Nonparametric estimation of Lévy processes with a view towards mathematical finance

A Thesis
Presented to
The Academic Faculty

by

José Enrique Figueroa-López

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Mathematics
Georgia Institute of Technology
April 2004

Nonparametric estimation of Lévy processes with a view towards mathematical finance

Approved by:

Christian Houdré, Committee Chair	Marcus C. Spruill
Christian Houdré, Advisor	Shijie Deng, ISyE
Robert P. Kertz	Richard Serfozo, ISyE

Date Approved: 26 November 2003

To God...

To my family and friends...

ACKNOWLEDGEMENTS

There have been many individuals without whom the completion of this work would not have been possible.

I would like to thank my advisor, Dr. Christian Houdré, for his support and orientation through all my studies at Georgia Tech. In addition to supporting and guiding my research efforts, he has also provided me with encouraging opportunities. I also thank him for his example of hard work and for setting high standard for me in research.

I would like to thank Professors Shijie Deng, Robert P. Kertz, Richard Serfozo, and Marcus C. Spruill for serving as members of my defense. I would like to specially recognize Professors Ted Hill, Robert Kertz, and Marcus Spruill for their teaching efforts. Also, I would like to express my gratitude to Patricia Reynaud-Bouret for enlightening discussions about her model selection methods.

I am grateful to the School of Mathematics at the Georgia Institute of Technology for providing the financial support for me throughout my doctoral studies. Several people at Georgia Tech are worth-mentioning. In particular, I want to thank Ms. Cathy Jacobson for all her work and dedication.

A lot of people contributed to my education and professional formation back in Mexico. I would like to acknowledge Professor Víctor Pérez-Abreu Carrión. He encouraged me to come to Georgia Tech and helped me in many different ways during my years at my second alma mater: CIMAT.

First and foremost, I would like to thank my family, and many beloved friends for their encouragement and support. Thanks to my parents: Alicia López Garcia and Enrique Figueroa Romero. Thanks to my brothers and sisters: Aideé, Ulises, MarcoTulio, and Xochitl. My gratitude goes to the most beloved person I have met: Claudia Antonini. A

very special dedication to my friends for more than fifteen years: Ayax R. Avila Ordoñez and Carlos A. Partida Andalon. At last but not least, I would like to thank my friends Lloyd Angulo, Luis Hernandez, Armel Kelome, Hamed Marrofi, Kasso Okoudjou, and Laura Rico.

TABLE OF CONTENTS

SUMMARY

Discrete-data based statistical methods for the calibration of financial models driven by Lévy processes are presented. The procedures rely on minimum contrast estimators for Poisson processes and on the short-time properties of Lévy processes. Therefore, the estimation is suitable for high-frequency data and for the analysis of the microstructure of stock prices.

Nonparametric estimation of the Lévy density s of a Lévy process is studied. Concretely, given a linear space $\mathcal{S}$ of possible Lévy densities, an asymptotically unbiased estimator for the orthogonal projection of s onto $\mathcal{S}$ is found. It is proved that the expected standard error of the proposed estimator realizes the smallest possible distance between the true Lévy density and the linear space $\mathcal{S}$ as the frequency of the data increases and as the sampling time period gets longer. Also, we develop data-driven methods to select a model among a collection of models $\{\mathcal{S}_m\}_{m\in\mathcal{M}}$. The method is designed to approximately realize the best trade-off between the error of estimation within the model and the distance between the model and the unknown Lévy density. As a result of this approach and of concentration inequalities for Poisson functionals, we prove *oracles inequalities* that guarantee us to reach the best expected error (using projection estimators) up to a constant.

A numerical study of our methods is presented for the case of histogram estimators and for Gamma Lévy processes as well as variance Gamma processes. To calibrate parametric models, a nonparametric estimation method with least-square errors is studied. Comparison with maximum likelihood estimation is provided.

On a separate problem, we review the theoretical properties of tempered stable processes, a class of processes of potential great use in Mathematical Finance.

CHAPTER I

INTRODUCTION

1.1 Opening thoughts

In its most primitive form, a "*thesis* is a formal (in depth) treatment of a subject based on *original research*". *Research* itself can be defined as a "methodical investigation into a subject in order to *discover facts*, to establish or revise a *theory*, or to develop a *plan of actions* based on the facts discovered". In this chapter, we intend to describe the subject of our dissertation and our main motivations, proceed to establish the novelty of our results by briefly reviewing the existing theories on the subject, and finally present our findings.

Let us give a short preview. The subject of the thesis is *the estimation of (pure jump) Lévy processes*. More formally, the statistical estimation of the "parameter" that controls the random evolution of the process (namely, the parameter in question is a measure, but for the sake of simplicity think of it as a nonnegative function that we call the Lévy density of the Lévy process). Our motivation comes from mathematical finance, and concretely, the recent application of pure jump Lévy processes for asset price modeling (see for instance [13], [11], [12], [4], [3], [14], [6], and references therein).

In our opinion, the existing theory and practice of estimation for pure jump Lévy processes do not provide reliable results and concrete objective measures of estimation errors, particularly when based on high frequency data where the microstructure of the market plays a fundamental role (see [10] for a good review of market microstructure and more recently [17]). Most of the existing works on the subject are informal and greatly intuitive in what estimation concerns. Even in the case where the theory behind the methods is sound and perfectly established, as maximum likelihood estimation and fast Fourier transforms are, no formal quantitative appreciation of the goodness of fit of the model or sensitivity to model mis-specifications has been discussed. These are real concerns that need to be dealt with since in most cases the likelihood function is intractable or does not have a close form at all, and thus, it is necessary to rely on numerical approximations and manipulation of the data to meet the assumptions of the method. On this matter, it seems that no detailed numerical analysis has been performed, and issues, like numerical stability and robustness to errors in data and *model*, have not been addressed (not even in numerical experiments).

The method that we use to deal with some of the previous issues relies precisely on the "microstructure" of Lévy processes; that is to say, the distributional and path properties of Lévy processes for small time spans. Locally, the dynamics of a Lévy process with non-continuous paths is better described as a (possibly infinite) "superposition of (compensated) jumps" plus a continuous Gaussian process with drift having independent and stationary increments (a Brownian motion with drift in the real case). This representation, called the Lévy-Itô decomposition of sample paths, associates to every Lévy process a unique process of jumps in time, mathematically described by a (marked) Poisson process or a Poisson process in $(0,\infty)\times\mathbb{R}^d$. Such a one-to-one relationship between the pure-jump Lévy process and the spatial Poisson process associated with the jump process, is the justification and motivation for our methods: we plan to estimate the Lévy density (that ultimately describes the jump nature of the process) using methods of estimation for Poisson processes.

The statistical inference for spatial Poisson processes has a long history (see for instance [23] and [21]), but to say the truth, our main incentive for this approach come from recent results on the nonparametric estimation of spatial Poisson processes and model selection methods (see [33]). There are two appealing accomplishments obtained by this theory: Oracle inequalities and competitive performance against minimax estimators. The theory assumes that the real model is not one of the models that are postulated, and content with selecting the model and a representative from it that approximately realizes the best tradeoff between the error of estimation within the model and the distance between the proposed models and the actual unknown model. Oracle inequalities precisely materialized this ideal. However, the methods presented in [33] seem to require finitely many jumps almost surely, and that condition is too much to ask for the general type of processes that one encounters in mathematical finance. We prove that this condition is actually superfluous, and properly modified the constructions in [33] to estimate nonparametrically the Lévy density, preserving Oracle inequalities.

To the best of our knowledge, the estimation of Lévy processes using statistical methods for point processes has not been formally considered so far. One can think of at least two reasons for this absence. On one hand, the fact that the increments of Lévy processes (on equally spaced time spans) constitute a random sample suggests to use standard statistical methods based on i.i.d. random variables; for instance, maximum likelihood estimation. The other reason is the apparent inaccessibility of the jumps of the process since we can only aspire to observe the process at finitely many times. The fact that the jumps are defined as limits of increments makes possible, at least theoretically, to approximate the jump at a particular point. However, that would require high frequency data, which in the old days was not a viable approach. Nowadays, we can access (almost in real time) financial data, and moreover data bases of intraday quotes are widely available.

The present work successfully combines both approaches, the microstructure of Lévy processes and the methods of estimation for Poisson processes in space, to estimate nonparametrically the Lévy density of the Lévy process. We believe that this approach will reduce the drawbacks of the standard methods. Being nonparametric, we do not rely on a particular model and hope that data itself validates the best model. Furthermore, we expect that the method will be more robust to departures from the assumptions of Lévy processes, since the representation of the jumps as a Poisson process is valid even if we do not have stationary increments. This last assumption is particularly suspicious for high-frequency financial returns due to intrinsic intraday seasonality. Moreover, hight-frequency returns (sometimes even daily returns) presents discreteness effects which contradicts the mathematical behavior of Lévy processes. However, we believe that our methods will still be informative to explain the probabilistic micro structure of the returns and to address the goodness of fit of Lévy based models in general.

1.2 General framework and background

The goal in this part is to introduce the framework of our results. In particular, we introduce the main object of study of this dissertation: Lévy processes. As a secondary objective, we present some results that are fundamental to our work and set some terminology used throughout all the thesis. Our review is by no means intended to be complete and we present results without proof. An excellent review of the subject is found in [39].

1.2.1 Lévy processes

Throughout this section, we assume the existence of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where all the random variables and stochastic processes are defined. The expectation with respect to the probability measure $\mathbb{P}$ is denoted by $\mathbb{E}$. The basic definitions and terminology of probability theory will be introduced as needed.

We say that $\{\mathbf{X}(t)\}_{t\geq 0}$ is stochastically continuous (in probability) if, for all $t \geq 0$ and $\varepsilon > 0$,

$$\lim_{h\to 0} \mathbb{P}[|\mathbf{X}(t+h) - \mathbf{X}(t)| > \varepsilon] = 0.$$

Here is our main object of study:

Definition 1.2.1 *A stochastic process $\{\mathbf{X}(t)\}_{t\geq 0}$ on $\mathbb{R}^d$ is a Lévy process if the following conditions are satisfied.*

(1) *For any $n \geq 1$ and reals $t_0 = 0 \leq t_1 \leq \cdots \leq t_n$, the random variables*

$$\mathbf{X}(t_1) - \mathbf{X}(t_0), \ldots, \mathbf{X}(t_n) - \mathbf{X}(t_{n-1})$$

are mutually independent;

(2) $\mathbf{X}(0) = 0$ *almost surely (a.s.) ;*

(3) *The distribution of* $\mathbf{X}(t+h) - \mathbf{X}(t)$ *does not depend on t;*

(4) *It is stochastically continuous;*

(5) *a.s. it has right continuous with left limits paths;*

As usual *right continuous with left limits* is written *càdlàg*. When a process satisfies (1) above, we say that the process has independent increments, while when (2) is satisfied we say the process has stationary increments. A process is called *additive* if the process satisfies (1), (2), (4), and (5).

The theory of Lévy processes is closely related to the concept of infinitely divisible distribution. Below, $\hat{\mu}$ stands for the characteristic function of the probability measure μ:

$$\hat{\mu}(\mathbf{z}) = \int_{\mathbb{R}^d} e^{i\mathbf{x}\cdot\mathbf{z}} \mu(d\mathbf{x}).$$

The notation $\hat{\mu}^t$ stands for the *distinguished t- th power* of the the complex valued function $\hat{\mu}$ (see pp. 33 of [39]).

Definition 1.2.2 *The distribution measure μ on $\mathbb{R}^d$ is infinitely divisible if, for any integer n, there is a probability measure μ_n such that*

$$\hat{\mu} = \hat{\mu}_n^n.$$

A more probabilistic characterization can be stated as follows. If μ is the distribution of a random variable $\mathbf{X}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then μ is infinitely divisible if on a possibly different probability space there exist independent identically distributed (i.i.d.) random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ such that

$$\mathbf{X} \overset{\mathfrak{D}}{=} \mathbf{X}_1 + \cdots + \mathbf{X}_n,$$

where $\overset{\mathfrak{D}}{=}$ means that the random variables on the right and left hand sides of the equality sign have the same distribution. We emphasize that having independent and stationary increments, the process $\{\mathbf{X}(t)\}_{t\geq 0}$ is characterized by the distribution of $\mathbf{X}(1)$. That is, if $\{\mathbf{Y}(t)\}_{t\geq 0}$ is another Lévy process, possibly defined on a different probability space, such that $\mathbf{X}(1)\overset{\mathfrak{D}}{=}\mathbf{Y}(1)$ then

$$\{\mathbf{X}(t)\}_{t\geq 0}\overset{\mathfrak{D}}{=}\{\mathbf{Y}(t)\}_{t\geq 0},$$

where the above notation means that their finite dimensional distributions are the same. Moreover, the distribution of $\mathbf{X}(1)$ is infinitely divisible, and reciprocally, for any infinitely divisible distribution μ there exists a stochastic process $\{\mathbf{X}(t)\}_{t\geq 0}$ defined on a probability space such that $\mu \sim \mathbf{X}(1)$ (here, $\sim$ means that μ is the distribution of $\mathbf{X}(1)$).

The following representation characterize the distribution of a Lévy process in terms of a measure ν, a matrix Σ, and a vector $\mathbf{b}$. This representation is called the Lévy-Khintchine representation.

Theorem 1.2.3 *(i) If $\{\mathbf{X}(t)\}_{t\geq 0}$ is a Lévy process, then there exist a $d \times d$ matrix Σ, a vector $\mathbf{b} \in \mathbb{R}^d$ and a measure ν on $\mathbb{R}^d_0 \equiv \mathbb{R}^d \backslash \{0\}$ such that*

$$\mathbb{E}\left[e^{i\mathbf{z}\cdot\mathbf{X}(t)}\right] = \exp\left(t\psi(\mathbf{z})\right), \quad \mathbf{z} \in \mathbb{R}^d$$

where

$$\psi(\mathbf{z}) = -\frac{1}{2}\,\mathbf{z}\cdot\Sigma\mathbf{z} + i\,\mathbf{z}\cdot\mathbf{b} + \int\limits_{\mathbb{R}^d_0} \left\{e^{i\,\mathbf{z}\cdot\mathbf{x}} - 1 - i\,\mathbf{z}\cdot\mathbf{x}\,I\left(\|\mathbf{x}\| \leq 1\right)\right\}\nu(d\mathbf{x}). \tag{1.2.1}$$

Moreover, Σ is nonnegative-definite symmetric and ν satisfies

$$\int_{\mathbb{R}^d_0} (\|\mathbf{x}\|^2 \wedge 1)\nu(d\mathbf{x}) < \infty. \tag{1.2.2}$$

(ii) The representation given in (i) via Σ, ν, and $\mathbf{b}$ is unique.
(iii) Conversely, if Σ is a symmetric nonnegative-definite matrix, ν is a measure satisfying 1.2.2, and $\mathbf{b} \in \mathbb{R}^d$, then there exists a Lévy process $\{\mathbf{X}(t)\}_{t\geq 0}$ on a probability space, possibly different from $(\Omega, \mathcal{F}, \mathbb{P})$, whose characteristic function is as in (i).

Definition 1.2.4 *We called ν the Lévy measure of the Lévy process $\mathbf{X}$. The triple $(\Sigma, \nu, \mathbf{b})$ is called the* generating triple *of the distribution of $\mathbf{X}(1)$ or the generating triple of the Lévy process $\{\mathbf{X}(t)\}_{t\geq 0}$. The function ψ is sometimes called the Lévy exponent or the characteristic exponent of the Lévy process. If in addition ν is absolutely continuous, we say that the function s satisfying $\nu(d\mathbf{x}) = s(\mathbf{x})d\mathbf{x}$ is the Lévy density of the process $\{\mathbf{X}(t)\}_{t\geq 0}$.*

It is illustrative to give a few snapshots of the proof of the above theorem that are relevant to our work. Concretely, we are looking for short-time characterizations of the generating triplet that are feasible for estimation based on high-frequency observations of X.

Remark 1.2.5 *(i) The uniqueness of the matrix Σ is a consequence of the following limit for $\mathbf{X}(t)$:*

$$\frac{1}{\sqrt{t}}\mathbf{X}(t)\xrightarrow{\mathfrak{D}}\mathbf{Y}, \quad t \to 0$$

where $\mathbf{Y} \equiv (\mathbf{Y}_1, \ldots, \mathbf{Y}_d)$ is a Multivariate Gaussian vector with variance-covariance matrix Σ (see pp. 40 in [39]). Above, $\xrightarrow{\mathfrak{D}}$ means limit in distribution, in this case as random elements of $\mathbb{R}^d$.

(ii) Another consequence of the proof for Theorem 1.2.3 is the following characterization of the Lévy measure ν. Namely, for any function f from $\mathbb{R}^d$ to $\mathbb{R}$ that is continuous, bounded and vanishes on a neighborhood of the origin:

$$\lim_{t\to 0}\frac{1}{t}\mathbb{E}\left[f\left(\mathbf{X}(t)\right)\right] = \int_{\mathbb{R}_0^d} f(x)\nu(dx).$$

We will use this type of limits in Section 2.3 to estimate the integral of f with respect to the random measure associated with the jumps of $\{\mathbf{X}(t)\}_{t\geq 0}$. In the same chapter we will also state related results for more general functions (these results are obtained in [36]). For now, let us state the next limit:

$$\lim_{n\to\infty} n\mathbb{E}\left[|\mathbf{X}(1/n)|^p\right] = \int_{\mathbb{R}_0^d} |x|^p\nu(dx),$$

where $p \geq 2$ and where it is assumed that $\mathbb{E}\left[|\mathbf{X}(1)|^p\right] < \infty$ and $\mathbb{E}\left[\mathbf{X}(1)\right] = 0$ (see [2] for a proof).

(iii) Finally, let us introduce the concept of drift of a Lévy process. If $\int_{||\mathbf{x}||\leq 1} ||\mathbf{x}||\nu(d\mathbf{x}) < \infty$, the vector $\mathbf{b}_0 \equiv \mathbf{b} - \int_{||\mathbf{x}||\leq 1} \mathbf{x}\nu(d\mathbf{x})$, where the integration is component wise, is called the drift of the Lévy process . If $\Sigma = 0$ and the drift exists,

$$\mathbb{P}\left[\lim_{t\to 0}\frac{1}{t}\mathbf{X}(t) = \mathbf{b}_0\right] = 1.$$

On the other hand, if $\int_{\|\mathbf{x}\|>1} \|\mathbf{x}\|\nu(d\mathbf{x}) < \infty$*, then* $\mathbf{X}(t)$ *has finite mean for any* t *(this is a necessary condition too) and*

$$\mathbb{E}[\mathbf{X}(t)] = t\left(\int_{\|\mathbf{x}\|>1} \mathbf{x}\nu(d\mathbf{x}) + \mathbf{b}\right).$$

Similarly, if $\int_{\|\mathbf{x}\|>1} \|\mathbf{x}\|^2\nu(d\mathbf{x}) < \infty$*, then* $\mathbf{X}(t)$ *has finite second moment for any* t *(this is a necessary condition too) and*

$$\mathbb{E}[(X_i(t) - X_i(t))(X_k(t) - X_k(t))] = t\left(\Sigma_{i,k} + \int_{\mathbb{R}_0^d} x_i x_k \nu(d\mathbf{x})\right),$$

for $i, k = 1, \ldots, d$*. Here,* $X_i(t)$ *and* x_i *refers to the* i^{th} *component of the vectors* $\mathbf{X}(t)$ *and* $\mathbf{x}$*, respectively.*

The Lévy-Khintchine representation states that the law of any Lévy process is characterized by three components: A Gaussian component, a "drift" component, and a "pure jump component". The celebrated Lévy-Itô decomposition extends this characterization to the sample paths of the process. We present below the Lévy-Itô decomposition for processes with independent increments. This version is taken from [20], Theorem 13.4. Throughout this section, the integrals of vector-valued functions with respect to measures are defined component wise.

Theorem 1.2.6 *Let* $\{\mathbf{X}(t)\}_{t\geq 0}$ *be a càdlàg process in* $\mathbb{R}^d$ *with* $\mathbf{X}(0) = 0$*. Then* $\mathbf{X}$ *has independent increments and no fixed jumps*[1] *if and only if, a.s.*

$$\mathbf{X}(t) = \mathbf{b}(t) + \mathbf{G}(t) + \int_0^t \int_{\|\mathbf{x}\|\leq 1} \mathbf{x}\,(\mathcal{J} - \mathbb{E}\mathcal{J})(ds\,d\mathbf{x}) + \int_0^t \int_{\|\mathbf{x}\|>1} \mathbf{x}\,\mathcal{J}(ds\,d\mathbf{x}), \qquad (1.2.3)$$

for every $t \geq 0$*, for some continuous function* $\mathbf{b}$ *with* $\mathbf{b}(0) = 0$*, some continuous centered Gaussian process* $\mathbf{G}$ *with independent increments and* $\mathbf{G}(0) = 0$*, and some independent Poisson process* $\mathcal{J}$ *on* $(0, \infty) \times \mathbb{R}_0^d$ *with*

$$\int_0^t \int_{\mathbb{R}_0^d} (\|\mathbf{x}\|^2 \wedge 1)\mathbb{E}\mathcal{J}(ds\,d\mathbf{x}) < \infty, \quad t > 0. \qquad (1.2.4)$$

[1] *We say the* $\mathbf{X}$ *has a fixed jump at some* $t > 0$ *if* $\mathbb{P}[\mathbf{X}(t) \neq \mathbf{X}(t^-)] > 0$*.*

In the special case when $\mathbf{X}$ *is real and nondecreasing, (1.2.3) simplifies to*

$$\mathbf{X}(t) = \mathbf{a}(t) + \int_0^t \int_0^\infty \mathbf{x}\,\mathcal{J}(ds\,d\mathbf{x})), \tag{1.2.5}$$

for some nondecreasing continuous function $\mathbf{a}$ *with* $\mathbf{a}(0) = 0$ *and some Poisson process* $\mathcal{J}$ *on* $(0,\infty)\times(0,\infty)$ *with*

$$\int_0^t \int_0^\infty (\|\mathbf{x}\| \wedge 1)\mathbb{E}\mathcal{J}(ds\,d\mathbf{x}) < \infty, \quad t > 0. \tag{1.2.6}$$

Both representations are almost surely unique, and all functions $\mathbf{b}$*,* $\mathbf{a}$ *and processes* $\mathbf{G}$*,* $\mathcal{J}$ *with the stated properties may occur in (1.2.3) or (1.2.5) for a process* $\{\mathbf{X}(t)\}_{t\geq 0}$ *defined on some probability space.*

As a corollary, we readily obtain the Lévy-Itô decomposition for Lévy processes:

Theorem 1.2.7 *A càdlàg process* $\{\mathbf{X}(t)\}_{t\geq 0}$ *in* $\mathbb{R}^d$ *is a Lévy process if and only a.s.*

$$\mathbf{X}(t) = \mathbf{b}t + \Sigma_0\mathbf{B}(t) + \int_0^t \int_{\|\mathbf{x}\|\leq 1} \mathbf{x}\,(\mathcal{J} - \mathbb{E}\mathcal{J})(ds\,d\mathbf{x}) + \int_0^t \int_{\|\mathbf{x}\|>1} \mathbf{x}\,\mathcal{J}(ds\,d\mathbf{x}), \tag{1.2.7}$$

for every $t \geq 0$*, for some vector* $\mathbf{b} \in \mathbb{R}^d$*, some* $d \times d$ *matrix* Σ_0*, some independent Poisson process* $\mathcal{J}$ *on* $(0,\infty)\times\mathbb{R}_0^d$ *with mean measure of the form* $\mathbb{E}\mathcal{J}(dt\,d\mathbf{x}) = dt\nu(d\mathbf{x})$*, and a standard Brownian Motion* $\mathbf{B}$ *in* $\mathbb{R}^d$ *independent of the process* $\mathcal{J}$*. Moreover, the measure* ν *satisfies*

$$\int_{\mathbb{R}_0^d} (\|\mathbf{x}\|^2 \wedge 1)\nu(d\mathbf{x}) < \infty. \tag{1.2.8}$$

The representations is unique, and all $\mathbf{b}$*,* Σ_0*. and* ν *with the stated properties may occur in (1.2.7) for a process* $\{\mathbf{X}(t)\}_{t\geq 0}$ *defined on some probability space.*

The Poisson integrals in (1.2.3) and (1.2.7) are defined in the sense of “approximation in probability” (see Definition 1.2.11 below). Now, let us give some other specifications of the representation. The following remarks are useful in the application of the Lévy-Itô representation (see Section 19 of [39]).

