (Cited on pages 38 and 108).

M. Trabs. Calibration of self-decomposable Lévy models. *Bernoulli*, 20(1):109–140, 2014. (Cited on page 49).

M. Trabs. Information bounds for inverse problems with application to deconvolution and Lévy models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 51(4): 1620–1650, 2015a. (Cited on pages 16, 24, 27, 38, 42, 50, 51, 53, 54, 55, 57, and 58).

M. Trabs. Quantile estimation for Lévy measures. *Stochastic Processes and their Applications*, 125(9):3484–3521, 2015b. (Cited on page 39).

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. (Cited on pages 7, 9, 73, and 80).

A. W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media, 1996. (Cited on pages 10, 62, 93, 98, 103, 104, and 111).

B. van Es, S. Gugushvili, and P. Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13(3):672–694, 2007. (Cited on pages 17, 34, 35, 36, 38, 40, 41, 42, 50, 59, 107, and 117).

M. Wand. Finite sample performance of deconvolving density estimators. *Statistics & Probability Letters*, 37(2):131–139, 1998. (Cited on page 117).

J. E. Yukich. Weak convergence of the empirical characteristic function. *Proceedings of the American Mathematical Society*, 95(3):470–473, 1985. (Cited on pages 73 and 74).