Remark 1.2.8 *(i) In the equation (1.2.3), we can take*

$$\int_0^t \int_{\|\mathbf{x}\|\le 1} \mathbf{x}\,(\mathcal{J} - \mathbb{E}\mathcal{J})(ds\,d\mathbf{x}) = \lim_{\varepsilon\downarrow 0} \int_0^t \int_{\varepsilon<\|\mathbf{x}\|\le 1} \mathbf{x}\,(\mathcal{J} - \mathbb{E}\mathcal{J})(ds\,d\mathbf{x}). \tag{1.2.9}$$

where the limit exists almost surely. Furthermore, a.s., the convergence is uniform in t on any bounded interval.

(ii) As a consequence of representation (1.2.3), the random measure $\mathcal{J}$ is almost surely determined by the sample paths of $\{\mathbf{X}(t)\}_{t\ge 0}$. Concretely, a.s.

$$\mathcal{J}(B) = \#\{t : (t, \Delta\mathbf{X}(t)) \in B\},$$

for every Borel set B of $[0,\infty)\times\mathbb{R}^d_0$. Here, # denotes cardinality, and $\Delta\mathbf{X}(t)$ is the "jump" of $\mathbf{X}$ at time t defined as $\Delta\mathbf{X}(t) \equiv \mathbf{X}(t) - \lim_{s\uparrow t}\mathbf{X}(s)$. In view of this, we usually call $\mathcal{J}$ the random measure associated with the jumps of $\mathbf{X}$ or simply the jump measure. One of the statements of the Lévy-Itô representation is that the random measure associated with the jumps of a Lévy process is a Poisson process on $[0,\infty)\times\mathbb{R}^d_0$. For a better understanding of equation (1.2.3), note it can be written in terms of the jump process as follows:

$$\begin{aligned}
\mathbf{X}(t) &= \mathbf{b}(t) + \mathbf{G}(t) \\
&\quad + \lim_{\varepsilon\downarrow 0}\left\{\sum_{s:s\le t}\Delta\mathbf{X}(s)\mathbf{1}(\varepsilon < \|\Delta\mathbf{X}(s)\| \le 1) - \int_{\varepsilon<\|\mathbf{x}\|\le 1}\mathbf{x}\,\nu_t(d\mathbf{x})\right\} \\
&\quad + \sum_{s:s\le t}\Delta\mathbf{X}(s)\mathbf{1}(\|\Delta\mathbf{X}(s)\| > 1),
\end{aligned}$$

where $\nu_t(B) \equiv \int_0^t\int_B \mathbb{E}\mathcal{J}(ds, d\mathbf{x})$. Here, $\mathbf{1}(C)$ is the indicator function of C, that takes the value 1 if C is true and takes the value 0 otherwise. The first two terms are called the continuous part *of the process, the last two terms are called* the pure jump part *of the process. Sometimes, the second term $\mathbf{G}(t)$ is called the* Gaussian component *of the process , the third term the* compensated Poisson part *and the last term simply the* compound Poisson component. *A* pure-jump Lévy process *is a Lévy process that has no continuous part.*

(iii) If in addition, we have

$$\int_0^t \int_{\|\mathbf{x}\|\le 1} \|\mathbf{x}\|\;\mathbb{E}\mathcal{J}(ds\,d\mathbf{x}) < \infty, \quad \forall t > 0, \tag{1.2.10}$$

then a.s.

$$\mathbf{X}(t) = \mathbf{b}(t) + \mathbf{G}(t) + \int_0^t \int_{\mathbb{R}^d} \mathbf{x}\, \mathcal{J}(ds\, d\mathbf{x}), \tag{1.2.11}$$

where $\mathbf{b}$ *and* $\mathbf{G}$ *are as in Theorem 1.2.6.*

(iv) Another nontrivial consequence of the Lévy-Itô decomposition is the fact that the marginal distributions of $\mathbf{X}$ *are infinitely divisible distributions and thus, admit the Lévy-Khintchine representation of Theorem 1.2.3. Concretely, we have that*

$$\mathbb{E}\left[e^{i\,\mathbf{z}\cdot\mathbf{X}(t)}\right] = \exp\left(t\psi_t(\mathbf{z})\right), \quad \mathbf{z} \in \mathbb{R}^d$$

where

$$\psi_t(\mathbf{z}) = -\frac{1}{2}\,\mathbf{z}\cdot\Sigma_t\mathbf{z} + i\,\mathbf{z}\cdot\mathbf{b}(t) + \int\limits_{\mathbb{R}^d_0} \left\{e^{i\,\mathbf{z}\cdot\mathbf{x}} - 1 - i\,\mathbf{z}\cdot\mathbf{x}\, I\left(\|\mathbf{x}\| \le 1\right)\right\} \nu_t(d\mathbf{x}), \tag{1.2.12}$$

where Σ_t *is the variance-covariance matrix of* $\mathbf{G}(t)$, $\mathbf{b}(t)$ *is as in (1.2.3), and* $\nu_t(d\mathbf{x}) \equiv \mathbb{E}\mathcal{J}(t, d\mathbf{x})$.

1.2.2 Poisson processes and Poisson integrals

There are excellent references about Poisson processes (see for instance [32] and [20]). In this section we give the very basic results that are used in the present dissertation. Below, $\bar{\mathbb{Z}}_+$ is the union of the set of all positive integers and $+\infty$. Also, as a convention, we say that a random variable X has Poisson distribution with mean 0 if $X = 0$ a.s. Similarly, we say that a random variable X has Poisson distribution with mean ∞ if $X = \infty$ a.s.

Definition 1.2.9 *Let* $(S, \mathcal{S}, \rho)$ *be a* σ*-finite measure space. A collection of* $\bar{\mathbb{Z}}_+$*-valued random variables* $\{\mathcal{J}(B) : B \in \mathcal{S}\}$ *is called a Poisson process (or Poisson random measure) with mean measure* ρ*, if the following hold:*

(1) for every B, $\mathcal{J}(B)$ *is a Poisson random variable with mean* $\rho(B)$*;*

(2) if $B_1, \ldots, B_n$ *are disjoint, then* $\mathcal{J}(B_1), \ldots, \mathcal{J}(B_1)$ *are independent;*

(3) for every ω, $\mathcal{J}(\cdot;\omega)$ *is a measure on* $\mathcal{S}$.

The proof of the following proposition can be found in [39]

Proposition 1.2.10 *For any σ-finite measure space $(S, \mathcal{S}, \rho)$, there exists, on some probability space $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$, a Poisson process $\{\mathcal{J}(B) : B \in \mathcal{S}\}$ on S with mean measure ρ.*

Note that for fixed $\omega \in \Omega$ the integral

$$I(\varphi; \omega) \equiv \int_S \varphi(x)\mathcal{J}(dx; \omega).$$

can be defined for all measurable functions $\varphi : S \to \mathbb{R}_+$, because of (3) in the definition of Poisson processes. For general measurable functions $\varphi : S \to \mathbb{R}$, the integral $I(\varphi; \omega)$ might not exist in the sense of Lebesgue integration. This is not a problem if we assume that ρ is finite (see Proposition 19.5 of [39]). In this case, we actually have that $I(\varphi)$ is infinitely divisible of the compound Poisson type with characteristic function:

$$\begin{aligned} \mathbb{E}\{\exp(i\mathbf{z} \cdot I(\varphi))\} &= \exp\left\{\int_S (e^{i\mathbf{z}\cdot\varphi(x)} - 1)\rho(dx)\right\} \\ &= \exp\left\{\int_{\mathbb{R}^d} (e^{i\mathbf{z}\cdot\mathbf{x}} - 1)(\rho\varphi^{-1})(d\mathbf{x})\right\}. \end{aligned} \tag{1.2.13}$$

Below, we give conditions to define Poisson integrals when S is a locally compact metric space with countable basis, $\mathcal{S}$ is the corresponding Borel σ-field, and ρ is a radon measure (see Chapter 10 of [20] for a detailed exposition). For our purposes, we can assume that S is an open subset of $\mathbb{R}^d$. Notice that $\int_S \varphi(x)\mathcal{J}(dx)$ exists a.s. if φ has compact support and is bounded. The following definition introduces a weaker type of integration.

Definition 1.2.11 *We say that $\int_S \varphi(x)\mathcal{J}(dx)$ exists (in the sense of approximation in probability) when, for any sequence $\{\varphi_n\}$ of bounded functions with compact support such that $|\varphi_n| \le |\varphi|$ and $\varphi_n \to \varphi$, the random variable $\int_S \varphi_n(x)\mathcal{J}(dx)$ converge in probability to the same limit (which is denoted by $\int_S \varphi(x)\mathcal{J}(dx)$).*

The next theorem states conditions for the existence of Poisson integrals and compensated Poisson integrals (see 10.2 and 10.15 of [20]).

Theorem 1.2.12 *Let $\mathcal{J}$ and $\mathcal{J}'$ be independent Poisson processes on S with common σ-finite radon mean measure ρ. Fix any measurable function $\varphi : S \to \mathbb{R}$. The following statements hold true:*

1. *If $\varphi : S \to \mathbb{R}_+$, then $\int_S \varphi(x)\mathcal{J}(dx;\omega)$ exists (in the sense of Lebesgue integration) for almost every ω and*
$$\mathbb{E}\left\{\exp\left(-\int_S \varphi(x)\mathcal{J}(dx)\right)\right\} = \exp\left\{-\int_S (1-e^{-\varphi(x)})\rho(dx)\right\}. \tag{1.2.14}$$

2. *If instead $\varphi : S \to \mathbb{R}$ and $\int_S(|\varphi(x)| \wedge 1)\rho(dx) < \infty$, then $\int_S \varphi(x)\mathcal{J}(dx;\omega)$ exists (in the sense of Lebesgue integration) for almost every ω, and*
$$\mathbb{E}\left\{\exp\left(i\int_S \varphi(x)\mathcal{J}(dx)\right)\right\} = \exp\left\{\int_S (e^{i\varphi(x)}-1)\rho(dx)\right\}. \tag{1.2.15}$$

3. *$\int_S \varphi(x)(\mathcal{J}-\mathcal{J}')(dx)$ exists in the sense of approximation in probability if and only if $\int_S(|\varphi(x)|^2 \wedge 1)\rho(dx) < \infty$.*

4. *The compensated integral $\int_S \varphi(x)(\mathcal{J}-\rho)(dx)$ exists in the sense of approximation in probability if and only if $\int_S(|\varphi(x)|^2 \wedge |\varphi(x)|)\rho(dx) < \infty$.*

The next proposition is well known and easy to derive.

Proposition 1.2.13 *If $\int_S(|\varphi(x)|\rho(dx) < \infty$, then $\mathbb{E}\left|\int_S \varphi(x)\mathcal{J}(dx;\omega)\right| < \infty$, and*
$$\mathbb{E}\int_S \varphi(x)\mathcal{J}(dx;\omega) = \int_S \varphi(x)\rho(dx).$$
If $\int_S \varphi^2(x)\rho(dx) < \infty$ then $\mathbb{E}\int_S \varphi^2(x)\mathcal{J}(dx;\omega) < \infty$, and
$$\mathrm{Var}\left(\int_S \varphi(x)\mathcal{J}(dx;\omega)\right) = \int_S \varphi^2(x)\rho(dx).$$

1.3 Motivation for our research

Consider a real Lévy process $\{X(t)\}_{t\geq 0}$ with unknown characteristic triplet (Σ, ν, b). Statistical inference for Lévy processes in principle should be a straightforward extensions of the standard statistical machinery for random samples. This appreciation is justified up to certain point. The mere definition of Lévy processes as processes with independent and stationary increment implies that we can readily construct the random sample

$$X_i^h \equiv X(t_i) - X(t_{i-1}), \quad i = 1, \ldots, n,$$

whenever $0 = t_0 < t_1 < \ldots$ are equally spaced sampling times with given time span h. In that case, by applying "standard" statistical methods to a finite sample $X(t_1), \ldots, X(t_n)$, statistical inferences for the distribution of $X(h)$ seem to be doable.

There are two problems with this paradigm. First, it is well known that parsimonious models in the "Fourier domain" do not corresponds to parsimonious models in the "space domain". Indeed, parsimonious parametric models for the Lévy measure can produce not only intractable but sometimes not even expressible density functions (assuming such density exists). We can give numerous examples of this phenomenon, but the most obvious and relevant one, in what statistical inference concerns, is a scaling of the Lévy measure. The density function, say f_h, corresponding to the characteristic triplet $(h\Sigma, h\nu, hb)$ has no general relationship to the density function f of (Σ, ν, b). In particular, the density function $f_t(\cdot)$ of $X(t)$ can "greatly" change in shape (even for small changes) in t. Consider the case of a Gamma Lévy process $\{X(t)\}_{t\geq 0}$ with Lévy measure of the form $\nu(dx) = x^{-1}\exp(-x)\mathbf{1}(x > 0)dx$ (see Section 3.2.2 for a more comprehensive description of this model). The density function of $X(1)$ is exponential. For any $t < 1$, the density is strictly decreasing with asymptote at $x = 0$. However, for any $t > 1$, the density of $X(t)$ is unimodal and approaches to 0 as $x \to 0$. From the point of view of the paradigm outlined above, this implies that the likelihood function is going to be "instable" as a function of h around $h = 1$. In other word, the likelihood function based on daily data will be completely different from the likelihood function based on weekly data. Another easy change in the Fourier space that leads to striking changes in the "space domain" is the superposition of Lévy measures. This will not be pursued any longer since we think that we already made our point here.

The other problem with the procedure above is more delicate to describe and harder to quantify. We pose the following question: what are the effects of small time increments in the results of standard statistical methods over finite time horizons?. This problem is relevant in applications involving high frequency data. The fact that the sample size of the data increases does not necessarily mean an improved "reliability" or precision of our results because the target distribution $\mathcal{L}(X(h))$ is changing as well.

Let us come back to the first situation where the Lévy process has simple Lévy density, but "intractable" density function. Such settings are particularly common in recent application of Lévy processes to asset price modeling. These financial models are driven by Lévy processes in the same way as Samuelson's geometric Brownian motion is driven by the Brownian motion (see [38]). Namely, the model represents the price $S(t)$ at time t of a risky asset by

$$S(t) = S(0)e^{X(t)}, \tag{1.3.1}$$

where $\{X(t)\}_{t\geq 0}$ is a Lévy process. One of the first to propose this geometric Lévy process was Mandelbrot [26]. He postulates that the prices of commodities is given by (1.3.1) with $\{X(t)\}_{t\geq 0}$ being a Symmetric α-stable Lévy motion with $\alpha < 2$. Later, Press [30] proposed a Brownian motion plus an independent compound Poisson process with normally distributed jumps. More recently, Madan and Seneta [24] introduced a model of this type that has influenced many future works in this area. They propose a Lévy process with "infinite activity of jumps" but bounded variation of paths, namely, the variance gamma model for log prices (see Section 3.2.3 for more on this model). This model has been increasingly specialized to better fit the empirically observed distributions and simultaneous fit the option prices (see [13], [11], [12]). The density function in most of these models do not have closed form expressions, and techniques using the characteristic function are inevitable. Another school of modeling considers the so called generalized hyperbolic distributions (see for instance [4], [3], [14], [6], and the references therein). The density functions in this class of models have closed form expressions, but in most case, they lead to intractable likelihood functions involving Bessel functions, exponentials, powers of x, etc. Statistical inference becomes numerically challenging and expensive.

In a technical report, Rosiński [35] studies the *tempered stable* model that encompasses the variance Gamma model and the CGMY model of [11]. One important contribution of this work is to recognize and stress the relationship of this type of models to the class of stable processes. Two fundamental connections were pointed out. On one hand, the scaling behavior of the process for both short time spans and long time spans are considered. It is found that in the short term the increments of the process behave (in the limit) as the increments of stable process, while in long spans the increments behave like the increments of a Brownian motion (up to a scaling in space and a shift). We found this property quite enlightening for financial applications. The second connection has to do with changes in the probability measure. In short, statistically a tempered stable process looks like a stable process under a suitable change in the probability measure. We study in detail this class of processes in Chapter 4 providing our own proofs to his results (so far we have not had the opportunity to see a complete version with proofs of this technical report). In its most general form, tempered stable processes involves a completely monotone function q (typical examples are e^{-x}, $1/(x+1)$, etc). This fact was another incentive for looking into nonparametric estimation methods for Lévy densities.

1.4 Brief overview of estimation methods for Lévy processes

Let $\{X(t)\}_{t\geq 0}$ be a real-valued Lévy process. The following problem is tackled: Assume that expressions for the characteristic function φ_h or for the Lévy density s_h of $X(h)$ are simple, but that the corresponding (marginal) density function f_h has either one of the following two shortcomings: it produces "intractable" likelihood functions, or it does not have a "closed expression". We mention below some methods found in the literature to deal with this problem.

The most wide spread approach is *Likelihood based methods*. This approach relies on an inversion formula for the characteristic function that evaluates the density function f_h at a point x from the characteristic function φ_h. The following fundamental result gives an inversion formula for probability densities:

Proposition 1.4.1 *Let φ be the characteristic function of a probability measure μ and suppose that $\int_{\mathbb{R}} |\varphi(z)|dz < \infty$. Then μ has a bounded continuous density f given by the Fourier transform of φ:*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-izx} \varphi(z) dz. \tag{1.4.1}$$

Therefore, based on sampling observations $x_1, \ldots, x_n$ of the increments $X(h)$, $X(2h) - X(h)$, ..., $X(nh) - X((n-1)h)$, the likelihood function of φ is implicitly given by

$$L_h(x_1, \ldots, x_n; \varphi) = \prod_{i=1}^{n} f_h(x_i) = \frac{1}{(2\pi)^n} \prod_{i=1}^{n} \int_{-\infty}^{\infty} e^{-izx_i} \varphi_h(z) dz. \tag{1.4.2}$$

As it is clear from the previous expression, any statistical inference for φ_h based on the likelihood function is highly computational expensive since it requires to compute a Fourier transform for each data point x_i, $i = 1, \ldots, n$. For instance, say that the characteristic function φ_θ is determined by one parameter $\theta \in \mathbb{R}$ that we wish to estimate. Let us write $L_h(\theta) = L_h(x_1, \ldots, x_n; \varphi_\theta)$. In order to find the maximum likelihood estimator of θ it is necessary to repeatedly evaluate the likelihood function $L_h(\theta)$ at different values of θ. Each evaluation requires n Fourier transforms. Therefore, the computational intensity of this approach appears non-viable for applications.

The following method is outlined in [11]. Fast Fourier transforms are applied to evaluate the levels of the function f at a regular lattice $\{x_k\}$ of an interval $[-b, b]$. In their own words, "fast Fourier transform effectively renders the level of the probability density at prespecified set of points", say $x_k = k\Delta x$, for $k = -N, \ldots, N$ and $\Delta = b/N$. Certainly, this statement deserves better justification or at least a reference. In any case, the next step is to "bin" the data to get an approximate likelihood function $L_N(\theta)$. Concretely, they consider the following multinomial like likelihood function:

$$L_N(\theta) = \prod_{k=-N}^{N} f_\theta(x_k)^{n_k},$$

where n_k is the number of points in the sample for which the closest grid point is x_k. Finally, "the parameter estimates that maximize the likelihood of this binned data are searched". To the best of our knowledge, there is no detailed numerical analysis of this method.

When explicit forms for the density functions are available, [6] considers a multinomial log likelihood function. More precisely, if $I_1, \ldots, I_k$ are disjoint intervals with union the entire real line and n_j is the number of observations in I_j , $j = 1, \ldots, k$, the multinomial log-likelihood function is given by $l(\theta) = \sum_{j=1}^{k} n_j \log p_j$, where p_j is the probability that the increment takes a value in I_j.

Another popular approach is *simulation based methods*. The very general idea behind this approach is to select a model (probably described via a parameter) that best "matches" the sampling observations and simulated observations using the model. To measure the closeness between the observed data and the simulated data it is necessary to look for concrete empirical characteristics like quantiles. So, the estimated model can be proposed to be one that minimize the distance between some empirical quantiles computed from the sample observations and the corresponding quantiles computed from simulated data using the model (see for instance [19]).

The most meaningful approach for the subject of our work is *jumps based methods*. As far as we know, this approach has not been fully explored in the context of estimation for Lévy densities before this dissertation. While the present thesis was in progressed, we became aware of a relevant result of S. Raible [31]. He considers the estimation of the Lévy density of Generalized Hyperbolic Lévy processes. It is proved there that some parameters in that model are invariant under equivalent changes in the probability measure. He also recognizes that this is a consequence of the fact that these parameters are "sample path dependent" and not "distributional dependent". In relation to that, the following general result is obtained in [31].

Proposition 1.4.2 *Let X be a Lévy process with finite second moments such that the Lévy measure has a density $s(x)$ with asymptotic behavior*

$$s(x) = \frac{a}{x^2} + o\left(x^{-2}\right), \quad \text{as} \quad x \downarrow 0.$$

Fixed an arbitrary time T and consider the sequence of random variables

$$S_n = \frac{1}{Tn} \# \left\{ s \leq T : \Delta X(s) \in \left[\frac{1}{k+1}, 1 \right) \right\}, \quad k \geq 1.$$

Then, a.s., S_n converges to the value a as $n \to \infty$.

It can be said that a is a “path dependent” parameter that can be determined from properties of a “typical path”. Using simulation, the convergence of the estimator as the number of jumps in the sample path gets larger was numerically illustrated . However, the problem of approximating S_n if the jump process is not observed is not considered.

1.5 Outline of the thesis

In Chapter 2, it is discussed the nonparametric estimation of the Lévy density s of a real Lévy process $X = \{X(t)\}_{t \geq 0}$ based on discrete observations of the process in a time period $[0, T]$. We devise two methods to accomplish this. The first method constructs estimators, say $\hat{s}(x)$, which can be written in terms of integrals of deterministic functions with respect to the random measure associated with the jumps of X. Concretely, the first method consists of two subparts:

(1) the selection of a good estimator $\hat{s}$ from a linear model $\mathcal{S}$ of possible estimators. Since the proposed estimator $\hat{s}$ is designed to be an unbiased estimator of the orthogonal projection of s onto $\mathcal{S}$, $\hat{s}$ is called the *projection estimator* of s on $\mathcal{S}$. It is proved that, when the time horizon T increases, the distance between the proposed estimator $\hat{s}$ and the Lévy density s realizes the smallest distance between s and any other estimator in $\mathcal{S}$;

(2) the selection of a linear model among a given collection of linear models. The proposed selection criteria is designed to approximately realize the best tradeoff between the error of estimation withinin the model and the distance between the model and the unknown Lévy density s. The resultant estimator is a type of *penalized projection estimator* that assesses the goodness of a projection estimator not only by its approximation quality inside the model but also by its complexity and variance. These last two characteristics are controlled by suitable penalty functions.

It is shown that the methodology of adaptive estimation, and model selection for non-homogeneous Poisson processes by [33] can be modified to estimate Lévy densities, while preserving desirable features like *Oracle type inequalities* and the convergence of the mean-square error to 0 as the time horizon increases.

The second proposed method contemplates the problem that the Poisson jump measure cannot be retrieved from discrete observation and finds an approximation procedure for Poisson integrals using time series of the form $\{X(t_i)\}_{i=1}^{N}$. This approximation is based on the "microstructure" of Lévy processes; that is to say, the distributional and path properties of Lévy processes for small time spans. It is proved the weak convergence of the approximation to the actual integrals when the mesh of the partition approaches 0. Also, it is proved that the mean-square error of the estimation based on the approximate integrals converges to the mean square-error of the estimation based on the unattainable Poisson integrals. Other connections to the non-parametric estimation of density functions are also considered.

In Chapter 3, we address the performance of penalized projection estimators and model selection methods based on computer simulations. The considered estimators are histogram projection estimators and their approximate versions based on increments. We analyze in detail two classes of Lévy processes that are relevant in financial applications: Variance Gamma processes and Gamma Lévy processes. A projection estimation method with least-squares errors is considered to calibrate parametric or semiparametric models based on nonparametric estimation outputs. This methodology is applied to calibrate Gamma Lévy processes and compare with standard methods of maximum likelihood estimation.

In Chapter 4, we study the class of *tempered stable processes* introduced by Rosiński in [35]. We presente a survey of his results, providing our own proofs, and make some additional remarks.

CHAPTER II

NONPARAMETRIC ESTIMATION OF LÉVY DENSITIES

We discuss the nonparametric estimation of the Lévy density p of a real Lévy process $X = \{X(t)\}_{t\geq 0}$ based on discrete observations of the process. We develop two methods to accomplish this. The first method construct estimators $\hat{p}(x)$ which can be written in terms of integrals of deterministic functions with respect to the random measure associated with the jumps of X. Moreover, the first method consists of two parts: (1) the selection of a good estimator from a linear space of proposed Lévy densities, and (2) a data-driven selection of a linear model among a given collection of linear models. It is shown that the methodology of adaptive estimation, and model selection for nonhomogeneous Poisson processes (see [33]) can be modified to estimate Lévy densities, while preserving desirable features of estimation like Oracle type inequalities. The second method contemplates the fact that the Poisson jump measure is never observed and proposes an approximation procedure for Poisson integrals using time series of the form $\{X(t_i)\}_{i=1}^{N}$. It is proved the weak convergence of the approximation to the actual integrals when the mesh of the partition approaches 0. Also, it is proved the convergence of the mean-square error of estimation based on the approximations to the mean square-error of estimation based on the Poisson integrals.

2.1 The basic method of estimation

Consider a real Lévy process $X = \{X(t)\}_{t\geq 0}$ with Lévy density p. That is, X is a càdlàg process with independent and stationary increments such that its characteristic function is given by

$$\mathbb{E}\left[e^{iuX(t)}\right] = \exp\left[t\left(iub - \frac{u^2\sigma^2}{2} + \int\limits_{\mathbb{R}_0}\left\{e^{iux} - 1 - iux1_{[|x|\le 1]}\right\} p(x)dx\right)\right], \tag{2.1.1}$$

where $\mathbb{R}_0 = \mathbb{R}\backslash\{0\}$ and p satisfies

$$\int_{\mathbb{R}_0}(1 \wedge x^2)p(x)dx < \infty. \tag{2.1.2}$$

Being a càdlàg process, the set of jump times $\{t > 0 : X(t) - X(t^-) > 0\}$ is countable and, for Borel subsets B of $[0,\infty) \times \mathbb{R}_0$,

$$\mathcal{J}(B) = \#\left\{t > 0 : (t, X(t) - X(t^-)) \in B\right\}, \tag{2.1.3}$$

is a well-defined random measure on $[0,\infty) \times \mathbb{R}_0$. The Lévy-Itô decomposition of sample functions (see Theorem 19.2 of [39]) implies that $\mathcal{J}$ is a Poisson process on the Borel sets of $\mathcal{B}([0,\infty) \times \mathbb{R}_0)$ with mean measure given by

$$\mu(B) = \iint\limits_{B} p(x)\, dt\, dx. \tag{2.1.4}$$

We study the problem of estimating the Lévy density p on a domain $D \in \mathcal{B}(\mathbb{R}_0)$, where p is bounded and $\int_D p^2(x)dx < \infty$. For instance, if p is bounded outside of any neighborhood of the origin, (2.1.2) implies that for any $\varepsilon > 0$:

$$\int_{|x|>\varepsilon} p^2(x)dx < \infty. \tag{2.1.5}$$

More generally, let us assume that the Lévy measure $\nu(dx) \equiv p(x)dx$ is absolutely continuous with respect to a known measure η on $\mathcal{B}(D)$ and that the Radon-Nikodym derivative

$$\frac{d\nu}{d\eta}(x) = s(x), \quad x \in D, \tag{2.1.6}$$

is positive, bounded, and satisfies

$$\int_D s^2(x)\eta(dx) < \infty. \tag{2.1.7}$$

Definition 2.1.1 *If (2.1.6) and (2.1.7) are verified, we say that η is a* regularization measure *for the Lévy density p. In that case, s is referred to as the* regularized Lévy density *of p on D (under η).*

Our goal will be to estimate the regularized Lévy density s, and using (2.1.6) to proceed to retrieve p on D from s. Notice that under the previous regularization assumption, the measure $\mathcal{J}$ of (2.1.3), when restricted to $\mathcal{B}([0,\infty)\times D)$, is a Poisson process with mean measure

$$\mu(B) = \iint_B s(x)\,dt\,\eta(dx), \quad B \in \mathcal{B}([0,\infty)\times D). \tag{2.1.8}$$

Example 2.1.2 *The statistician could be interested in continuous densities p such that*

$$p(x) = O\left(\frac{1}{x}\right), \quad \text{as} \quad x \to 0.$$

This type of densities admit the regularization measure $\eta(dx) = x^{-2}dx$ on domains of the form $D = \{x : 0 < |x| < b\}$. Indeed, $s(x) = x^2 p(x)$ will be bounded and fulfills (2.1.7). The problem reduces to first estimate s and to subsequently estimate p by $x^{-2}s(x)$.

The general methodology we use is motivated by the recent procedures of *model selection* and *adaptive estimation* of the intensity function of non-homogeneous Poisson processes (see [33]). In this paper, the *projection estimator* is proposed as a plausible candidate for the intensity among a set of functions that constitute a finite dimensional subspace, whereas *penalized projection estimation* is devised as a data driven criterion for model selection among a family of linear models. One of the advantages of this approach compared to previous ones is to accomplish *Oracle inequalities* under quite general conditions (see Section 2.2 for a brief explanation of this type of inequalities). However, there are some drawbacks when facing domains of estimation with infinite measure as the domain $D = \{x : |x| > \varepsilon\}$ is under the Lebesgue measure or $D = \{x : 0 < |x| < b\}$ is under the measure $\eta(dx) = x^{-2}dx$. Actually, the total measure of the domain plays a key role in the definitions of projection estimators, contrast functions, and penalization. Our job in this section is to develop and heuristically justify a methodology that does not relies on the finiteness of the domain.

Let us describe the main ingredients of our procedure. Consider the random functional

$$\gamma_D(f) \equiv -\frac{2}{T} \iint\limits_{[0,T]\times D} f(x)\,\mathcal{J}(dt, dx) + \int\limits_D f^2(x)\,\eta(dx), \tag{2.1.9}$$

well defined for any function $f \in L^2((D,\eta))$, where $D \in \mathcal{B}(\mathbb{R}_0)$ and η is as in equations (2.1.6)-(2.1.8). Following [33], we call γ_D the *contrast function*. Throughout this section,

$$\|f\|^2 \equiv \int\limits_D f^2(x)\,\eta(dx),$$

for any $f \in L^2((D,\eta))$. Let $\mathcal{S}$ be a finite dimensional subspace of $L^2 = L^2((D,\eta))$. The *projection estimator* of s on S is defined by

$$\hat{s}(x) \equiv \sum_{i=1}^{d} \hat{\beta}_i \varphi_i(x), \tag{2.1.10}$$

where $\{\varphi_1, \ldots, \varphi_d\}$ is any orthonormal basis of $\mathcal{S}$ and

$$\hat{\beta}_i \equiv \frac{1}{T} \iint\limits_{[0,T]\times D} \varphi_i(x)\mathcal{J}(dt, dx). \tag{2.1.11}$$

Let us give another characterization of the projection estimator.

Remark 2.1.3 *The projection estimator is the unique minimizer of the contrast function γ_D over S. Indeed, plug $f = \sum_{i=1}^{d} \beta_i \varphi_i$ in (2.1.9) to get that $\gamma_D(f) = \sum_{i=1}^{d}\left(-2\beta_i\hat{\beta}_i + \beta_i^2\right)$, and thus, $\gamma_D(f) \geq -\sum_{i=1}^{d}\hat{\beta}_i^2$, for all $f \in S$. In particular, this characterization implies that $\hat{s}$ does not depend on the choice of the orthonormal basis, and suggests a mechanism to numerically approximate $\hat{s}$ when we do not have an explicit orthonormal basis for $\mathcal{S}$.*

The remark above helps to make sense of $\hat{s}$ as an estimator of the regularized Lévy density s because the minimizer of $\mathbb{E}[\gamma_D(f)]$ over all $f \in \mathcal{S}$ is precisely the closest function in $\mathcal{S}$ to s. Concretely, the *orthogonal projection* of s on the subspace $\mathcal{S}$, namely

$$s^{\perp} = \sum_{i=1}^{d} \left(\int_D \varphi_i(y) s(y) \eta(dy) \right) \varphi_i(x), \tag{2.1.12}$$

is such that

$$-\|s^{\perp}\|^2 = \mathbb{E}\left[\gamma_D(s^{\perp})\right] \le \mathbb{E}\left[\gamma_D(f)\right], \quad \forall f \in \mathcal{S}. \tag{2.1.13}$$

Moreover, by Proposition 1.2.13, we can corroborate that $\hat{s}$ is an unbiased estimator of the orthogonal projection $s^{\perp}$. In order to assess the quality of estimation, we compute the "square error" of $\hat{s}$:

$$\chi^2 \equiv \|s^{\perp} - \hat{s}\|^2 = \sum_{i=1}^{d} \left[\iint_{[0,T]\times D} \varphi_i(x) \frac{\mathcal{J}(dt, dx) - s(x)\, dt\, \eta(dx)}{T} \right]^2. \tag{2.1.14}$$

Then, using Proposition 1.2.13, the mean square error takes the form

$$\mathbb{E}\left[\chi^2\right] = \frac{1}{T} \sum_{i=1}^{d} \int_{D} \varphi_i^2(x) s(x)\, \eta(dx). \tag{2.1.15}$$

The quantity $\mathbb{E}\left[\chi^2\right]$ is called the *variance term* and the equation above shows that this term will shrink to 0 when the time horizon T goes to infinity. It is not hard to see that the *risk* of $\hat{s}$, $\mathbb{E}\left[\|s - \hat{s}\|^2\right]$, can be decomposed into a nonrandom term plus the variance term:

$$\mathbb{E}\left[\|s - \hat{s}\|^2\right] = \|s - s^{\perp}\|^2 + \mathbb{E}\left[\chi^2\right]. \tag{2.1.16}$$

The first term, called the *bias term*, accounts for the distance of the unknown function s to the model $\mathcal{S}$ and does not depend on the estimation criteria we use within the model.

The next natural problem to tackle is to design a data-driven scheme for selecting a "good" model from a collection of linear models $\{\mathcal{S}_m, m \in \mathcal{M}\}$. Namely, we wish to select a model that approximately realizes the best trade-off between the risk of estimation within the model and the distance of the unknown Lévy density to the model. Let $\hat{s}_m$ and $s_m^{\perp}$ be respectively the projection estimator and the orthogonal projection of s on $\mathcal{S}_m$. For each $m \in \mathcal{M}$, let χ_m^2 be as in (2.1.14). The following simplifications of the equation (2.1.16) will give us insight on a possible solution:

$$\begin{aligned}\mathbb{E}\left[\|s-\hat{s}_m\|^2\right] &= \|s-s_m^{\perp}\|^2+\mathbb{E}\left[\chi_m^2\right] \\ &= \|s\|^2-\|s_m^{\perp}\|^2+\mathbb{E}\left[\chi_m^2\right] \\ &= \|s\|^2-\mathbb{E}\left[\|\hat{s}_m\|^2\right]+2\mathbb{E}\left[\chi_m^2\right] \\ &= \|s\|^2+\mathbb{E}\left[\gamma_D\left(\hat{s}_m\right)+\mathrm{pen}(m)\right],\end{aligned} \tag{2.1.17}$$

where pen(m) is defined in terms of an orthonormal basis $\{\varphi_{1,m},\ldots,\varphi_{d_m,m}\}$ for $\mathcal{S}_m$ by the equation:

$$\mathrm{pen}(m)=\frac{2}{T^2}\iint_{[0,T]\times D}\left(\sum_{i=1}^{d_m}\varphi_{i,m}^2(x)\right)\mathcal{J}(dt,dx). \tag{2.1.18}$$

Equation (2.1.17) shows that the risk of $\hat{s}_m$ moves "parallel" to the expectation of the *observable statistics* $\gamma_D\left(\hat{s}_m\right)+\mathrm{pen}(m)$. This fact heuristically justifies to choose the model that minimizes such a penalized contrast value. We will show in a subsequent section that it is possible to take simpler penalty functions pen : $\mathcal{M}\to[0,\infty)$. In general, a **penalized projection estimator** (p.p.e.) is of the form

$$\tilde{s}\equiv\hat{s}_{\hat{m}}, \tag{2.1.19}$$

where $\hat{s}_m$ is defined as in (2.1.10) for each $m\in\mathcal{M}$, and where $\hat{m}$ is chosen so that $\gamma_D\left(\hat{s}_m\right)+$ pen(m) is minimal:

$$\hat{m}\equiv\mathrm{argmin}_{m\in\mathcal{M}}\left\{\gamma_D\left(\hat{s}_m\right)+\mathrm{pen}(m)\right\}.$$

Methods of estimation based on the minimization of penalty functions have a long history in the literature of regression and density estimation (for instance, [1], [25], and [40]). The general idea is to choose among a given collection of parametric models the model that minimizes a loss function plus a penalty term that controls the complexity of the model. The *nonparametric point of view* of penalized estimation has been promoted in the context of density estimation by Birgé and Massart (see [7] and references herein). In fact, the work on non-homogeneous Poisson processes by [33] is directly inspired by them. There are two main accomplishments obtained in these works both in the context of density estimation and intensity estimation of nonhomogeneous Poisson processes: Oracles inequalities and competitive performance against minimax estimators. The following section shows that the method outlined here preserves Oracle inequalities.

2.2 Oracle inequalities

Consider the problem of model selection among a collection of linear models, $\{\mathcal{S}_m, m \in \mathcal{M}\}$, for the regularized Lévy density s on D as outlined in Section 2.1. We showed through (2.1.17) that a sensible criterion to decide for a projection estimator is to penalize its contrast value with a properly chosen penalty function pen : $\mathcal{M} \to [0, \infty)$. Of course, the "best" model, namely

$$\bar{m} \equiv \operatorname{argmin}_{m \in \mathcal{M}} \mathbb{E}\left[\|s - \hat{s}_m\|^2\right], \tag{2.2.1}$$

is not accessible, but we can aspire to achieve the smallest possible risk up to a constant. In other words, it is desirable that our estimator $\tilde{s}$ comply with an inequality of the form

$$\mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C \inf_{m \in \mathcal{M}} \mathbb{E}\left[\|s - \hat{s}_m\|^2\right], \tag{2.2.2}$$

for a constant C "independent" of the linear models. The model that achieves the minimal risk of projection estimation is called the *Oracle model* and inequalities of the type (2.2.2) are called **Oracle inequalities**. "Approximate" Oracle inequalities were proved by Reynaud-Bouret [33] for the intensity function of a nonhomogeneous Poisson process $\{N_A\}_{A \in \mathcal{V}}$ on a measurable space $(\mathrm{V}, \mathcal{V})$. Concretely, she defines projection estimators $\hat{s}_m$ and penalized projection estimators $\tilde{s}$ satisfying

$$\mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C \inf_{m \in \mathcal{M}} \mathbb{E}\left[\|s - \hat{s}_m\|^2\right] + \frac{C'}{\zeta(V)},$$

where s is a bounded function and ζ is finite measure on V such that

$$\mu(A) \equiv \mathbb{E}[N_A] = \int_A s(\mathbf{x}) d\zeta(\mathbf{x}).$$

The finiteness of ζ plays an important role in her definitions and results, and it is not necessarily satisfied by the mean measure of the Poisson process $\mathcal{J}(\cdot)$ of (2.1.3) on $\mathcal{B}([0,T]\times D)$ (for instance, if $D = \{|x| > \varepsilon\}$ under $\zeta(d\mathbf{x}) = dx\,dt$ as in (2.1.4), or if $D = \{0 < |x| < b\}$ and $\zeta(d\mathbf{x}) = x^{-2}dx\,dt$ as in Example 2.1.2). In this section we show that, based on one sample of the Lévy process X on $[0,T]$, the projection estimators $\{\hat{s}_m\}_{m\in\mathcal{M}}$ and the penalized projection estimator $\tilde{s}$ described in Section 2.1 yield the approximate Oracle inequality

$$\mathbb{E}\left[\|s-\tilde{s}\|^2\right] \le C \inf_{m\in\mathcal{M}} \mathbb{E}\left[\|s-\hat{s}_m\|^2\right] + \frac{C'}{T}.$$

for the Lévy measure s under suitable chosen penalization functions. The proof we present essentially follows the same line of reasoning as [33]; however, to overcome the possible lack of finiteness in η and avoid unnecessary use of upper bounds, we include a new element in the penalization functions which is also appealing: the dimension of the linear model. We also address the problem of estimating the order of the constants C and C' appearing in the Oracle inequality.

The following regularity condition was introduced by [33] to make a distinction between not too "large" families of linear models and wavelet-type linear models. We will focus here on the simplest case.

Definition 2.2.1 *A collection of models $\{\mathcal{S}_m, m\in\mathcal{M}\}$ is said to be polynomial if there exist constants $\Gamma > 0$ and $R \ge 0$ such that for every positive integer n*

$$\#\{m\in\mathcal{M} : d_m = n\} \le \Gamma n^R,$$

where d_m stands for the dimension of the model S_m, while # denotes cardinality.

We assume below the setting of Section 2.1; that is to say, $X = \{X(t)\}_{t\ge 0}$ is a Lévy process with Lévy density p and *regularized* Lévy density s on a domain $D \in \mathcal{B}(\mathbb{R}_0)$ under a *regularization measure* η (see Definition 2.1.1). Define

$$D_m = \sup\left\{\|f\|_\infty^2 : f\in S_m, \|f\|^2 \equiv \int_D f^2(x)\eta(dx) = 1\right\}. \tag{2.2.3}$$

Remark 2.2.2 *If $\{\varphi_{1,m}, \ldots, \varphi_{d_m,m}\}$ is an orthonormal basis of $\mathcal{S}_m$, then $D_m = \|\sum_{i=1}^{d_m} \varphi_{i,m}^2\|_\infty$ (see Section 2.5 for a verification).*

Here is the main result of this section:

Theorem 2.2.3 *Let $\{\mathcal{S}_m, m \in \mathcal{M}\}$ be a polynomial family of finite dimensional linear subspaces of $L^2((D, \eta))$. Let T be large enough so that $D_m \leq T$, for all $m \in \mathcal{M}$. If $\hat{s}_m$ and $s_m^\perp$ are respectively the projection estimator and the orthogonal projection of the regularized Lévy density s on $\mathcal{S}_m$ then, the penalized projection estimator $\tilde{s}$ of (2.1.19) is such that*

$$\mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C \inf_{m \in \mathcal{M}} \left\{\|s - s_m^\perp\|^2 + \mathbb{E}\left[\mathrm{pen}(m)\right]\right\} + \frac{C'}{T}, \tag{2.2.4}$$

whenever $\mathrm{pen} : \mathcal{M} \to [0, \infty)$ *takes one of the following forms for constants $c > 1$, $c' > 0$, and $c'' > 0$:*

(a) $\mathrm{pen}(m) \geq c\frac{D_m \mathcal{N}}{T^2} + c'\frac{d_m}{T}$, *where $\mathcal{N} \equiv \mathcal{J}([0, T] \times D)$ is the number of jumps prior to T with sizes falling in D and where it is assumed $\rho \equiv \int_D s(x)\eta(dx) < \infty$;*

(b) $\mathrm{pen}(m) \geq c\frac{\hat{V}_m}{T}$, *where $\hat{V}_m$ is defined in terms of an orthonormal basis $\{\varphi_{1,m}, \ldots, \varphi_{d_m,m}\}$ of $\mathcal{S}_m$ by the equation:*

$$\hat{V}_m \equiv \frac{1}{T} \iint_{[0,T] \times D} \left(\sum_{i=1}^{d_m} \varphi_{i,m}^2(x)\right) \mathcal{J}(dt, dx), \tag{2.2.5}$$

and where it is assumed $\beta \equiv \inf_{m \in \mathcal{M}} \frac{\mathbb{E}[\hat{V}_m]}{D_m} > 0$ and $\phi \equiv \inf_{m \in \mathcal{M}} \frac{D_m}{d_m} > 0$;

(c) $\mathrm{pen}(m) \geq c\frac{\hat{V}_m}{T} + c'\frac{D_m}{T} + c''\frac{d_m}{T}$.

Moreover, the constant C depends only on c, c' and c'', while C' varies with c, c', c'', Γ, R, $\|s\|$, $\|s\|_\infty$, ρ, β, and ϕ.

The next corollary immediately follows from the first equality in (2.1.17), equation (2.1.15), and part (b) above:

Corollary 2.2.4 *In the setting of Theorem 2.2.3, if the penalty function is of the form* $\mathrm{pen}(m) \equiv c\frac{\hat{V}_m}{T}$, *for every $m \in \mathcal{M}$, $\beta > 0$, and $\phi > 0$, then*

$$\mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C_1 \inf_{m \in \mathcal{M}} \left\{\mathbb{E}\left[\|s - \hat{s}_m\|^2\right]\right\} + \frac{C_2}{T}, \tag{2.2.6}$$

for a constant C_1 depending only on c, and a constant C_2 depending on c, Γ, R, $\|s\|$, $\|s\|_\infty$, β, and ϕ.

We will break the proof of Theorem 2.2.3 into several preliminary results.

Lemma 2.2.5 *For any penalty function* $\mathrm{pen} : \mathcal{M} \to [0, \infty)$ *and any $m \in \mathcal{M}$, the penalized projection estimator $\tilde{s}$ satisfies*

$$\|s - \tilde{s}\|^2 \leq \|s - s_m^{\perp}\|^2 + 2\chi_{\hat{m}}^2 + 2\nu_D\left(s_{\hat{m}}^{\perp} - s_m^{\perp}\right) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}), \tag{2.2.7}$$

where $\chi_m^2 \equiv \|s_m^{\perp} - \hat{s}_m\|^2$ and where the functional $\nu_D : L^2\left((D, \eta)\right) \to \mathbb{R}$ is defined by

$$\nu_D(f) \equiv \iint_{[0,T]\times D} f(x)\, \frac{\mathcal{J}(dt, dx) - s(x)\, dt\, \eta(dx)}{T}. \tag{2.2.8}$$

The general idea to obtain (2.2.4) is to bound the unattainable terms $\chi_{\hat{m}}^2$ and $\nu_D\left(s_{\hat{m}}^{\perp} - s_m^{\perp}\right)$ in the right hand side of (2.2.7) by observable statistics. Then, the form of $\mathrm{pen}(\cdot)$ will be determined by this observable statistics so that the right hand side in (2.2.7) does not involve $\hat{m}$. To carry out this plan, we use concentration inequalities for $\chi_{\hat{m}}^2$ and for the compensated Poisson integrals $\nu_D(f)$. The following result gives a concentration inequality for general compensated Poisson integrals.

Theorem 2.2.6 *Let N be a Poisson process on a measurable space* $(\mathrm{V}, \mathcal{V})$ *with mean measure μ and let $f : \mathrm{V} \to \mathbb{R}$ be an essentially bounded measurable function satisfying $0 < \int_{\mathrm{V}} f^2(v)\mu(dv)$ and $\int_{\mathrm{V}} |f(v)|\mu(dv) < \infty$. Then, for any $u > 0$,*

$$\mathbb{P}\left[\int_{\mathrm{V}} f(v)(N(dv) - \mu(dv)) \geq \|f\|_{L^2(\mu)} \sqrt{2u} + \frac{1}{3}\|f\|_\infty u\right] \leq e^{-u}, \tag{2.2.9}$$

where $\|f\|_{L^2(\mu)}^2 = \int_{\mathrm{V}} f^2(v)\mu(dv)$. In particular, if $f : \mathrm{V} \to [0, \infty)$ then, for any $\epsilon > 0$ and $u > 0$,

$$\mathbb{P}\left[(1+\varepsilon)\left(\int_{\mathrm{V}} f(v)N(dv) + \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right)\|f\|_\infty u\right) \geq \int_{\mathrm{V}} f(v)\mu(dv)\right] \geq 1 - e^{-u}. \tag{2.2.10}$$

The inequality (2.2.9) is proved in [18] (see also Proposition 7 of [33]). A verification of (2.2.10) is provided in Section 2.5.

The next result allow us to bound the Poisson functional χ_m^2. This results is essentially Proposition 9 of [33].

Lemma 2.2.7 *Let N be a Poisson process on a measurable space $(\mathrm{V}, \mathcal{V})$ with mean measure $\mu(dv) = p(v)\eta(dv)$ and intensity function $p \in L^2(\mathrm{V}, \mathcal{V}, \eta)$. Let $\mathcal{S}$ be a finite dimensional subspace of $L^2(\mathrm{V}, \mathcal{V}, \eta)$ with orthonormal basis $\{\tilde{\varphi}_1, \dots, \tilde{\varphi}_d\}$, and define*

$$\hat{p}(v) \equiv \sum_{i=1}^{d} \left(\int_{\mathrm{V}} \tilde{\varphi}_i(w) N(dw) \right) \tilde{\varphi}_i(v) \tag{2.2.11}$$

$$p^{\perp}(v) \equiv \sum_{i=1}^{d} \left(\int_{\mathrm{V}} p(w) \tilde{\varphi}_i(w) \eta(dw) \right) \tilde{\varphi}_i(v). \tag{2.2.12}$$

Then, $\chi^2(\mathcal{S}) \equiv \|\hat{p} - p^{\perp}\|^2_{L^2(\eta)}$ is such that for any $u > 0$ and $\varepsilon > 0$

$$\mathbb{P}\left[\chi(\mathcal{S}) \geq (1+\varepsilon)\sqrt{\mathbb{E}\left[\chi^2(\mathcal{S})\right]} + \sqrt{2k\mathrm{M}_{\mathcal{S}}u} + k(\varepsilon)\mathrm{B}_{\mathcal{S}}u\right] \leq e^{-u}, \tag{2.2.13}$$

where we can take $k = 6$, $k(\varepsilon) = 1.25 + 32/\varepsilon$, and where

$$M_{\mathcal{S}} \equiv \sup\left\{ \int_{\mathrm{V}} f^2(v) p(v) \eta(dv) : f \in \mathcal{S}, \|f\|_{L^2(\eta)} = 1 \right\}, \tag{2.2.14}$$

$$B_{\mathcal{S}} \equiv \sup\left\{ \|f\|_{\infty} : f \in \mathcal{S}, \|f\|_{L^2(\eta)} = 1 \right\}. \tag{2.2.15}$$

Following the same strategy as [33], the idea is to obtain first concentration inequality of the form

$$\mathbb{P}\left[\|s - \tilde{s}\|^2 \leq C\left(\|s - s_m^{\perp}\|^2 + \mathrm{pen}(m)\right) + h(\xi)\right] \geq 1 - C'e^{-\xi},$$

for constants C and C', and a function $h(\xi)$ (all independent of m). This will prove to be enough in view of the following result (see Section 2.5 for a proof).

Lemma 2.2.8 *Let $h : [0, \infty) \to \mathbb{R}_+$ be an strictly increasing function with continuous derivative and such that $h(0) = 0$ and $\lim_{\xi \to \infty} e^{-\xi} h(\xi) = 0$. If Z is random variable satisfying*

$$\mathbb{P}\left[Z \geq h(\xi)\right] \leq Ke^{-\xi},$$

for every $\xi > 0$, then

$$\mathbb{E}Z \leq K \int_0^\infty e^{-u} h(u) du.$$

We are now in position to prove the main result of this section. Throughout the proof, we shall have to introduce various constants and inequalities that will hold with high probability. In order to clarify the role that the constants play in these inequalities, we shall make some conventions and give to the letters x, y, f, a, b, ξ, $\mathcal{K}$, c, and C, with various sub- or superscripts, special meaning. The letters with x are reserved to denote positive constants that can be chosen arbitrarily. The letters with y denote arbitrary constants greater than 1. $f, f_1, f_2, \ldots$ denote quadratic polynomials of a variable ξ whose coefficients (denoted by $a's$ and $b's$) are determined by the values of the $x's$ and $y's$. The inequalities will be true with probabilities greater that $1 - \mathcal{K}e^{-\xi}$, where $\mathcal{K}$ is determined by the values of the $x's$ and the $y's$. Finally, $c's$ and $C's$ are used for constants constrained by the $x's$ and $y's$. *It is important to remember that the constants in a given inequality are only used in that inequality*. The pair of equivalent inequalities below will be repeatedly used through the proof:

$$\begin{aligned} &\text{(i)}\;\; 2ab \leq xa^2 + \tfrac{1}{x} b^2, \quad \text{and} \\ &\text{(ii)}\;\; (a+b)^2 \leq (1+x)\, a^2 + \left(1 + \tfrac{1}{x}\right) b^2, \quad (\text{for } x > 0). \end{aligned} \tag{2.2.16}$$

Proof of Theorem 2.2.3: We consider successive improvements of the inequality (2.2.7):

Inequality 1: For any positive constants x_1, x_2, x_3, and x_4, there is a positive number $\mathcal{K}$ and an increasing quadratic function $f(\xi)$ (both independent of the family of linear models and of T) such that, with probability larger than $1 - \mathcal{K}e^{-\xi}$,

$$\begin{aligned} \|s - \tilde{s}\|^2 \;\; &\leq \|s - s_m^\perp\|^2 + 2\chi_{\hat{m}}^2 + 2x_1 \|s_{\hat{m}}^\perp - s_m^\perp\|^2 \\ &\quad + x_2 \tfrac{D_{\hat{m}}}{T} + x_3 \tfrac{D_m}{T} + x_4 \tfrac{d_{\hat{m}}}{T} \\ &\quad + \text{pen}(m) - \text{pen}(\hat{m}) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{2.2.17}$$

Verification: Let us find an upper bound for $\nu_D\left(s_{m'}^\perp - s_m^\perp\right)$, $m', m \in \mathcal{M}$. Since the operator ν_D defined by (2.2.8) is just a compensated integral with respect to a Poisson process with mean measure $\mu(dtdx) = dt\eta(dx)$, we can apply Theorem 2.2.6 to obtain that, for any $x'_{m'} > 0$, and with probability larger than $1 - e^{-x'_{m'}}$

$$\nu_D\left(s^{\perp}_{m'} - s^{\perp}_{m}\right) \leq \left\|\frac{s^{\perp}_{m'} - s^{\perp}_{m}}{T}\right\|_{L^2(\mu)} \sqrt{2x'_{m'}} + \frac{\|s^{\perp}_{m'} - s^{\perp}_{m}\|_{\infty} x'_{m'}}{3T}. \tag{2.2.18}$$

In that case, the probability that (2.2.18) holds for every $m' \in \mathcal{M}$ is larger than $1 - \sum_{m' \in \mathcal{M}} e^{-x_{m'}}$ because $P(A \cap B) \geq 1 - a - b$, whenever $P(A) \geq 1 - a$ and $P(B) \geq 1 - b$. Clearly,

$$\begin{aligned}\left\|\frac{s^{\perp}_{m'} - s^{\perp}_{m}}{T}\right\|^2_{L^2(\mu)} &= \iint_{[0,T]\times D} \left(\frac{s^{\perp}_{m'}(x) - s^{\perp}_{m}(x)}{T}\right)^2 s(x)dt\eta(dx) \\ &\leq \|s\|_{\infty} \frac{\|s^{\perp}_{m'} - s^{\perp}_{m}\|^2}{T}.\end{aligned}$$

Using (2.2.16-i), the first term on the right hand side of (2.2.18) is then bounded as follows:

$$\left\|\frac{s^{\perp}_{m'} - s^{\perp}_{m}}{T}\right\|_{L^2(\mu)} \sqrt{2x'_{m'}} \leq x_1 \|s^{\perp}_{m'} - s^{\perp}_{m}\|^2 + \frac{\|s\|_{\infty} x_{m'}}{2Tx_1}, \tag{2.2.19}$$

for any $x_1 > 0$. Using (2.2.3) and (2.2.16-i),

$$\begin{aligned}\|s^{\perp}_{m'} - s^{\perp}_{m}\|_{\infty} x'_{m'} &\leq \left(\|s^{\perp}_{m'}\|_{\infty} + \|s^{\perp}_{m}\|_{\infty}\right) x'_{m'} \\ &\leq \left(\sqrt{D_{m'}}\|s^{\perp}_{m'}\| + \sqrt{D_m}\|s^{\perp}_{m}\|\right) x_{m'} \\ &\leq \sqrt{D_{m'}}\|s\| x'_{m'} + \sqrt{D_m}\|s\| x'_{m'} \\ &\leq 3x_2 D_{m'} + 3x_3 D_m + \frac{\|s\|^2 x'^2_{m'}}{12}\left(\frac{1}{x_2} + \frac{1}{x_3}\right),\end{aligned}$$

for all $x_2 > 0$, $x_3 > 0$. It follows that, for any $x_1 > 0$, $x_2 > 0$, and $x_3 > 0$,

$$\begin{aligned}\nu_D\left(s^{\perp}_{m'} - s^{\perp}_{m}\right) \leq x_1 \|s^{\perp}_{m'} - s^{\perp}_{m}\|^2 + x_2 \frac{D_{m'}}{T} + x_3 \frac{D_m}{T} \\ + \frac{\|s\|_{\infty} x'_{m'}}{2Tx_1} + \frac{\|s\|^2 x'^2_{m'}}{36Tc},\end{aligned}$$

where we set $\frac{1}{c} = \frac{1}{x_2} + \frac{1}{x_3}$. Next, take

$$x'_{m'} \equiv x_4 \sqrt{d_{m'}} \left(\frac{1}{\|s\|} \wedge \frac{1}{\|s\|_{\infty}}\right) + \xi.$$

Then, for any positive x_1, x_2, x_3, and x_4, there is a $\mathcal{K}$ and a function f such that, with probability greater than $1 - \mathcal{K}e^{-\xi}$,

$$\begin{aligned} \nu_D\left(s_{m'}^{\perp} - s_m^{\perp}\right) &\leq x_1\|s_{m'}^{\perp} - s_m^{\perp}\|^2 + x_2\tfrac{D_{m'}}{T} + x_3\tfrac{D_m}{T} \\ &\quad + \left(\tfrac{x_4^2}{18c} + \tfrac{x_4}{2x_1}\right)\tfrac{d_{m'}}{T} + \tfrac{f(\xi)}{T}, \quad \forall m' \in \mathcal{M}. \end{aligned} \tag{2.2.20}$$

Concretely,

$$\begin{aligned} f(\xi) &= \tfrac{\|s\|}{18c}\xi^2 + \tfrac{\|s\|_\infty}{2x_1}\xi, \\ \mathcal{K} &= \Gamma \textstyle\sum_{n=1}^{\infty} n^R \exp\left(-\sqrt{n}x_4\left(\tfrac{1}{\|s\|} \wedge \tfrac{1}{\|s\|_\infty}\right)\right). \end{aligned} \tag{2.2.21}$$

Here, we use the assumption of polynomial models (Definition2.2.1) to come up with the constant $\mathcal{K}$. Pluging (2.2.20) in (2.2.7), and renaming the coefficient of $d_{m'}/T$, we can corroborate inequality 1.

Inequality 2: For any positive constants $y_1 > 1$, x_1, x_2, and x_3, there are positive constants $C_1 < 1$, $C_1' > 1$, and $\mathcal{K}$, and a strictly increasing quadratic polynomial f (all independent of the class of linear models and T) such that with probability larger than $1 - \mathcal{K}e^{-\xi}$,

$$\begin{aligned} C_1\|s - \tilde{s}\|^2 &\leq C_1'\|s - s_m^{\perp}\|^2 + y_1\chi_{\hat{m}}^2 \\ &\quad + x_2\tfrac{D_{\hat{m}}}{T} + x_3\tfrac{D_m}{T} + x_4\tfrac{d_{\hat{m}}}{T} \\ &\quad + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{2.2.22}$$

Moreover, if $1 < y_1 < 2$, then $C_1' = 3 - y_1$ and $C_1 = y_1 - 1$. If $y_1 \geq 2$, then $C_1' = 1 + 4x_1$ and $C_1 = 1 - 4x_1$, where x_1 is any positive constant related to f according to equation (2.2.21).

Verification: Let us combine the term on the left hand side of (2.2.17) with the first three terms on the right hand side. Using the triangle inequality followed by (2.2.16-ii),

$$\|s_{\hat{m}}^{\perp} - s_m^{\perp}\|^2 \leq 2\|s - s_m^{\perp}\|^2 + 2\|s_{\hat{m}}^{\perp} - s\|^2.$$

Then, since $\chi_{\hat{m}}^2 = \|s_{\hat{m}}^{\perp} - \hat{s}_{\hat{m}}\|^2$, and $\|s_{\hat{m}}^{\perp} - s\|^2 = \|s - \hat{s}_{\hat{m}}\|^2 - \|s_{\hat{m}}^{\perp} - \hat{s}_{\hat{m}}\|^2$, it is found that

$$\begin{aligned} &\|s - s_m^{\perp}\|^2 + 2\chi_{\hat{m}}^2 + 2x_1\|s_{\hat{m}}^{\perp} - s_m^{\perp}\|^2 - \|s - \tilde{s}\|^2 \\ &\leq (1 + 4x_1)\,\|s - s_m^{\perp}\|^2 + (2 - 4x_1)\,\|s_{\hat{m}}^{\perp} - \hat{s}_{\hat{m}}\|^2 \\ &\quad + (4x_1 - 1)\,\|s - \tilde{s}\|^2, \end{aligned}$$

for every $x_1 > 0$. Then, for any $y_1 > 1$, there are positive constant C, $C_1' > 1$, and $C_1 < 1$ such that

$$\begin{aligned} &\|s - s_m^\perp\|^2 + 2\chi_{\hat{m}}^2 + 2C\|s_{\hat{m}}^\perp - s_m^\perp\|^2 - \|s - \tilde{s}\|^2 \\ \leq\ & C_1'\|s - s_m^\perp\|^2 + y_1\chi_{\hat{m}}^2 - C_1\|s - \tilde{s}\|^2. \end{aligned} \tag{2.2.23}$$

Combining (2.2.17) and (2.2.23), we obtain (2.2.22).

Inequality 3: For any $y_2 > 1$ and positive constants x_i, $i = 2, 3, 4$, there exist positive numbers $C_2 < 1$, $C_2' > 1$, an increasing quadratic polynomial of the form $f_2(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K}_2 > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}_2 e^{-\xi}$,

$$\begin{aligned} C_2\|s - \tilde{s}\|^2 \quad & \leq C_2'\|s - s_m^\perp\|^2 \\ & +y_2\tfrac{V_{\hat{m}}}{T} + x_2\tfrac{D_{\hat{m}}}{T} + x_3\tfrac{d_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ & +x_4\tfrac{D_m}{T} + \mathrm{pen}(m) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{2.2.24}$$

Verification: We bound $\chi_{m'}^2$ using Lemma 2.2.7 with $\mathrm{V} = \mathbb{R}_+ \times D$ and $\mu(d\mathbf{x}) = s(x)dt\eta(dx)$. We regard the linear model $\mathcal{S}_m$ as a subspace of $L^2(\mathbb{R}_+ \times D, dt\eta(dx))$ with orthonormal basis $\left\{\frac{\varphi_{1,m}}{\sqrt{T}}, \ldots, \frac{\varphi_{d_m,m}}{\sqrt{T}}\right\}$. Recall that

$$\chi_m^2 = \|s_m^\perp - \hat{s}_m\|^2 = \sum_{i=1}^{d}\left[\ \iint_{[0,T]\times D} \varphi_{i,m}(x)\frac{\mathcal{J}(dt, dx) - s(x)dt\eta(dx)}{T}\ \right]^2.$$

Then, with probability larger than $1 - \sum_{m' \in \mathcal{M}} e^{-x'_{m'}}$,

$$\sqrt{T}\chi_{m'} \leq (1 + x_1)\sqrt{V_{m'}} + \sqrt{2k\mathrm{M}_{m'}x'_{m'}} + k(x_1)\mathrm{B}_{m'}x'_{m'}, \tag{2.2.25}$$

for every $m' \in \mathcal{M}$, where $B_{m'} = \sqrt{D_{m'}/T}$,

$$\begin{aligned} V_{m'} &\equiv \int_D \left(\sum_{i=1}^{d_m} \varphi_{i,m}^2(x)\right) s(x)\eta(dx), \quad \text{and} \\ M_{m'} &\equiv \sup\left\{\int_D f^2(x)s(x)\eta(dx) : f \in \mathcal{S}_{m'}, \|f\| = 1\right\}. \end{aligned} \tag{2.2.26}$$

Since $\int_D f^2(x)s(x)\eta(dx) \leq \|f\|_\infty \|s\|$, $M_{m'}$ is bounded above by $\|s\|\sqrt{D_{m'}}$. In that case, we can use (2.2.16-i) to obtain

$$\sqrt{2k\mathrm{M}_{m'}x'_{m'}} \leq x_2\sqrt{D_{m'}} + \frac{k\|s\|}{2x_2}x'_{m'},$$

for any $x_2 > 0$. On the other hand, by hypothesis $D_{m'} \leq T$, and (2.2.25) implies that

$$\sqrt{T}\chi_{m'} \leq (1+x_1)\sqrt{V_{m'}} + x_2\sqrt{D_{m'}} + \left(\frac{k\|s\|}{2x_2} + k(x_1)\right) x'_{m'},$$

where the constants $x'_{m'}$ are chosen as

$$x'_{m'} = \frac{x_3\sqrt{d_{m'}}}{\frac{k\|s\|}{2x_2} + k(x_1)} + \xi.$$

Then, for any $x_1 > 0$, $x_2 > 0$, $x_3 > 0$, and $\xi > 0$,

$$\sqrt{T}\chi_{m'} \leq (1+x_1)\sqrt{V_{m'}} + x_2\sqrt{D_{m'}} + x_3\sqrt{d_{m'}} + f_1(\xi), \tag{2.2.27}$$

with probability larger than $1 - \mathcal{K}_1 e^{-\xi}$, where $\mathcal{K}_1$ is determined by the Polynomial property and where

$$f_1(\xi) = \left(\frac{k\|s\|}{2x_2} + k(x_1)\right)\xi.$$

Squaring (2.2.27) and using (2.2.16-ii) repeatedly, we conclude that, for any $y > 1$, $x_2 > 0$, and $x_3 > 0$, there are both a constant $\mathcal{K}_1 > 0$ and a quadratic function of the form $f_2(\xi) = a\xi^2$ (independent of T, m', and the family of linear models) such that, with probability greater than $1 - \mathcal{K}_1 e^{-\xi}$,

$$\chi^2_{m'} \leq y\frac{V_{m'}}{T} + x_2\frac{D_{m'}}{T} + x_3\frac{d_{m'}}{T} + \frac{f_2(\xi)}{T}, \quad \forall m' \in \mathcal{M}. \tag{2.2.28}$$

Then, (2.2.24) immediately follows from (2.2.28) and (2.2.22).

Proof of (2.2.4) for case (c):

By the inequality (2.2.10), we can upper bound $V_{m'}$ by $\hat{V}_{m'}$ on an event of large probability. Namely, for every $x'_{m'} > 0$ and $x > 0$, with probability greater than $1 - \sum_{m' \in \mathcal{M}} e^{-x'_{m'}}$

$$(1+x)\left(\hat{V}_{m'} + \left(\frac{1}{2x} + \frac{5}{6}\right)\frac{D_{m'}}{T}x'_{m'}\right) \geq V_{m'}, \quad \forall m' \in \mathcal{M}, \tag{2.2.29}$$

(recall that $D_m = \| \sum_{i=1}^{d_m} \varphi^2_{i,m} \|_\infty$). Since by hypothesis $D_{m'} < T$, and choosing

$$x'_{m'} = x' d_{m'} + \xi, \quad (x' > 0),$$

it is seen that for any $x > 0$ and $x_4 > 0$, there are a positive constant $\mathcal{K}_2$ and a function $f(\xi) = b\xi$ (independent of T and of the linear models) such that with probability greater than $1 - \mathcal{K}_2 e^{-\xi}$

$$(1 + x)\hat{V}_{m'} + x_4 d_{m'} + f(\xi) \geq V_{m'}, \quad \forall m' \in \mathcal{M}. \tag{2.2.30}$$

Here, we get $\mathcal{K}_2$ from the Polynomial assumption on the class of models. Combining (2.2.30) and (2.2.24), it is clear that for any $y > 1$, and positive x_i, $i = 1, 2, 3$, we can choose a pair of positive constants $C_1 < 1$, $C_1' > 1$, an increasing quadratic polynomial of the form $f(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}e^{-\xi}$

$$\begin{aligned} C_1\|s - \tilde{s}\|^2 \quad & \leq C_1'\|s - s_m^\perp\|^2 \\ & +y\tfrac{\hat{V}_{\hat{m}}}{T} + x_1\tfrac{D_{\hat{m}}}{T} + x_2\tfrac{d_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ & +x_3\tfrac{D_m}{T} + \mathrm{pen}(m) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{2.2.31}$$

Next, we take $y = c$, $x_1 = c'$, and $x_2 = c''$ to cancel $-pen(\hat{m})$ in (2.2.31). By Lemma 2.2.8, it follows that

$$C_1\mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C_1'\|s - s_m^\perp\|^2 + \left(1 + \frac{x_3}{c'}\right)\mathbb{E}\left[\mathrm{pen}(m)\right] + \frac{C_1''}{T}. \tag{2.2.32}$$

Since m is arbitrary, we obtain the case (c) of (2.2.4).

Proof of (2.2.4) for case (a):

By Remark 2.2.2, we can bound $V_{m'}$, as given in (2.2.26), by $D_{m'}\rho$ (assuming that $\rho < \infty$). On the other hand, (2.2.10) implies that

$$(1 + x_1)\left(\frac{\mathcal{N}}{T} + \left(\frac{1}{2x_1} + \frac{5}{6}\right)\frac{\xi}{T}\right) \geq \rho, \tag{2.2.33}$$

with probability greater than $1 - e^{-\xi}$. Using these bounds for $V_{m'}$ and the assumption that $D_{m'} \leq T$, (2.2.24) reduces to

$$\begin{aligned} C_1\|s - \tilde{s}\|^2 \quad & \leq C_1'\|s - s_m^\perp\|^2 \\ & +y\tfrac{D_{\hat{m}}\mathcal{N}}{T^2} + x_1\tfrac{d_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ & +x_2\tfrac{D_m\mathcal{N}}{T^2} + \mathrm{pen}(m) + \tfrac{f(\xi)}{T}, \end{aligned} \tag{2.2.34}$$

which is valid with probability $1 - \mathcal{K}e^{-\xi}$. In (2.2.34), $y > 1$, $x_1 > 0$ and $x_2 > 0$ are arbitrary, while C_1, C_1', the increasing quadratic polynomial of the form $f(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ are determined by y, x_1, and x_2 independently of the family of linear models and of T. We point out that we divided and multiplied by ρ the terms $D_{\hat{m}}/T$ and D_m/T in (2.2.24), and then applied (2.2.33) to get (2.2.34). It is now clear that $y = c$, and $x_1 = c'$ will produce the desired cancelation.

Proof of (2.2.4) for case (b):

We first upper bound $D_{\hat{m}}$ by $\beta^{-1}V_{\hat{m}}$ and $d_{\hat{m}}$ by $(\beta\phi)^{-1}V_{\hat{m}}$ in the inequality (2.2.24):

$$\begin{aligned} C_1\|s - \tilde{s}\|^2 \quad &\leq C_1'\|s - s_m^{\perp}\|^2 + \left(y + x_1\beta^{-1} + x_2(\beta\phi)^{-1}\right)\tfrac{V_{\hat{m}}}{T} \\ &\qquad -\mathrm{pen}(\hat{m}) + x_3\beta^{-1}\tfrac{V_m}{T} + \mathrm{pen}(m) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{2.2.35}$$

Then, using $d_{m'} \leq (\beta\phi)^{-1}V_{m'}$ in (2.2.30) and letting $x_4(\beta\phi)^{-1}$ vary on $(0, 1)$, we verify that for any $x' > 0$, a positive constant $\mathcal{K}_4$ and a polynomial f can be found so that with probability greater than $1 - \mathcal{K}_4e^{-\xi}$,

$$(1 + x')\hat{V}_{m'} + f(\xi) \geq V_{m'}, \quad \forall m' \in \mathcal{M}. \tag{2.2.36}$$

Putting together (2.2.36) and (2.2.35), it is clear that for any $y > 1$ and $x_1 > 0$, we can find a pair of positive constants $C_1 < 1$, $C_1' > 1$, an increasing quadratic polynomial of the form $f(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}e^{-\xi}$,

$$\begin{aligned} C_1\|s - \tilde{s}\|^2 \quad &\leq C_1'\|s - s_m^{\perp}\|^2 + y\tfrac{\hat{V}_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ &\qquad + x_1\tfrac{V_m}{T} + \mathrm{pen}(m) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{2.2.37}$$

In particular, by taking $y = c$, the term $-pen(\hat{m})$ cancels out. Lemma 2.2.8 implies that

$$C_1\mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C_1'\|s - s_m^{\perp}\|^2 + (1 + x_1)\,\mathbb{E}\left[\mathrm{pen}(m)\right] + \frac{C_1''}{T}. \tag{2.2.38}$$

Finally, (2.2.4) (b) follows since m is arbitrary. $\square$

Remark 2.2.9 *Let us analyze more carefully the values that the constants C and C' can take in the inequality (2.2.4). For instance, consider the penalty function of part (c). As* *we saw in (2.2.32), the constants C and C' are determined by C_1, C_1', C_1'', and x_3. The constant C_1 was proved to be $y_1 - 1$ if $y_1 < 2$, while it can be made arbitrarily close to one otherwise (see the comment immediately after (2.2.22)). On the other hand, y_1 itself can be made arbitrarily close to the c of (2.2.31) by taking x small enough in (2.2.29) and y close to 1 in (2.2.28). Then, when $c \geq 2$, C_1 can be made arbitrarily close to one at the cost of increasing C_1'' in (2.2.32). Similarly, paying the same cost, we are able to select C_1' as close to one as we wish and x_3 arbitrarily small. Therefore, it is possible to find for any $\varepsilon > 0$, a constant $C'(\varepsilon)$ (increasing in ε) so that*

$$\mathbb{E}\|s - \tilde{s}\|^2 \leq (1+\varepsilon) \inf_{m \in \mathcal{M}} \left\{ \|s - s_m^{\perp}\|^2 + \mathbb{E}\left[\mathrm{pen}(m)\right] \right\} + \frac{C'(\varepsilon)}{T}. \tag{2.2.39}$$

A more thorough inspection shows that

$$\lim_{\varepsilon \to 0} C'(\varepsilon)\varepsilon = K,$$

where K depends only c, c', c'', Γ, R, $\|s\|$, and $\|s\|_\infty$. The same reasoning apply to the other two types of penalty functions when $c \geq 2$. In particular, we point out that C_1 can be made arbitrarily close to 2 in the Oracle inequality (2.2.6) at the price of having a large C_2 constant.

2.3 Calibration based on discrete time data: approximation of Poisson integrals

One drawback to the method outlined in Section 2.1 is that in general we do not observe the jumps of a Lévy process $X = \{X(t)\}_{t \geq 0}$. In practice, we can aspire to sample the process $X(t)$ at discrete times, but we are neither able to measure the size of the jumps $\Delta X(t) \equiv X(t) - X(t^-)$ nor the times of jumps $\{t : \Delta X(t) > 0\}$. Poisson integrals of the type

$$I(f) \equiv \iint_{[0,T]\times\mathbb{R}_0} f(x)\mathcal{J}(dt, dx) = \sum_{t \leq T} f(\Delta X(t)), \tag{2.3.1}$$

are simply not accessible. In this section, we discuss the approximation of the integral (2.3.1) based on time series of the form $\left\{X(t_k^n)\right\}_{k=0}^{n}$, where $t_k^n = \frac{kT}{n}$.

Let us motivate our approximation scheme. The natural way of interpolating the sample path of a Lévy process from the sampling observations $\left\{X(t_k^n)\right\}_{k=0}^{n}$ is to take a càdlàg piecewise constant approximation of the form

$$X^n(t) \equiv \sum_{k=1}^{n} X\left(t_{k-1}^n\right) \mathbf{1}\left(t \in [t_{k-1}^n, t_k^n)\right), \quad t \in [0, T). \tag{2.3.2}$$

Above, $\mathbf{1}$ denotes the indicator function of the corresponding set. It is quite simple to prove that we have the convergence of X^n to X at finitely many points with probability one (a quality shared by any right-continuous process X). Furthermore, the approximated process X^n, having independent increments, converges to $\mathbf{X}$ in $D[0, \infty)$, under the Skorohod metric (see VI of [29] and in concrete Example VI.18). Hence, a first guess is that

$$I_n(f) \equiv \sum_{t \leq T} f(\Delta X^n(t)) = \sum_{k=1}^{n} f\left(X\left(t_k^n\right) - X\left(t_{k-1}^n\right)\right), \tag{2.3.3}$$

converges to (2.3.1) as $n \to \infty$. We are able to prove the weak convergence of (2.3.3) to (2.3.1) using well-know facts on the transition distributions of X in small time (see for instance pp. 39 of [5], Corollary 8.9 of [39], or Corollary 3 of [36]).

Lemma 2.3.1 *Let $X = \{X(t)\}_{t \geq 0}$ be a Lévy process with Lévy measure ν. Then:*

1) *For each $a > 0$,*

$$\lim_{t \to 0} \frac{1}{t} \mathbb{P}\left(X(t) > a\right) = \nu([a, \infty)), \quad \text{and} \quad \lim_{t \to 0} \frac{1}{t} \mathbb{P}\left(X(t) \leq -a\right) = \nu((-\infty, -a]). \tag{2.3.4}$$

2) *For any continuous bounded function h vanishing on a neighborhood of the origin,*

$$\lim_{t \to 0} \frac{1}{t} \mathbb{E}\left[h\left(X(t)\right)\right] = \int_{\mathbb{R}_0} h(x) \nu(dx). \tag{2.3.5}$$

Remark 2.3.2 *In particular, the two parts in the previous Lemma imply (2.3.5) when $h(x) = \mathbf{1}_{(a,b]}(x) f(x)$, where $(a, b]$ is an interval of $\mathbb{R}_0$ and f is a continuous function.*

It is worth mentioning that [36] provides stronger conclusions on the distribution of $X(t)$ for small time t. The following theorem summarizes some of their results.

Theorem 2.3.3 *Let $X = \{X(t)\}_{t\geq 0}$ be a Lévy process with Lévy measure ν. Let F_t be the distribution function of $X(t)$ and G the* spectral function *of ν; i.e. $G(x) = \nu([x,\infty))$ for $x > 0$ and $G(x) = \nu((-\infty, x])$ for $x < 0$. The following properties hold:*

(i) *If F_t and G have densities*[1] *f_t and g, then for $x \neq 0$*

$$\lim_{t\to 0} \frac{1}{t} f_t(x) = \frac{\partial}{\partial t} f_t(x)\Big|_{t=0} = g(x), \tag{2.3.6}$$

where we additionally assume that $F_t(x)$ is continuous in a neighborhood of $(t = 0, x)$ and that moreover $(\partial/\partial t)F_t(x)$, $(\partial/\partial x)F_t(x)$, and $(\partial/\partial t)(\partial/\partial x)F_t(x)$ exist and are continuous in $(t = 0, x)$.

(ii) *For a fixed $N \geq 1$, there exist $\varepsilon'(N) > 0$ and $t_0 > 0$ such that, for all $\varepsilon \in (0, \varepsilon'(N))$ and $t \in (0, t_0)$, and for $x > \eta > 0$,*

$$1 - F_t(x) = \sum_{i=1}^{N-1} \frac{t^i}{i!} G_\varepsilon^{\star i}(x) + O_{\varepsilon,\eta}(t^N), \tag{2.3.7}$$

where $G_\varepsilon(x) = \mathbf{1}(|x| \geq \varepsilon)G(x)$. Similarly, for $x < -\eta < 0$,

$$F_t(x) = \sum_{i=1}^{N-1} \frac{t^i}{i} G_\varepsilon^{\star i}(x) + O_{\varepsilon,\eta}(t^N). \tag{2.3.8}$$

(iii) *If h is continuous and bounded and if $\lim_{|x|\to 0} h(x)|x|^{-2} = 0$, then*

$$\lim_{t\to 0} \frac{1}{t}\mathbb{E}\left[h\left(X(t)\right)\right] = \int_{\mathbb{R}_0} h(x)\nu(dx).$$

Moreover, if $\int_{\mathbb{R}_0}(|x| \wedge 1)\nu(dx) < \infty$, it is enough to postulate that $h(x)(|x| \wedge 1)^{-1}$ is continuous and bounded.

Limiting results like (2.3.5) are useful to establish the convergence in distribution of $\mathrm{I}_n(f)$ since

$$\mathbb{E}\left[e^{\mathrm{i}u I_n(f)}\right] = \left(\mathbb{E}\left[e^{\mathrm{i}u f\left(X\left(\frac{T}{n}\right)\right)}\right]\right)^n = \left(1 + \frac{a_n}{n}\right)^n,$$

where $a_n = n\mathbb{E}\left[h\left(X\left(\frac{T}{n}\right)\right)\right]$ with $h(x) = e^{\mathrm{i}uf(x)} - 1$. So, if f is such that

$$\lim_{t\to 0} \frac{1}{t}\mathbb{E}\left[e^{\mathrm{i}uf(X(t))} - 1\right] = \int_{\mathbb{R}_0} \left(e^{\mathrm{i}uf(x)} - 1\right)\nu(dx), \tag{2.3.9}$$

[1] *The function $g \geq 0$ is said to be the density of the spectral function G if $G'(x) = g(x)$ for $x < 0$ and $G'(x) = -g(x)$ for $x > 0$.*

then a_n converges to $a \equiv T \int_{\mathbb{R}_0} h(x)\nu(dx)$, and thus

$$\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n = \lim_{n\to\infty} e^{n\log\left(1+\frac{a_n}{n}\right)} = e^a.$$

We thus have the following result (see Section 2.5 for verification):

Proposition 2.3.4 *Let $X = \{X(t)\}_{t\geq 0}$ be a Lévy process with Lévy measure ν. Then,*

$$\lim_{n\to\infty}\mathbb{E}\left[e^{\mathrm{i}uI_n(f)}\right] = \exp\left\{T\int_{\mathbb{R}_0}\left(e^{\mathrm{i}uf(x)}-1\right)\nu(dx)\right\},$$

if f satisfies either one of the following:

1) $f(x) = \mathbf{1}_{(a,b]}(x)h(x)$ for an interval $(a,b]\subset\mathbb{R}_0$ and a continuous function h;

2) $f(x)$ is continuous on $\mathbb{R}_0$ and $\lim_{|x|\to 0} f(x)|x|^{-2} = 0$.

In particular, $I_n(f)$ converges in distribution to $I(f)$ under any of the two previous conditions.

Remark 2.3.5 *Notice that if $\mathcal{S}$ is a linear space of functions such that every $f\in\mathcal{S}$ fulfill (2.3.9), then the stochastic process $\{I_n(f)\}_{f\in\mathcal{S}}$, with "time-space" $\mathcal{S}$, converges in law to $\{I(f)\}_{f\in\mathcal{S}}$. The convergence is in the sense of finite dimensional distributions; i.e.*

$$\left(I_n(f_1),\ldots,I_n(f_d)\right)\xrightarrow{\mathfrak{D}}\left(I(f_1),\ldots,I(f_d)\right),$$

as $n\to\infty$, for all $f_1,\ldots,f_d\in\mathcal{S}$. This results is a direct consequence of the fact that the "random functionals" $I_n(\cdot)$ and $I(\cdot)$ are both linear. Proposition 2.3.4 describes two possibilities for the space $\mathcal{S}$.

Example 2.3.6 *Consider the case where f is the indicator function in an interval $(a,b]\subset\mathbb{R}_0$ so that $I_n(f)$ counts the number of increments $\{X(t_k^n)-X(t_{k-1}^n)\}_{k=1}^n$ that fall on that interval. As we will see, this type of statistics is relevant for the estimation of the Lévy density by histograms (piece-wise constant functions). The distribution of $I_n(f)$ is Binomial with parameters n and "success" probability $p_n\equiv\mathbb{P}\left[X(T/n)\in(a,b]\right]$. In that case, Proposition 2.3.4 merely asserts the elementary "Poisson approximation to Binomial", namely the* *distribution of* $I_n(f)$ *converges to a Poisson distribution with mean*

$$\lim_{n\to\infty} np_n = T\nu((a,b]).$$

Also, notice that

$$\mathrm{Var}(I_n(f)) = np_n(1-p_n) \to T\nu((a,b]) \tag{2.3.10}$$

as $n \to \infty$. *In general, if* $f^{(1)}, \ldots, f^{(d)}$ *are indicator functions on* mutually exclusive *Borel sets of* $\mathbb{R}_0$, *the vector* $\left(I_n\left(f^{(1)}\right), \cdots, I_n\left(f^{(d)}\right)\right)$ *has a multinomial distribution with parameters* n *and "membership" probabilities* $p_n^{(i)} = \mathbb{E}\left[f^{(i)}\left(X\left(\frac{T}{n}\right)\right)\right]$, *for* $i = 1, \ldots, d$. *In that case,*

$$\lim_{n\to\infty} n\mathrm{Cov}\left(I_n\left(f^{(i)}\right), I_n\left(f^{(j)}\right)\right) = -T^2 \int_{\mathbb{R}_0} f^{(i)}(x)\nu(dx) \int_{\mathbb{R}_0} f^{(j)}(x)\nu(dx), \quad i \neq j.$$

Moreover, the random variables $I_n\left(f^{(1)}\right), \cdots, I_n\left(f^{(d)}\right)$ *happen to be asymptotically uncorrelated as seen from*

$$\lim_{n\to\infty} n\rho\left(I_n\left(f^{(i)}\right), I_n\left(f^{(j)}\right)\right) = -T\left(\int_{\mathbb{R}_0} f^{(i)}(x)\nu(dx) \int_{\mathbb{R}_0} f^{(j)}(x)\nu(dx)\right)^{1/2},$$

valid for $i \neq j$.

Remark 2.3.7 *Clearly, if* f *and* f^2 *satisfy (2.3.5), then the mean and variance of* $I_n(f)$ *obey the asymptotics:*

$$\lim_{n\to\infty} \mathbb{E}\left[I_n(f)\right] = T\int_{\mathbb{R}_0} f(x)\nu(dx);$$
$$\lim_{n\to\infty} \mathrm{Var}\left[I_n(f)\right] = T\int_{\mathbb{R}_0} f^2(x)\nu(dx).$$

2.4 Estimation Method

Let us summarize the previous sections and outline the proposed algorithm of estimation:

Statistician's parameters: The procedure is fed with a Borel *window of estimation* $D \subset \mathbb{R}_0$, a collection $\{\mathcal{S}_m\}_{m\in\mathcal{M}}$ of finite dimensional *linear models* of $L^2((D,\eta))$, and a *level of penalization* $c > 1$.

Model and data: It is assumed that a Lévy process $\{X(t)\}_{t\in[0,T]}$ is monitored at equally spaced times $t_k^n = k\frac{T}{n}$, $k = 1, \ldots, n$, during the time period $[0, T]$. The data consists of the time series $\left\{X\left(t_k^n\right)\right\}_{k=1}^n$. The Lévy process admits a Lévy density p with regularization s under the measure η on D (see Definition 2.1.1).

Estimators: Inside the linear model $\mathcal{S}_m$, the estimator of s is the *approximated projection estimator*:

$$\hat{s}_m^n(x) \equiv \sum_{i=1}^{d_m} \hat{\beta}_{i,m}^n \varphi_{i,m}(x), \tag{2.4.1}$$

where $\{\varphi_{1,m}, \ldots, \varphi_{d_m,m}\}$ is an orthonormal basis for $\mathcal{S}_m$, and

$$\hat{\beta}_{i,m}^n \equiv \frac{1}{T}\sum_{k=1}^n \varphi_{i,m}\left(X\left(t_k^n\right) - X\left(t_{k-1}^n\right)\right), \tag{2.4.2}$$

is the estimator of the inner product $\beta_{i,m} \equiv \int_D \varphi_{i,m}(x)s(x)\eta(dx)$, for $i = 1, \ldots, d_m$. Across the collection of linear models $\{\mathcal{S}_m : m \in \mathcal{M}\}$, the estimator $\hat{s}_m^n$ which minimizes $-\|\hat{s}_m^n\|^2 + c \operatorname{pen}^n(m)$, is selected, where

$$\operatorname{pen}^n(m) = \frac{1}{T^2}\sum_{k=1}^n \left(\sum_{i=1}^{d_m} \varphi_{i,m}^2\left(X\left(t_k^n\right) - X\left(t_{k-1}^n\right)\right)\right).$$

Remark 2.4.1 *It is worthwhile to point out the great similarity of the scheme above to the methods of density estimation given by L. Birgé and P. Massart [7]. In this article, the authors estimate the probability density function f of a random sample $X_1, \cdots, X_n$ by projection estimators of the type:*

$$\hat{f}(x) = \sum_{i=1}^d \left\{\frac{1}{n}\sum_{k=1}^n \varphi_i\left(X_k\right)\right\}\varphi_i(x), \tag{2.4.3}$$

where $\{\varphi_i\}_{i=1}^d$ is an orthonormal basis of a linear model $\mathcal{S}$ of $L^2((\mathbb{R}, dx))$. To solve the problem of model selection, they propose penalized projection estimators with penalty function:

$$\operatorname{pen}(\mathcal{S}) = \frac{2}{n(n+1)}\sum_{k=1}^n \sum_{i=1}^d \varphi_i^2\left(X_k\right).$$

Then, it is intuitive that when estimating the Lévy density p of (2.1.1), the method outlined at the beginning of this section "works" as a byproduct of the small time qualities of Lévy *processes and of the standard methods of nonparametric estimation of probability densities. Concretely, consider the statistics*

$$\hat{\beta}_{i,m}^{n,j} \equiv \frac{n}{Tj} \sum_{k=1}^{j} \varphi_{i,m} \left(X\left(t_k^n\right) - X\left(t_{k-1}^n\right) \right).$$

Notice that n takes charge of the time length of the increments and j determines the number of increments. By the methods of Birgé and Massart, as j progresses the (penalized) projection estimator

$$\hat{s}_m^{n,j}(x) \equiv \sum_{i=1}^{d_m} \hat{\beta}_{i,m}^{n,j} \, \varphi_{i,m}(x),$$

will estimate $\frac{n}{T} f_{T/n}(x)$, where f_t stands for the probability density function of $X(t)$ (if it exists). By the small time properties of $\{X(t)\}_{t\geq 0}$ as summarized by Theorem 2.3.3, this will be enough to estimate the Lévy density p, if n is large enough. Notice that in general a.s.

$$\lim_{n\to\infty} \lim_{j\to\infty} \frac{n}{Tj} \sum_{k=1}^{j} \varphi \left(X\left(t_k^n\right) - X\left(t_{k-1}^n\right) \right) = \int_{\mathbb{R}_0} \varphi(x) \nu(dx),$$

whenever φ satisfies (2.3.5). Our method essentially conjectures that we can do both operations simultaneously and simply take $n = j$. Below, we prove that some asymptotically nice properties are still preserved with such a simplification.

Let $\mathcal{R}(X)$ be the linear space of measurable functions φ such that $\mathbb{E}\left[\varphi\left(X(t)\right)\right] < \infty$, for t in some $(0, \varepsilon)$, and

$$\lim_{t\to 0} \frac{1}{t} \mathbb{E}\left[\varphi\left(X(t)\right)\right] = \int_{\mathbb{R}_0} \varphi(x) \nu(dx), \tag{2.4.4}$$

where ν is the Lévy measure of the Lévy process $\{X(t)\}_{t\geq 0}$. Let $\mathcal{S}$ be a linear space of functions f such that

$$\lim_{t\to 0} \frac{1}{t} \mathbb{E}\left[e^{\mathrm{i}u f(X(t))} - 1 \right] = \int_{\mathbb{R}_0} \left(e^{\mathrm{i}u f(x)} - 1 \right) \nu(dx),$$

for every $u \in \mathbb{R}$. The following holds (see Section 2.5 for a proof).

Proposition 2.4.2 *Let $s_m^{\perp}$ be the orthogonal projection of s on $\mathcal{S}_m$. If $\varphi_{i,m}$ and $\varphi_{i,m}^2$ belong to $\mathcal{R}(X)$ for every $m \in \mathcal{M}$ and $i = 1, \ldots, d_m$, then the approximated projection estimator $\hat{s}_m^n$*

of s on $\mathcal{S}_m$ (based on n discrete observations) satisfies:

$$\lim_{n\to\infty} \mathbb{E}\left[\|\hat{s}_m^n - s_m^\perp\|^2\right] = \mathbb{E}\left[\|\hat{s}_m - s_m^\perp\|^2\right]. \tag{2.4.5}$$

Furthermore,

$$\lim_{n\to\infty} \mathbb{E}\left[\|\hat{s}_m^n - s\|^2\right] = \mathbb{E}\left[\|\hat{s}_m - s\|^2\right].$$

2.5 Some additional proofs

Verification of Remark 2.2.2: Suppose that D_m is finite, and thus each $f \in S$, with $\|f\| = 1$ is bounded. It follows using Lagrange multipliers that, for each $x \in D$,

$$D(x) \equiv \sup\left\{\Big|\sum_{i=1}^{d_m} c_i\varphi_i(x)\Big|^2 : \sum_{i=1}^{d_m} c_i^2 = 1\right\} = \sum_{i=1}^{d_m} \varphi_i^2(x).$$

Since $D_m \geq D(x)$ for every $x \in D$, we obtain $D_m \geq \|\sum_{i=1}^{d_m} \varphi_i^2\|_\infty$. On the other hand, for every $\varepsilon > 0$, there are $b_1, \ldots, b_n$ satisfying $\sum_{i=1}^{d_m} b_i^2 = 1$ and an $x \in D$ such that

$$D_m - \varepsilon < \Big|\sum_{i=1}^{d_m} b_i\varphi_i(x)\Big|^2 \leq D(x) = \sum_{i=1}^{d_m} \varphi_i^2(x) \leq \Big\|\sum_{i=1}^{d_m} \varphi_i^2\Big\|_\infty.$$

Letting $\varepsilon \to 0$, it follows that $D_m = \big\|\sum_{i=1}^{d_m} \varphi_i^2\big\|_\infty$. □

Proof of Lemma 2.2.5: Clearly, γ_D as defined by (2.1.9) can be written as

$$\mathbb{E}\gamma_D(f) = \|f\|^2 - 2f\cdot s_D - 2\nu_D(f) = \|f - s_D\|^2 - \|s_D\|^2 - 2\nu_D(f).$$

By the very definition of $\tilde{s}$ as the penalized projection estimator and by Remark 2.1.3,

$$\gamma_D(\tilde{s}) + \mathrm{pen}(\hat{m}) \leq \gamma_D(\hat{s}_m) + \mathrm{pen}(m) \leq \gamma(s_m^\perp) + \mathrm{pen}(m),$$

for any $m \in \mathcal{M}$. Using the previous two equations:

$$\begin{aligned}
\|\tilde{s} - s_D\|^2 &= \gamma_D(\tilde{s}) + \|s_D\|^2 + 2\nu_D(\tilde{s}) \\
&\leq \gamma(s_m^\perp) + \|s_D\|^2 + 2\nu_D(\tilde{s}) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}) \\
&= \|s_m^\perp - s_D\|^2 + 2\nu_D(\tilde{s} - s_m^\perp) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}).
\end{aligned}$$

Finally, notice that $\nu_D(\tilde{s} - s_m^{\perp}) = \nu_D(\tilde{s} - s_{\hat{m}}^{\perp}) + \nu_D(s_{\hat{m}}^{\perp} - s_m^{\perp})$ and $\nu_D(\hat{s}_m - s_m^{\perp}) = \chi_m^2$. □

Verification of inequality (2.2.10): Notice just that for any $a, b, \varepsilon > 0$:

$$a - \sqrt{2ab} - \frac{1}{3}b \geq \frac{a}{1+\varepsilon} - \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right)b. \tag{2.5.1}$$

Evaluating the integral in (2.2.9) for $-f$, we can write

$$\mathbb{P}\left[\int_{\mathbb{X}} f(x)N(dx) \geq \int_{\mathbb{X}} f(x)\mu(dx) - \|f\|_\mu \sqrt{2u} - \frac{1}{3}\|f\|_\infty u\right] \geq 1 - e^{-u}.$$

Using that $\|f\|_\mu^2 \leq \|f\|_\infty \int_{\mathbb{X}} f(x)\mu(dx)$ and (2.5.1),

$$\mathbb{P}\left[\int_{\mathbb{X}} f(x)N(dx) \geq \frac{1}{1+\varepsilon}\int_{\mathbb{X}} f(x)\mu(dx) - \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right)\|f\|_\infty u\right] \geq 1 - e^{-u},$$

which is precisely inequality (2.2.10). □

Proof of Lemma 2.2.8:

Let Z^+ be the positive part of Z. First,

$$\mathbb{E}[Z] \leq \mathbb{E}[Z^+] = \int_0^\infty \mathbb{P}[Z > x]dx.$$

Since h is continuous and strictly increasing, $\mathbb{P}[Z > x] \leq K\exp(-h^{-1}(x))$, where h^{-1} is the inverse of h. Then, changing variables to $u = h^{-1}(x)$,

$$\int_0^\infty \mathbb{P}[Z > x]dx \leq K\int_0^\infty e^{-h^{-1}(x)}dx = K\int_0^\infty e^u h'(u)du.$$

Finally, integration by parts yields $\int_0^\infty e^u h'(u)du = \int_0^\infty h(u)e^{-u}du$. □

Proof of Proposition 2.3.4:

It suffices to prove (2.3.9). Clearly, f and $h(x) \equiv e^{\mathrm{i}uf(x)} - 1$ both have the identical support and set of continuity. For the first case, the limit follows from Lemma 2.3.1 applied to the real and the imaginary parts of h (see also Remark 2.3.2). Next, if $\lim_{|x|\to 0} f(x)|x|^{-2} = 0$,

$$\lim_{|x|\to 0}\frac{h(x)}{x^2} = \lim_{|x|\to 0}\frac{e^{\mathrm{i}uf(x)} - 1}{x^2} = iu\lim_{|x|\to 0}\frac{f(x)}{x^2} = 0.$$

By part (iii) of Theorem 2.3.3, we get (2.3.9). The last statement in Proposition 2.3.4 follows from the characteristic function of a Poisson integral (Proposition 1.2.12). □

Proof of Proposition 2.4.2:

From the orthonormality property,

$$\begin{aligned}\mathbb{E}\left[\|\hat{s}_m^n - s_m^\perp\|^2\right] &= \sum_{i=1}^{d_m} \mathbb{E}\left[\left(\hat{\beta}_{i,m}^n - \beta_{i,m}\right)^2\right] \\ &= \sum_{i=1}^{d_m} \left\{\mathrm{Var}\left(\hat{\beta}_{i,m}^n\right) + \left(\mathbb{E}\left[\hat{\beta}_{i,m}^n\right] - \beta_{i,m}\right)^2\right\}.\end{aligned}$$

By remark 2.3.7,

$$\lim_{n\to\infty} \mathbb{E}\left[I_n(\varphi_{i,m})\right] = T\int_{\mathbb{R}_0} \varphi(x)s(x)\eta(dx) \quad \text{and} \quad \lim_{n\to\infty} \mathrm{Var}\left[I_n(\varphi_{i,m})\right] = T\int_{\mathbb{R}_0} \varphi_{i.m}^2(x)s(x)\eta(dx).$$

Then, (2.4.5) is true from (2.1.14) and (2.1.15). The second statement in the proof is straightforward since

$$\mathbb{E}\left[\|\hat{s}_m^n - s\|^2\right] = \mathbb{E}\left[\|\hat{s}_m^n - s_m^\perp\|^2\right] + \|s_m^\perp - s\|^2.$$

□

CHAPTER III

NUMERICAL TESTS OF THE METHOD

This part studies the performance of penalized projection estimators and model selection methods based on computer simulations. The first section shows some procedures, due to Rosiński [34], to simulate pure jump Lévy processes. These procedures are based on *shot-noice processes* and has the advantage of providing us with a set of jumps. Concretely, the method generated the jumps of the Lévy process. The second section analyzes the performance of projection estimators and approximated projection estimators to simulated data. We consider two relevant classes of Lévy processes for our numerical experiments: Gamma, and Variance Gamma models. A projection estimation method with least-squares errors is used to calibrate parametric or semiparametric models.

3.1 Simulation of Lévy processes

3.1.1 Brief overview

Accurate path simulation of a pure jump Lévy processes $\mathbf{X} = \{\mathbf{X}(t)\}_{t\in[0,T]}$, regardless of the relatively simple statistical structure of their increments, present some challenging problems when dealing with *infinite activity* (namely, processes with infinite Lévy measure). Just try to conceive that in this case the jump times are in fact dense on $[0,\infty)$ (see Theorem 21.3 of [39]).

One of the most popular simulation schemes is based on the generation of *discrete skeletons*. Namely, given a partition $t_0 = 0 < t_1 < \cdots < t_n \to \infty$ of $[0,\infty)$, the discrete skeleton of $\mathbf{X}$ (based on this partition) is defined by

$$\widetilde{\mathbf{X}}(t) = \sum_{k=1}^{\infty} \mathbf{X}(t_{k-1})\,\mathbf{1}(t \in [t_{k-1}, t_k)) = \sum_{k=1}^{\infty} \Delta_k \mathbf{1}(t \geq t_k),$$

where $\Delta_k = \mathbf{X}(t_k) - \mathbf{X}(t_{k-1})$. Moreover, if the partition is regular (that is, $t_k = k/n$ for some positive integer n), then $\widetilde{\mathbf{X}}(t) = \sum_{k=1}^{[nt]} \Delta_k$, and the increments $\{\Delta_k\}_{k\geq 1}$ are i.i.d. with common distribution $\mathcal{L}(\mathbf{X}(1/n))$. In that case, the discrete skeletons can readily be generated with any accuracy if the marginal distribution of $\mathbf{X}(t)$ can be simulated for any t. This type of approximation for the Lévy process was the motivation behind our approach to estimate the integrals with respect to the jump measure of $\mathbf{X}$ in Section 2.3. The main drawback to the previous scheme is the fact that most often the marginal distributions are not easily generated (at least for some popular financial models). The approximated process $\tilde{\mathbf{X}}$, having independent increments, converges to $\mathbf{X}$ in $D[0, \infty)$ under the Skorohod metric (see VI of [29] and in concrete Example VI.18).

The second easiest scheme would be to approximate the Lévy process by finite activity Lévy processes. That is, the Lévy-Itô decomposition of sample paths establishes that a.s. the process

$$\mathbf{X}_\varepsilon(t) \equiv t\left(\mathbf{b} - \int_{\|\mathbf{x}\|\geq\varepsilon} \mathbf{x}\, \nu(d\mathbf{x})\right) + \sum_{s\leq t} \Delta\mathbf{X}(s)\mathbf{1}(\|\Delta\mathbf{X}(s)\| \geq \varepsilon) \tag{3.1.1}$$

converges uniformly on any bounded interval, and a.s. the limiting process coincides with the paths of $\mathbf{X}$ (above, $\Delta\mathbf{X}(t) \equiv \mathbf{X}(t) - \mathbf{X}(t^-)$). The process $\sum_{s\leq t} \Delta\mathbf{X}(s)\mathbf{1}(\|\Delta\mathbf{X}(s)\| \geq \varepsilon)$ can easily be simulated by a *compound Poisson process* of the form $\sum_{k=1}^{\mathcal{J}^\varepsilon(t)} Y_k^\varepsilon$, where $\mathcal{J}^\varepsilon(t)$ is a homogeneous Poisson process with intensity $\nu(\|\mathbf{x}\| \geq \varepsilon)$ and where $\left\{Y_k^\varepsilon\right\}_{k=1}^\infty$ are i.i.d with common distribution

$$\nu_\varepsilon(d\mathbf{x}) \equiv \mathbf{1}(\|\mathbf{x}\| \geq \varepsilon)\frac{\nu(d\mathbf{x})}{\nu(\|\mathbf{x}\| \geq \varepsilon)}.$$

Clearly, such a scheme is unsatisfactory because all jumps smaller than ε are ignored. An alternative method of simulation is based on time series representations of the form

$$\mathbf{X}(t) = \mathbf{b}t + \sum_{i=1}^{\infty} \left[\mathbf{H}\left(\Gamma_i, \mathbf{V}_n\right) \mathbf{1}(U_i \leq t) - t\mathbf{c}_i\right],$$

which will be explained in the next section. We shall simulate the Lévy processes for our numerical experiments using this method. We decide on this method because it generates directly a sample of the process jumps, which are needed for the basic method of estimation described in 2.1. It is worth mentioning that the previous representation still exhibit some difficulties regarding the small jumps of the process. Indeed, in practice the series need to be truncated to finitely many terms, and since the *response function* **H** is decreasing in the first variable, this operation has a tendency to remove small jump sizes. Asmussen and Rosiński [2] introduce a further improvement to the jump-based method above that relies on a Brownian motion approximation for the small jumps of the process.

3.1.2 Simulations based on series representations

Throughout, let again $\mathbf{X} = \{\mathbf{X}(t)\}_{t\in[0,T]}$ be a Lévy process on $\mathbb{R}^d$ with characteristic function

$$\mathbb{E}\left[e^{i\mathbf{u}\cdot\mathbf{X}(t)}\right] = e^{t\psi(\mathbf{u})},$$

where ψ is characteristic exponent defined by

$$\psi(\mathbf{u}) = i\mathbf{u}\cdot\mathbf{b} + \int_{\mathbb{R}_0^d} \left\{e^{i\mathbf{u}\cdot\mathbf{x}} - 1 - i\mathbf{u}\cdot\mathbf{x}\,\mathbf{1}\left(\|\mathbf{x}\|\le 1\right)\right\}\nu(d\mathbf{x}). \tag{3.1.2}$$

We now introduce a methodology to simulate the process **X** based on series representations for the Lévy process **X**. The results below are presented in [34] and are given here for the sake of completeness. A series representation for **X** can be derived from a series representations for the random measure associated with the jumps of **X**. In general terms, if the random measure

$$\mathcal{J}(B) = \#\left\{t > 0 : (t, \mathbf{X}(t) - \mathbf{X}(t^-)) \in B\right\}, \tag{3.1.3}$$

has the representation

$$\mathcal{J}(\cdot) = \sum_{i=1}^{\infty} \delta_{(U_i,\mathbf{J}_i)}(\cdot), \tag{3.1.4}$$

for a sequence of i.i.d uniform random variables $\{U_i\}_{i\ge1}$ on $[0,T]$ and a sequence of random vectors $\{\mathbf{J}_i\}_{i\ge1}$, then a.s. **X** has the *shot-noise series* representation

$$\mathbf{X}(t) = \mathbf{b}t + \sum_{i=1}^{\infty}\left[\mathbf{J}_i\mathbf{1}(U_i \le t) - t\mathbf{c}_i\right], \quad 0\le t\le T, \tag{3.1.5}$$

for suitable centers $\mathbf{c}_i$ that compensate the jumps. The random variables U's govern the times of the jumps, while the **J**'s give the size of the jumps.

Let us first describe the techniques used to obtain representations of the form (3.1.4). Think of the jumps $\mathbf{J}$'s as random responses to fictitious "shots" occurring in the past according to a homogeneous Poisson process. The distribution of the jumps are dictated by a probability measure $\sigma(u;\cdot)$ on $\mathcal{B}(\mathbb{R}^d)$ which might depend on the elapsed time u between the jump and the shot. Moreover, if $\mathbf{J}_i$ is the jump originating from a shot Γ_i time units ago, assume that

$$\mathbb{P}\left(\mathbf{J}_i \in B \mid \left\{\Gamma_j\right\}_{j\geq 1}, \left\{\mathbf{J}_j\right\}_{j\neq i}\right) = \sigma\left(\Gamma_i, B\right), \quad B \in \mathcal{B}(\mathbb{R}^d). \tag{3.1.6}$$

It follows that, under the measurability of $\sigma(\cdot; B)$ for any $B \in \mathcal{B}(\mathbb{R}^d)$, the jumps $\mathbf{J}_1,\ \mathbf{J}_2,\ldots$ form a Poisson process in $\mathbb{R}^d$ with mean measure $T\int_0^\infty \sigma(u;B)du$, whenever the elapsed times among shots $0 < \Gamma_1 < \Gamma_2 < \ldots$ form themselves a homogeneous Poisson process on $(0,\infty)$ with intesity T (see Proposition 3.8 of [32]). Consequently, the marked point process $\sum_{i=1}^\infty \delta_{(U_i,\mathbf{J}_i)}(\cdot)$ will have mean measure of the form

$$\mu(dt, d\mathbf{x}) = \int_0^\infty \sigma(u; d\mathbf{x})du dt.$$

Combining the previous arguments with the Lévy-Itô decomposition for Lévy processes, we conclude that the measure $\mathcal{J}$ of (3.1.3) has the same law as $\sum_{i=1}^\infty \delta_{(U_i,\mathbf{J}_i)}(\cdot)$ whenever the Lévy measure ν has the representation

$$\nu(B) = \int_0^\infty \sigma(u;B)du. \tag{3.1.7}$$

Under the additional assumption that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{J}$ is defined is rich enough to be equipped with an independent uniform random variable, [34] shows that the sequences $\{\Gamma_i\}_{i\geq 1}$, $\{\mathbf{J}_i\}_{i\geq 1}$, and $\{U_i\}_{i\geq 1}$ can be defined in $(\Omega, \mathcal{F}, \mathbb{P})$ and the representation (3.1.4) holds a.s.

There are other considerations we need to think about for the representation (3.1.5) to hold. It has to do with the probabilistic structure of the jumps $\{\mathbf{J}_i\}$. Roughly speaking, to avoid divergence problems and to guarantee the existence of compensating centers $\mathbf{c}_i$, it is necessary that the magnitude of the jumps decreases as the elapsed time between the jump and the shot increases (an appealing physical assumption as well). This is better explained if we notice that (3.1.6) is equivalent to having

$$\mathbf{J}_i = \mathbf{H}(\Gamma_i, \mathbf{V}_i),$$

for a sequence of random elements $\{\mathbf{V}_i\}_{i\geq 1}$ independent of $\{U_i, \mathbf{V}_i\}_{i\geq 1}$ (see Lemma 2.22 of [20]). Then, we expect that $\|\mathbf{H}(r, \mathbf{v})\|$ should decrease in r for (3.1.5) to be true. Let us summarize the conditions and the main theorem for the simulation of Lévy processes.

Condition 3.1.1 *The jump measure of* $\mathbf{X}$ *can be written as*

$$\mathcal{J}(\cdot) = \sum_{i=1}^{\infty} \delta_{(U_i, \mathbf{H}(\Gamma_i, \mathbf{V}_i))}(\cdot), \quad a.s. \tag{3.1.8}$$

for a measurable function $\mathbf{H} : (0, \infty) \times S \to \mathbb{R}^d$*, where* S *is an arbitrary measurable space. Here,* $\{\Gamma_i\}_{i=1}^{\infty}$ *is a homogeneous Poisson process on* $\mathbb{R}_+$ *with intensity* T*,* $\{U_i\}_{i=1}^{\infty}$ *is an independent random sample with uniform distribution on* $(0, T)$*, and* $\{\mathbf{V}_i\}_{i=1}^{\infty}$ *is an independent random sample* $\{\mathbf{V}_i\}_{i=1}^{\infty}$ *with common distribution* F *on the space* S.

Condition 3.1.2 *For any Poisson process* $\{\Gamma_i^1\}_{i=1}^{\infty}$ *on* $\mathbb{R}_+$ *with unit rate,*

$$\mathbf{A}(\Gamma_n^1) - \mathbf{A}(n) \to 0 \ \ a.s., \tag{3.1.9}$$

where

$$\mathbf{A}(s) \equiv \int_0^s \int_S \mathbf{H}(r, \mathbf{v}) \mathbf{1} \left(\|\mathbf{H}(r, v)\| \leq 1 \right) F(d\mathbf{v}) dr. \tag{3.1.10}$$

The next lemma gives sufficient conditions for (3.1.9) (see pp. 409 [34]):

Lemma 3.1.3 *The limit in (3.1.9) holds true if either one of the following conditions is satisfied:*

i) $\mathbf{a} \equiv \lim_{s\to\infty} \mathbf{A}(s)$ *exists in* $\mathbb{R}^d$*;*

ii) *the mapping* $r \to \|\mathbf{H}(r, v)\|$ *is nonincreasing for each* $v \in S$.

The following result establishes the series representations for Lévy processes.

Proposition 3.1.4 *If the conditions 3.1.1 and 3.1.2 are satisfied then, a.s.*

$$\mathbf{X}(t) = \mathbf{b}t + \sum_{i=1}^{\infty} \left[\mathbf{H}(\Gamma_i, \mathbf{V}_i)\mathbf{1}\left(U_i \leq t\right) - t\mathbf{c}_i\right], \tag{3.1.11}$$

for all $t \in [0, T]$*, where* $\mathbf{c}_i \equiv \mathbf{A}(i) - \mathbf{A}(i-1)$.

Proof: Notice that the Lévy-Itô representation (1.2.3) takes the form

$$\mathbf{X}(t) = \mathbf{b}t + \int_{[0,t]\times\mathbb{R}_0^d} \mathbf{x}\mathbf{1}(\|\mathbf{x}\| \leq 1)(\mathcal{J} - E\mathcal{J})(du, d\mathbf{x}) + \int_{[0,t]\times\mathbb{R}_0^d} \mathbf{x}\mathbf{1}(\|\mathbf{x}\| > 1)\mathcal{J}(du, d\mathbf{x}).$$

Define

$$M(\cdot) \equiv \sum_{i=1}^{\infty} \delta_{(U_i, \Gamma_i, \mathbf{V}_i)}(\cdot).$$

From Proposition 3.8. of [32], M is a (marked) Poisson process on $R \equiv [0, T] \times \mathbb{R}_+ \times \mathbb{R}_0^d$ with mean measure $du\, dr\, F(d\mathbf{v})$. By a "change of variables",

$$\begin{aligned}\mathbf{X}(t) = \mathbf{b}t &+ \int_{R_t} \mathbf{H}(r, \mathbf{v})\mathbf{1}(\|\mathbf{H}(r, \mathbf{v})\| \leq 1)(M(du, dr, d\mathbf{v}) - dudrF(d\mathbf{v})) \\ &+ \int_{R_t} \mathbf{H}(r, \mathbf{v})\mathbf{1}(\|\mathbf{H}(r, \mathbf{v})\| > 1)(M(du, dr, d\mathbf{v}) - dudrF(d\mathbf{v})),\end{aligned}$$

where $R_t \equiv [0, t] \times \mathbb{R}_+ \times \mathbb{R}_0^d$. Define

$$\begin{aligned}\mathbf{X}_s(t) = \mathbf{b}t &+ \int_{R_t^s} \mathbf{H}(r, \mathbf{v})\mathbf{1}(\|\mathbf{H}(r, \mathbf{v})\| \leq 1)(M(du, dr, d\mathbf{v}) - dudrF(d\mathbf{v})) \\ &+ \int_{R_t^s} \mathbf{H}(r, \mathbf{v})\mathbf{1}(\|\mathbf{H}(r, \mathbf{v})\| > 1)(M(du, dr, d\mathbf{v}) - dudrF(d\mathbf{v})),\end{aligned}$$

where $R_t^s \equiv [0, t]\times[0, s]\times\mathbb{R}_0^d$. Using that the Poisson process M is an independently scatter measure (that is, $M(A_1), \ldots, M(A_n)$ are mutually independent for disjoint sets $A_1, \ldots, A_n$), we can verify in a standard way that $\mathbf{X}_s(t)$ has independent increments both with respect to $s \in [0, \infty)$ and $t \in [0, T]$. Also, notice that

$$\mathbf{X}_s(t) = \mathbf{b}t + \sum_{i:\Gamma_i \leq s} \mathbf{H}(\Gamma_i, \mathbf{V}_i)\mathbf{1}\left(U_i \leq t\right) - t\mathbf{A}(s), \tag{3.1.12}$$

implying that $\mathbf{X}_s(t)$ enjoys càdlàg paths in s for each t. We claim that almost surely,

$$\lim_{s\to\infty} \mathbf{X}_s(t) = \mathbf{X}(t),$$

for all $t \in [0, T]$. Fix $t \in [0, T]$ and take a sequence $s_n \uparrow \infty$. Since $\mathbf{X}_\cdot(t)$ has càdlàg paths, it suffices to check that $\lim_{n\to\infty} \mathbf{X}_{s_n}(t) = \mathbf{X}(t)$ a.s. Furthermore, since $\mathbf{X}_{s_n}(t) = \sum_{i=1}^{n}(\mathbf{X}_{s_i}(t) - \mathbf{X}_{s_{i-1}}(t))$ and since $\mathbf{X}_\cdot(t)$ has independent increment, it is enough to have convergence is in distribution. The later can be deduced from arguments based on characteristic function. □

Remark 3.1.5 *We finally point out that if condition (3.1.9) is true, and the representation (3.1.8) holds in distribution, then the representation (4.5.12) is valid in the sense of finite dimensional distributions. In view of our opening arguments in the present Section, (3.1.8) can be obtained in law if if and only if the Lévy measure has the decomposition*

$$\nu(B) = \int_0^\infty \sigma(u; B)du, \tag{3.1.13}$$

where $\sigma(u; B) = \mathbb{P}\left[\mathbf{H}(u, \mathbf{V}) \in B\right]$.

The following remark considers the case of Lévy processes with paths of *bounded variation*.

Remark 3.1.6 *The series (4.5.12) simplifies further when* $\int_{\|\mathbf{x}\|\leq 1} \|\mathbf{x}\|\nu(d\mathbf{x}) < \infty$, *namely, when* $\mathbf{X}$ *has paths of bounded variation a.s. (see Theorem 21.9 of [39]). Concretely, a.s.*

$$\mathbf{X}(t) = (\mathbf{b} - \mathbf{a})\, t + \sum_{i=1}^{\infty} \mathbf{J}_i \mathrm{I}\left(U_i \leq t\right), \tag{3.1.14}$$

where $\mathrm{a} = \sum_{i=1}^{\infty} \mathbf{c}_i = \lim_{i\to\infty} \mathbf{A}(i)$. *Such a constant is finite and equals*

$$\mathbf{a} = \int_0^\infty \int_S \mathbf{H}(r, \mathbf{v})\mathbf{1}\left(\|\mathbf{H}(r, v)\| \leq 1\right) F(d\mathbf{v})dr = \int_{\|\mathbf{x}\|\leq 1} \mathbf{x}\nu(d\mathbf{x}).$$

The vector $\mathbf{d} \equiv \mathbf{b} - \mathbf{a}$ *is called the* drift *of the Lévy process and, when it exists, is uniquely determined by the following form of the characteristic function of* $\mathbf{X}$*:*

$$\mathbb{E}\left[e^{i\,\mathbf{u}\cdot\mathbf{X}(t)}\right] = \exp\left\{t\left(i\mathbf{u}\cdot\mathbf{d} + \int_{\mathbb{R}_0^d} \left\{e^{i\mathbf{u}\cdot\mathbf{x}} - 1\right\}\nu(d\mathbf{x})\right)\right\}$$

The previous methodology can be applied to generate series representation for a wide range of Lévy processes. We illustrate this technique in the next Chapter, Section 4.2, to obtain series for the *tempered stable processes* to be introduced in Chapter 4. Besides its importance as technique of simulations, Rosiński [34] suggests to use such representation in obtaining path sample properties of the Lévy process.

3.2 Numerical tests of projection estimators

This section addresses the performance of some projection estimators of Lévy densities based on simulations of the Lévy processes. For this initial analysis, we essentially take piece-wise constant estimation functions on regular partitions. Two relevant classes of Lévy processes are considered for our numerical tests: Gamma, and variance Gamma processes. For purposes of comparison, a method of least-squares errors is then used to generate the parametric Lévy density that best fit our nonparametric results.

3.2.1 Specifications of the statistical methods

Let us describe in greater detail the projection estimators we consider. Below, we write $\mathcal{J}(A)$ instead of the notation $\mathcal{J}([0,T]\times A)$ of (2.1.3) when referring to the number of jumps of sizes in $A \in \mathcal{B}(\mathbb{R}_0)$ occurring prior to T, and we write $\chi_A(x)$ for the indicator function on A. Let $C : a = x_0 < x_1 < \cdots < x_m = b$ be a partition of the interval $D \equiv [a,b]$ ($0 < a < b$), and let $\mathcal{S}_C$ be the span of the indicator functions $\chi_{[x_0,x_1)},\ldots,\chi_{[x_{m-1},x_m)}$. In other words, the linear model $\mathcal{S}_C$ consists of "histogram functions" on the window D with cutoff points in C. We assume that the Lévy process has a Lévy density s bounded outside of any neighborhood of the origin. This assumption is very mild, and yet good enough for the integral $\int_D s^2(x)dx$ to be finite. In that case, the orthogonal projection of s onto $\mathcal{S}_C$ exists (under the standard inner product of $L^2(D,dx)$), and thus the projection estimation on $\mathcal{S}_C$ is meaningful. In the terminology of Section 2.1, the regularization measure is simply dx, the regularized Lévy density coincides with the Lévy density, and the orthonormal basis $\{\varphi_1,\ldots,\varphi_m\}$ for $\mathcal{S}_C$ is

$$\varphi_i(x) = \frac{1}{\sqrt{x_i - x_{i-1}}}\chi_{[x_{i-1},x_i)}(x), \quad i = 1,\ldots,m.$$

Following the *basic estimation method* outlined in Section 2.1, the projection estimator, previously defined in (2.1.10), on the linear model $\mathcal{S}_C$ is given by

$$\hat{s}_C(x) = \frac{1}{T}\sum_{i=1}^{m} \frac{\mathcal{J}([x_{i-1},x_i))}{x_i - x_{i-1}}\,\chi_{[x_{i-1},x_i)}(x). \tag{3.2.1}$$

Remember that the projection estimator on a linear model is characterized by having minimal contrast value on that model (see Remark 2.1.3). Under the previous notation, the contrast value of (3.2.1) takes the form:

$$\gamma(\hat{s}_C) = -\frac{1}{T^2}\sum_{i=1}^{m}\frac{(\mathcal{J}([x_{i-1}, x_i)))^2}{x_i - x_{i-1}}.$$

Similarly, the statistic $\hat{V}$ of (2.2.5), needed for the penalty function, is given by

$$\hat{V}_C = \frac{1}{T}\sum_{i=1}^{m}\frac{\mathcal{J}([x_{i-1}, x_i))}{x_i - x_{i-1}}. \tag{3.2.2}$$

In that case, following the heuristics of Section 2.1 and Theorem 2.2.3 part (b), an appealing procedure to select a projection estimator of the form (3.2.1) is to look for the minimization of the following penalized contrast value

$$\frac{1}{T^2}\sum_{i=1}^{m}\frac{1}{x_i - x_{i-1}}\left\{c\,\mathcal{J}([x_{i-1}, x_i)) - [\mathcal{J}([x_{i-1}, x_i))]^2\right\}. \tag{3.2.3}$$

Here, $c > 1$ is a constant that controls the level of penalization. In fact, Theorem 2.2.3 and Corollary 2.2.6 ensure us that, for large enough T, the previous procedure will yield competitive results against the best projection estimator. For that to happen it is necessary to restrict ourselves to models C satisfying that $D_C \leq T$, where D_C is defined as in (2.2.3). In this case, the constant D_C is $1/\min_{1\leq i\leq m}\{x_i - x_{i-1}\}$ as seen from Remark 2.2.2. Notice also that the mean square error (2.1.15) of $\hat{s}_C$ is

$$\mathrm{E}\left[\|s^{\perp}_{[a,b]} - \hat{s}_C\|^2\right] = \frac{1}{T}\sum_{i=1}^{m}\frac{1}{x_i - x_{i-1}}\int_{x_{i-1}}^{x_i} s(x)dx,$$

which goes to infinity when $a = x_0 \downarrow 0$.

The simplest case is to take regular partitions $\{x_i = a + i\Delta x\}_{i=0}^{m}$, where $\Delta x = (b-a)/m$ is the mesh of the partition. Then, the projection estimators of (3.2.1) becomes

$$\hat{s}_m(x) \equiv \frac{m}{T(b-a)}\sum_{i=1}^{m}\mathcal{J}([x_{i-1}, x_i))\,\chi_{[x_{i-1},x_i)}(x), \tag{3.2.4}$$

and penalized projection estimation will look to minimize

$$\frac{m}{T^2(b-a)}\left(c\mathcal{J}([a,b)) - \sum_{i=1}^{m}(\mathcal{J}([x_{i-1}, x_i)))^2\right), \tag{3.2.5}$$

over all m such that $D_m = m/(b-a)$ is smaller than T. Observe that (3.2.4) is simply a scaling of a histogram of the jumps of X. There are three new ingredients though: the base of the histograms is deliberately taken off the origin; the number of intervals in the partition is restricted to be at most $T(b-a)$; the use of penalized contrast values to choose an appropriate partition. Let us emphasize this point: the three previous conditions are not only perfectly objective, but also have very well-defined consequence: The Oracle inequality. This is in contrast to common practitioner methods of histogram construction, where the choice of the partition is made qualitatively and usually to try to match an assumption that bias our results and "don't let the data speak by itself". It is still open the choice of the estimation window D and penalization parameter c, but the arbitrariness of the method is reduced.

For comparison against other procedures and to assess the goodness of fit to specific parametric models, it is useful to find the parametric model of a given type that "best fits" our non-parametric estimators; for instance, suppose the we want to assess whether or not the nonparametric results supports the parametric Gamma model for the Lévy density. The method of least square errors provides an easy solution to this problem. For instance, if $s_\theta(x)$ is the parametric form of the Lévy density, a plausible estimator of θ can be defined by

$$\hat{\theta} = \operatorname{argmin}_\theta d(s_\theta, \hat{s}),$$

where $\hat{s}$ is the (penalized) projection estimator on a given family of linear models, and d is a function that accounts for the difference between s_θ and $\hat{s}$. For instance, for a fixed set of points $\{x_i\}_{i=1}^k \subset D$, $d(\cdot,\cdot)$ can simply be defined for functions f and g as

$$d(f,g) \equiv \sum_{i=1}^k [f(x_i) - g(x_i)]^2 .$$

If it is preferable to have a least-square method that is linear in the parameters to avoid *ill-poseness*[1] numerical problems, we can look for a functional T so that $T(s_\theta)$ is linear in

[1]In numerical methods, the term ill-conditioned or ill-posed refers to problems were small changes in the input data can cause "large" errors in the final solution

θ and define

$$d(f,g) \equiv \sum_{i=1}^{k} \left[T(f)(x_i) - T(g)(x_i)\right]^2 .$$

Example 3.2.1 *To illustrate the previous least-square scheme, let us go ahead and consider the case of Gamma Lévy densities*

$$s(x) = \frac{\alpha}{x} e^{-x/\beta},$$

for $x > 0$. *Given a (penalized) projection estimator* $\hat{s}$, *the least-square estimates of* α *and* β *based on the values at* $\{x_i\}_{i=1}^{k}$ *are the solution* $\hat{\alpha}$ *and* $\hat{\beta}$ *to the minimization problem:*

$$\min_{\alpha,\beta} \sum_{i=1}^{m} \left(\frac{\alpha}{x_i} \exp\left(-\frac{x_i}{\beta}\right) - \hat{s}(x_i) \right)^2 . \tag{3.2.6}$$

However, the estimation would be very susceptible to points with small x_i *and ill-conditioned problems might arise. We could try instead a regression method that is linear in the parameters using a logarithmic transformation:*

$$\min_{\alpha,\beta} \sum_{i=1}^{m} \left(-\frac{1}{\beta} x_i + \log(\alpha) - \log(x_i \hat{s}(x_i)) \right)^2 . \tag{3.2.7}$$

The simulation method we use (described in Section 3.1.2) generates a sample of n jumps of the Lévy process $\{X(t)\}_{t\in[0,T]}$ by truncating the series (3.1.5) or (4.5.12) to n terms. In our numerical experiments below, we will use this sample to approximate the integrals with respect to the jump measure $\mathcal{J}$ as follows

$$\iint_{[0,T]\times\mathbb{R}_0} f(x)\mathcal{J}(dt,dx) = \sum_{t\le T} f(\Delta X(t)) \approx \sum_{i=1}^{n} f(J_i), \tag{3.2.8}$$

where J_i is the i^{th} jump in the sample (see Proposition 3.1.4 for more detailed description). In particular, the projection estimators $\hat{s}_C$ of (3.2.1) and $\hat{s}_m$ of (3.2.4) will be approximated by

$$\hat{s}_C(x) \approx \frac{1}{T} \sum_{i=1}^{m} \frac{\#\{i : J_i \in [x_{i-1}, x_i)\}}{x_i - x_{i-1}} \chi_{[x_{i-1},x_i)}(x); \tag{3.2.9}$$

$$\hat{s}_m(x) \approx \frac{m}{T(b-a)} \sum_{i=1}^{m} \#\{i : J_i \in [x_{i-1}, x_i)\} \chi_{[x_{i-1},x_i)}(x). \tag{3.2.10}$$

If instead we apply discrete skeletons to simulate the path of the Lévy process or if we apply the method developed in Section 2.4, the integrals with respect to $\mathcal{J}$ shall be approximated using the increments $\left\{X(t_i^n) - X(t_{i-1}^n)\right\}_{i=1}^n$. From an algorithmic point of view, this can be achieved by taking $J_i = X(t_i^n) - X(t_{i-1}^n)$ in the previous expressions. Notice that in the previous case, the approximation of the estimator $\hat{s}_m$ is simply a scaling of a histogram of the i.i.d random variables $\left\{X(t_i^n) - X(t_{i-1}^n)\right\}_{i=1}^n$. In particular, we expect that such a histogram will be similar to the density function of $\mathcal{L}(X(T/n))$. This is not a contradiction since when n is large, the density of $\frac{n}{T}X\left(\frac{T}{n}\right)$ converges to the Lévy density under some regularity conditions (see Section 2.3 and Remark 2.4.1 for a discussion of this matter).

3.2.2 Estimation of Gamma Lévy densities.

3.2.2.1 The model

As a first example, we discuss the calibration of Gamma Lévy processes. These processes are fundamental building blocks in the construction of other Lévy processes like the variance Gamma model [13] and the generalized Gamma convolutions [8]. Moreover, by Berstein's theorem, any Lévy density of the form $u(x)/|x|$, where u is a completely monotone function, is the limit of superpositions of Gamma Lévy densities.

The Gamma Lévy process $X = \{X(t)\}_{t\geq 0}$ is determined by two positive parameters α and β so that the probability density function of $X(t)$ is

$$f_t(x) = \frac{x^{\alpha t-1}e^{-x/\beta}}{\Gamma(\alpha t)\beta^{\alpha t}}, \tag{3.2.11}$$

for $x > 0$. In this case, the characteristic function of X is

$$E\left[e^{iuX(t)}\right] = (1 - i\beta t)^{\alpha t} = \exp\left[t\left(\alpha\int_0^\infty \left(e^{iux} - 1\right)\nu(dx)\right)\right],$$

where the Lévy measure ν is

$$\nu(dx) = \frac{\alpha}{x}\exp\left(-\frac{x}{\beta}\right)dx, \quad \text{for} \quad x > 0; \tag{3.2.12}$$

see [16] pp. 87 or Example 8.10. of [39]. From the point of view of the marginal densities, β is a *scale parameter* and α is a *shape parameter*. In terms of the jump activity, α controls the overall activity of the jumps, while β takes charge of the heaviness of the Lévy density tail, and hence, of the frequency of big jumps. Notice that changes in the time units is statistically equivalent to changes of the parameter α, while changes in the units at which the values of X are measured are statistically reflected on changes of the parameter β. That is to say, the scaled process $\{cX(ht)\}_{t\geq 0}$ is also a Gamma Lévy process with shape parameter αh and scale parameter βc. This property is consistent with the previous remark on α taking charge of the jump activity and on β taking charge of the frequency of large jumps.

3.2.2.2 The simulation procedure

Let us specialize the simulation procedure of Section 3.1.2 to Gamma Lévy processes. That is to say, we need to find random element V on a measurable space $\mathcal{S}$ and a function $H : (0,\infty)\times\mathcal{S} \to \mathbb{R}_0$ such that

$$\nu(B) = \int_0^\infty \mathbb{P}\left[H(u,V) \in B\right] du.$$

It is not hard to check that V can be made exponentially distributed with mean 1 and $H(u,v) = \beta v e^{-u/\alpha}$. Indeed, for all $b > 0$,

$$\begin{aligned}\int_0^\infty \mathbb{P}\left[H(u,V) > b\right] du &= \int_0^\infty \int_0^\infty \mathbf{1}\left[\beta v e^{-\frac{u}{\alpha}} > b\right] e^{-v} dv du \\ &= \int_0^\infty \exp\left(-\frac{b}{\beta} e^{\frac{u}{\alpha}}\right) du. \\ &= \int_b^\infty \frac{\alpha}{x} e^{-\frac{x}{\beta}} dx,\end{aligned}$$

by changing variables to $x = be^{u/\alpha}$ in the last equality. $H(u,v)$ being non-increasing in v, Remark 3.1.4 implies that the conditions of Proposition 4.5.12 are satisfied and thus, the process

$$X(t) \equiv \beta \sum_{i=1}^{\infty} V_i \exp\left(-\frac{\Gamma_i}{\alpha}\right), \tag{3.2.13}$$

is a Gamma Lévy process on $[0,T]$ with shape parameter α and scale parameter β (see also Remark 3.1.6). Below, we shall truncate the series to n terms (corresponding to n jumps in the process X) to simulate a path of X and moreover, to simulate the jump process $\mathcal{J}$ by

$$\mathcal{J}_n(\cdot) \equiv \sum_{i=1}^{n} \delta_{J_i}(\cdot), \tag{3.2.14}$$

where $J_i \equiv \alpha V_i \exp\left(-\frac{\Gamma_i}{\alpha}\right)$.

3.2.2.3 *The numerical results*

We now present a few cases to illustrate the technique of projection estimation on histogram functions based on regular partitions (see Section 3.2.1 for the specifications of the estimation method). Figure 3.1 shows the Gamma density with $\alpha = 1$ and $\beta = 1$ and its (approximate) penalized projection histogram on regular partitions of the form (3.2.10). The estimation is based on 2000 jumps of the Gamma Lévy process on $[0, 365]$, and their approximated Poisson integrals (3.2.8). The least-square method (3.2.7), taking the x_i's as the mid points of the partition intervals, yields the estimators $\hat{\alpha} = 0.932$ and $\hat{\beta} = 1.055$. The maximum likelihood estimators based on the increments of the sample path of time length 1 are 1.015 for α and 0.949 for β (we do not observe real improvement if the time length of the increments is reduced).

In the next simulation, we consider a Gamma density with a lighter tail ($\beta = 0.5$) and more jump activity ($\alpha = 2$). The opposite setting was also studied: a heavier tail determined by a $\beta = 2$ and a lower jump activity given by an $\alpha = 0.5$ (see Figures 3.2 and 3.3). In the first scenario, the least-square method estimators are $\hat{\alpha} = 1.907$ and $\hat{\beta} = 0.472$, while the maximum likelihood estimators are 1.924 and 0.527, respectively. For this second Gamma density, the least-square method (3.2.6), taking the x_i's as the midpoints of the partition intervals, produce estimators $\hat{\alpha} = 0.5$ and $\hat{\beta} = 1.72$, while the maximum likelihood estimators are 0.55 and 1.99, respectively.

Approximate histogram estimation on regular partitions is less successful in case of high activity levels. This problem is particularly evident when we have in addition heavy tails in the Lévy density. For instance, if $\alpha = 3$ and $\beta = 3$, the method requires a large sample size to satisfactorily retrieve the behavior around the origin (see Figures 3.4 and

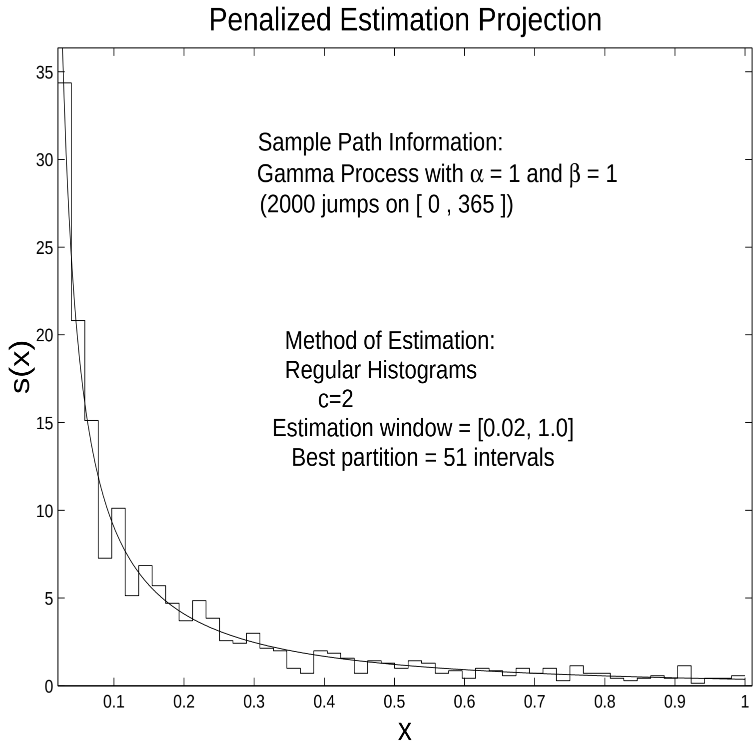

Figure 3.1: Penalized projection estimation of $\frac{e^{-x}}{x}$.

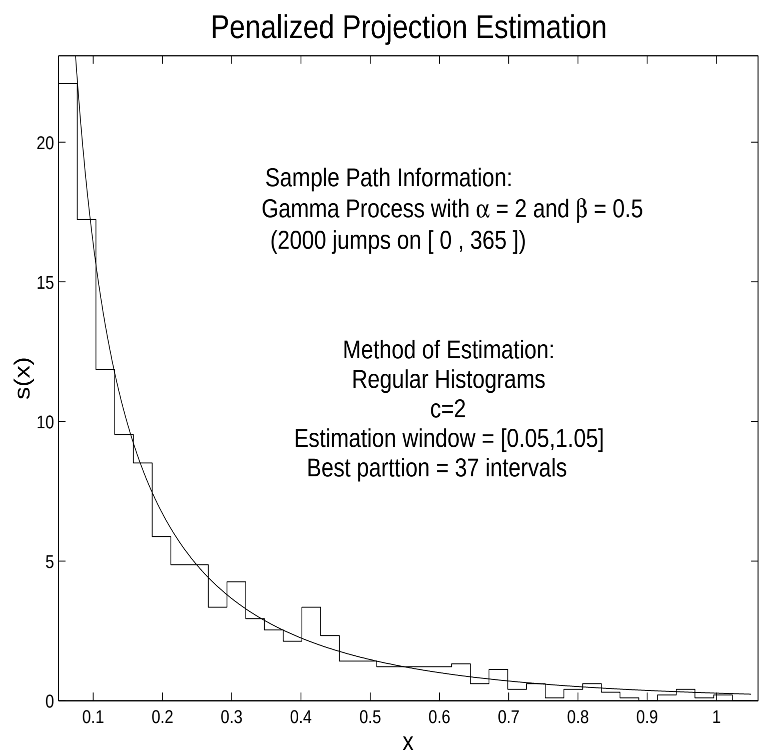

Figure 3.2: Penalized projection estimation of $\frac{2}{x}\exp(-2x)$.

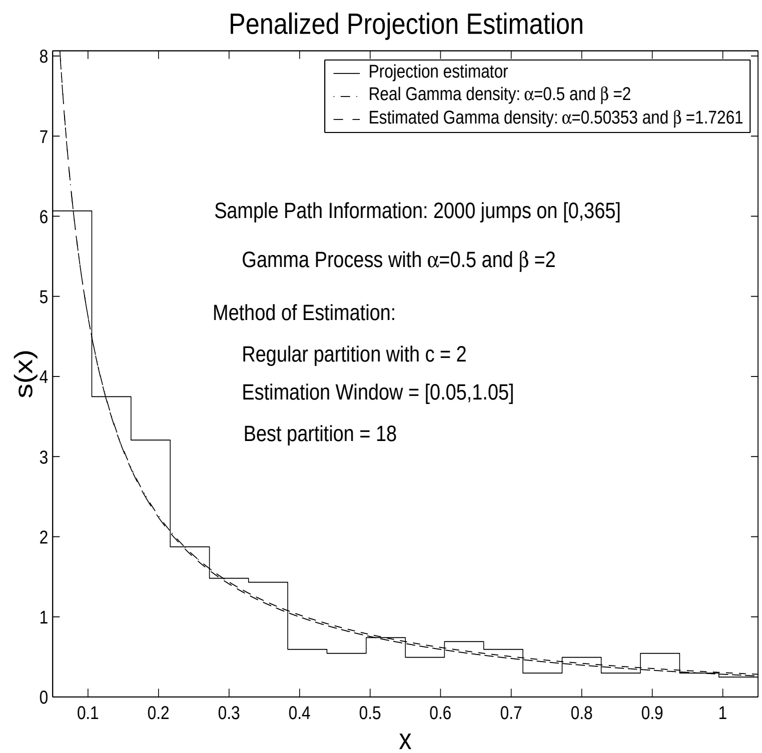

Figure 3.3: Penalized projection estimation of $\frac{1}{2x}\exp\left(-\frac{x}{2}\right)$.

3.5). For 2000 jumps, the least square estimates are $\hat{\alpha} = 1.87$ and $\hat{\beta} = 4.45$, while the estimates are $\hat{\alpha} = 2.8893$ and $\hat{\beta} = 2.9268$ for twice as many jumps. The maximum likelihood estimators based on the increments of time length .5 are 2.4134054 for α and 3.30971 for β when the approximate process is made out of 2000 jumps, while when the process is approximated using the 4000 jumps, these estimates are 2.8281 and 3.1007 for α and β, respectively. We also notice in our experiments that the estimates for the first simulation improve considerably if the window of estimation is taken "far away" from the origin (for example, $\hat{\alpha} = 3.20944$ and $\hat{\beta} = 2.68775$ on $[a, b] = [1.5, 5]$; see Figure 3.6).

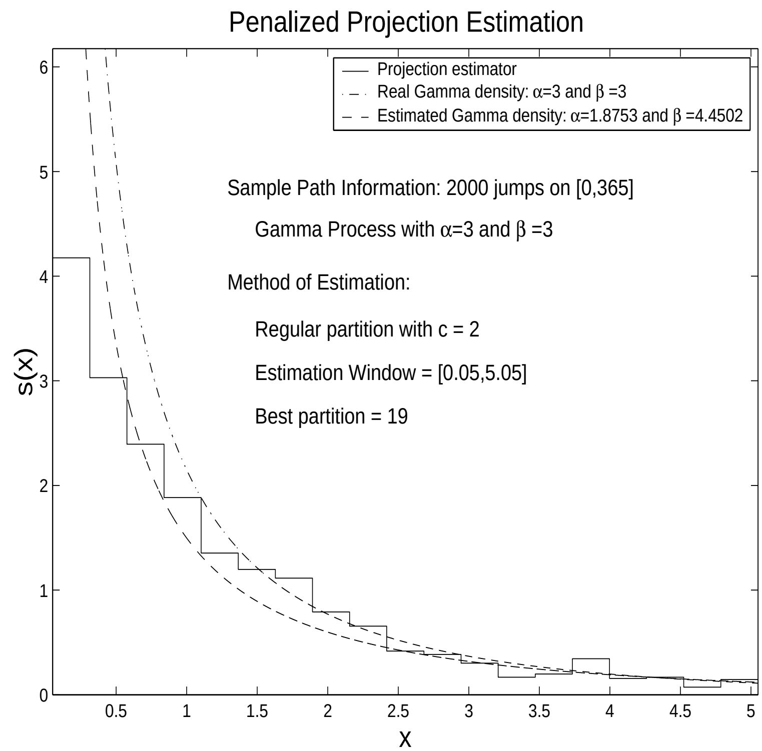

Figure 3.4: Penalized projection estimation of $\frac{3}{x}\exp\left(-\frac{x}{3}\right)$ on the interval $[.05, 5]$.

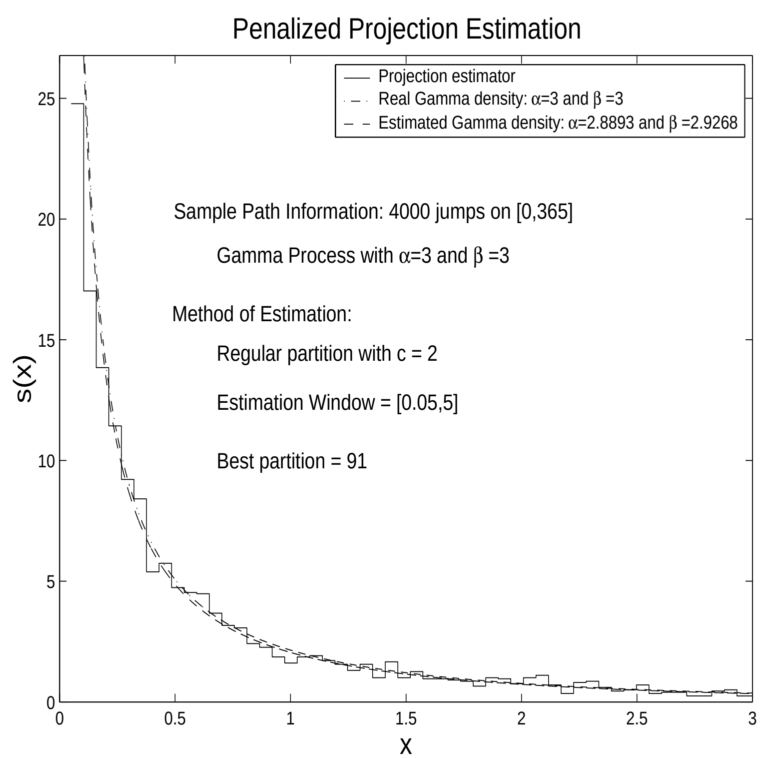

Figure 3.5: Penalized projection estimation of $\frac{3}{x}\exp\left(-\frac{x}{3}\right)$ on the interval $[.05, 5]$.

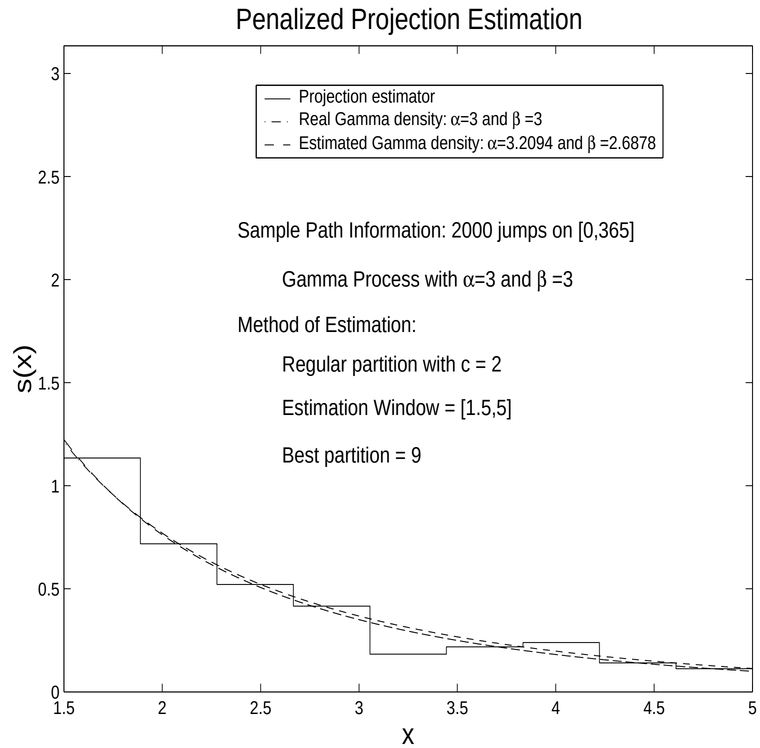

Figure 3.6: Penalized projection estimation of $\frac{3}{x}\exp\left(-\frac{x}{3}\right)$ on the interval $[1.5, 5]$.

3.2.2.4 Regularized projection estimation around the origin

We present another way to estimate the Gamma Lévy density even around the origin based on the regularization technique described in Section 2.1. The key observation is the following: the Gamma Lévy measure (3.2.12) can be written as

$$\nu(dx) = \alpha x \exp\left(-\frac{x}{\beta}\right)\eta(dx), \tag{3.2.15}$$

where $\eta(dx) = \frac{1}{x^2}dx$. Then, $s(x) \equiv \alpha x \exp\left(-\frac{x}{\beta}\right)$ is square integrable with respect to η, opening the possibility to use the projection estimation of s on a linear space $\mathcal{S}$ of $L^2\left((0,\infty),\eta\right)$. Once an estimator $\hat{s}$ for s has been obtained, $\hat{p}$ defined by $\hat{p}(x) = \hat{s}(x)/x^2$ can work as an estimator for the Lévy density $p(x) \equiv \alpha \exp\left(-x/\beta\right)/x$. In the terminology introduced in Section 2.1, η is a regularization measure for the Gamma Lévy density p, and s is the respective regularized Lévy density (see Definition 2.1.1).

Let us specify this method for the linear model

$$\mathcal{S}_C = \left\{ f : [0,\infty) \to \mathbb{R} \ : \ f(x) = c_1 x \chi_{[x_0,x_1)}(x) + \sum_{i=2}^{m} c_i \, \chi_{[x_i,x_{i+1})}(x), \text{ for } c_1,\ldots,c_m \in \mathbb{R} \right\},$$

where $C : 0 = x_0 < x_1 < \cdots < x_m = b$ is a partition of an chosen interval $D = [0,b]$. The projection estimator $\hat{s}_C$ onto $\mathcal{S}_C$, under the standard inner product of $L^2\left((0,\infty),\eta\right)$, takes on the value

$$\hat{s}_C(x) = x \frac{1}{T x_1} \sum_{t \le T} \Delta X(t) \mathrm{I}\left[\Delta X(t) < x_1\right],$$

if $x < x_1$, while if $x_{i-1} \le x < x_i$, for some $i \in \{2,\ldots,m\}$, then

$$\hat{s}_C(x) = \frac{x_{i-1} x_i}{T(x_i - x_{i-1})} \mathcal{J}([x_{i-1}, x_i)).$$

We shall use the penalty function of Theorem 2.2.3 part (b) to perform model selection. That is, among different partitions C that satisfy

$$D_C = \max\left\{ \frac{1}{x_1}, \frac{x_2 x_1}{x_2 - x_1}, \ldots, \frac{x_m x_{m-1}}{x_m - x_{m-1}} \right\} \le T,$$

we choose the projection estimator $\hat{s}_C$ that minimize

$$\gamma(\hat{s}_C) + \hat{V}_C = \frac{1}{T^2} \sum_{i=2}^{m} \frac{x_i x_{i-1}}{x_i - x_{i-1}} \left[c\, \mathcal{J}([x_{i-1}, x_i)) - \left(\mathcal{J}([x_{i-1}, x_i))\right)^2 \right]$$
$$+ \frac{c}{T^2 x_1} \sum_{\substack{t \le T: \\ \Delta X(t) < x_1}} (\Delta X(t))^2 - \frac{1}{x_1} \left(\sum_{\substack{t \le T: \\ \Delta X(t) < x_1}} \Delta X(t) \right)^2 .$$

The previous formulas are found directly from the definitions and results given in Section 2.1 (see for instance formulas (2.1.9), (2.1.10), (2.2.3), and (2.2.5)).

Let us also point out that the risk of estimation inside the linear model S_C is given by

$$\mathrm{E}\left[\|s^{\perp}_{[0,b]} - \hat{s}_C\|_\eta^2 \right] = \frac{1}{T} \left\{ \frac{1}{x_1} \int_0^{x_1} x^2 s(x) dx + \sum_{i=2}^{m} \frac{x_{i-1} x_i}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} s(x) dx \right\},$$

where $\|\cdot\|_\eta$ stands for the L^2-norm with respect to η.

Remark 3.2.2 *Observe that the previous procedure is appropriate to estimate the density function* $s(x) = \frac{\alpha}{x}\exp(-\frac{x}{\beta})$ *around the origin as far as*

$$\hat{\alpha} \equiv \frac{1}{T x_1} \sum_{t \le T} \Delta X(t) \mathrm{I}\left[\Delta X(t) < x_1\right],$$

is a good estimator of α. It is not hard to check that the bias of $\hat{\alpha}$ tend to zero as $x_1 \searrow 0$. However, the variance of $\hat{\alpha}$ converges to $\frac{\alpha}{2T}$, suggesting that the method works better when T is "large" and α is "small".

We apply the above method to the simulated Lévy process used in Figure 3.1; i.e. a Gamma process with $\alpha = 1$ and $\beta = 1$. Figure 3.7 shows the estimator $\hat{p}_2(x) = \hat{s}(x)/x^2$ and the actual Lévy density $p(x) = \exp(-x)/x$ for $x \in [0.02, 1]$ (we used regular partitions on $[0, 1]$). From Figure 3.1, the improvement is notorious, and moreover, we accomplish a good estimation around the origin of $\hat{p}_2(x) = 0.9/x$, for $x \in [0, 0.2)$.

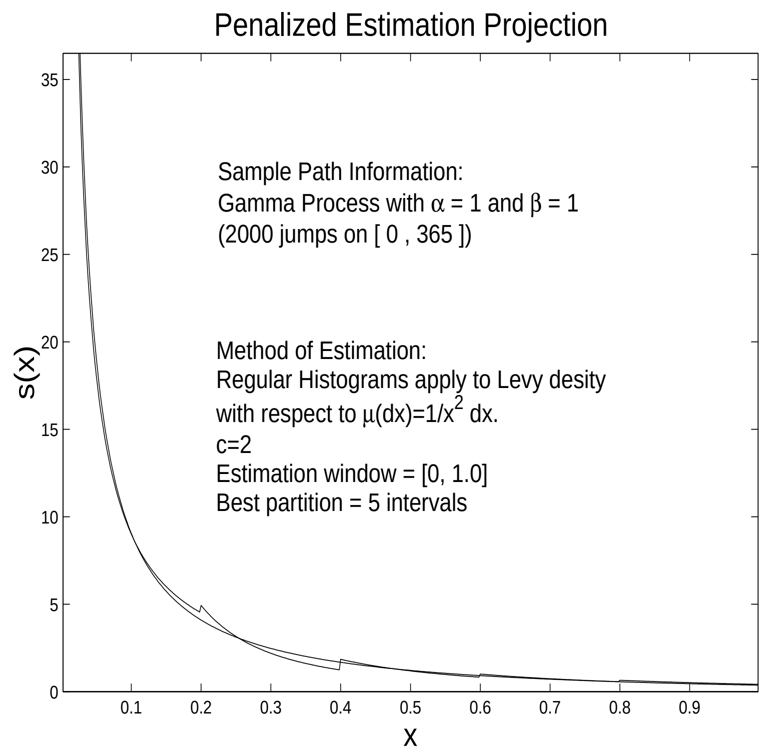

Figure 3.7: Penalized projection estimation of $\frac{e^{-x}}{x}$ using $\frac{\hat{s}(x)}{x^2}$.

This procedure was also applied to the simulations of Gamma Lévy process with ($\alpha = 3, \beta = 3$) and with ($\alpha = 1/2, \beta = 2$) (see the results of projection estimation for these two cases in Figures 3.3 and 3.5). The results are plotted in Figures 3.8 and 3.9 below. We observe an improvement under both sample data. For instance, for $\alpha = \beta = 3$, the nonparametric estimator $\hat{s}(x)/x^2$ combined with a method of least-squares errors estimate

α by 2.7296 and β by 3.2439. Similarly, when $\alpha = .5$ and $\beta = 2$, least-square errors estimates $\hat{\alpha} = .4825$ and $\hat{\beta} = 2.1131$.

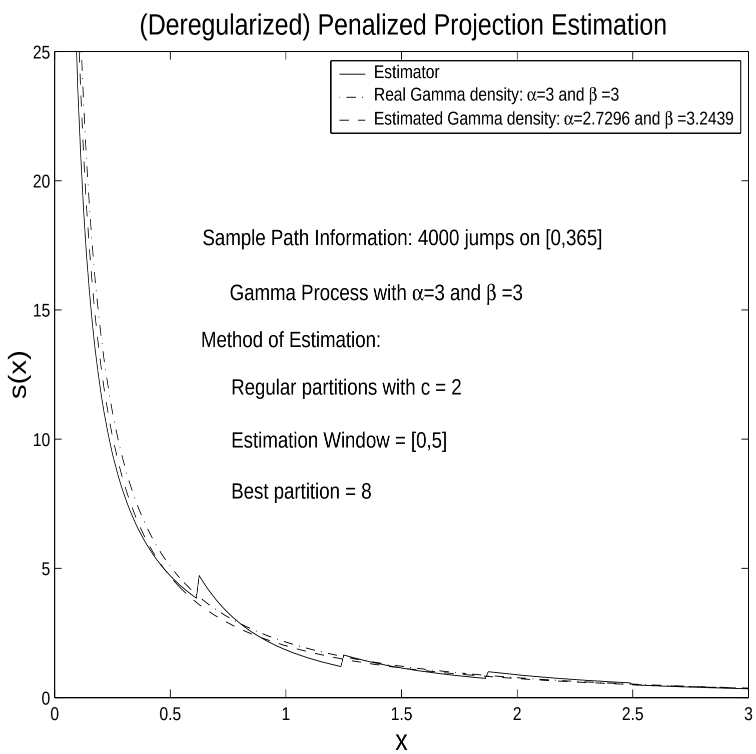

Figure 3.8: Regularized penalized projection estimation of $\frac{3}{x}\exp\left(-\frac{x}{3}\right)$.

3.2.2.5 Performance of projection estimation based on finitely many observation

In this part, we study the performance of the (approximate) projection estimators introduced in Section 2.3, and formally stated in Section 2.4. Namely, the method obtained by approximating the Poisson process of jumps $\mathcal{J}$ by

$$\mathcal{J}^n(\cdot) = \sum_{i=1}^{n} \delta_{J_i}(\cdot),$$

where J_i is the i^{th} increment of X from t_{i-1}^n to t_i^n and $t_i^n = iT/n$. The time span between increments is denoted by $\Delta t = T/n$. Concretely, the estimators we consider are histogram estimators as defined in Section 3.2.1 and applied in Section 3.2.2.3.

Table 3.1 compares (approximate) projection estimators with least-square errors (PPE-LSE) to maximum likelihood estimators (MLE) for the Gamma Lévy process with $\alpha = \beta =$

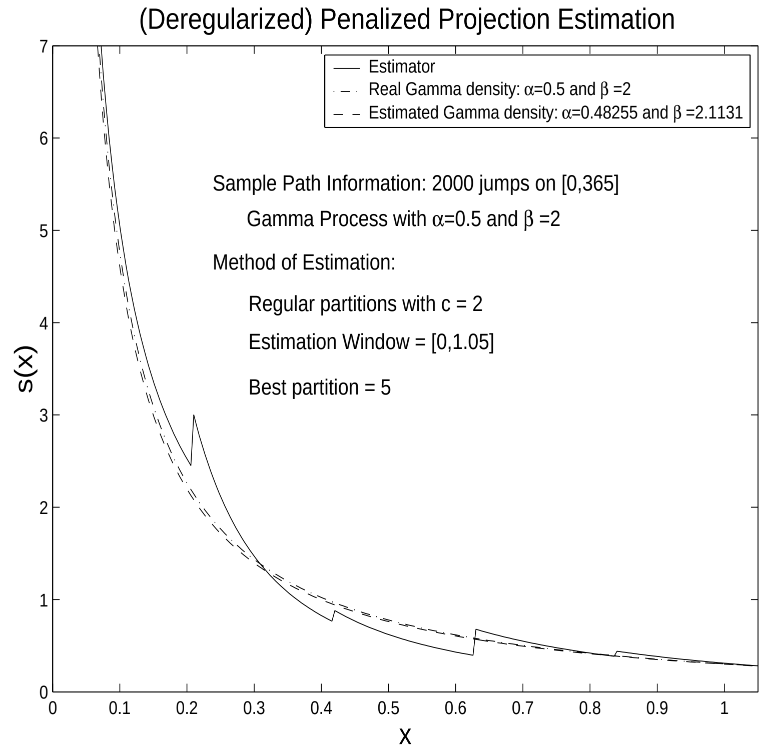

Figure 3.9: Regularized penalized projection estimation of $\frac{1}{2x}\exp\left(-\frac{x}{2}\right)$.

1 using different time spans Δt. We also consider two simulation procedures: jump-based and increment-based. The jump-based method uses series representation with $n = 36500$ jumps occurring during the time period $[0, 365]$ (notice that if we think of 365 as days, the number of jumps corresponds to an average of about 1 jump every minute). The increment-based method is a *discrete skeleton* (see Section 3.1.1) with mesh of 0.001.

	Jump-based Simulation				Increment-based Simulation			
Δt	PPE-LSE		MLE		PPE-LSE		MLE	
1	1.01	1.46	.997	.995	.73	1.78	1.09	.99
.5	1.03	1.09	.972	.978	.9	1.49	1.01	1.06
.1	.944	.995	1.179	.837	.923	1.03	.989	1.09
.01	.969	.924	6.92	.5	.955	1.019	.9974	1.083

Table 3.1: Estimation of a Lévy Gamma process with $\alpha = \beta = 1$. Two types of simulation are considered: series-representation based and increments-based. The estimations are based on equally spaced sampling observation at the time span Δt. Results for the approximate penalized projection estimators with least-squares errors, and for the maximum likelihood estimators are given.

Notice that maximum likelihood estimation does not perform well for small time spans when the approximate sample path is based on jumps. Similarly, penalized projection estimation does not provide good results for long time spans when the approximate sample path is based on increments.

3.2.3 Estimation of variance Gamma processes.

Variance Gamma processes have been introduced in [13] as a substitute to Brownian Motion in the Black-Scholes model. There are two useful representations for this type of processes. In short, a variance Gamma process $X = \{X(t)\}_{t\geq 0}$ is a Brownian motion with drift, time changed by a Gamma Lévy process. Concretely,

$$X(t) = \theta S(t) + \sigma W(S(t)), \tag{3.2.16}$$

where $\{W(t)\}_{t\geq 0}$ is a standard Brownian motion, $\theta \in \mathbb{R}$, $\sigma > 0$, and $S = \{S(t)\}_{t\geq 0}$ is a Gamma Lévy process with density at time t given by

$$f(x) = \frac{x^{t/\beta - 1} \exp\left(-\frac{x}{\beta}\right)}{\beta^{t/\beta}\Gamma\left(\frac{t}{\beta}\right)}. \tag{3.2.17}$$

Notice that $\mathrm{E}[S(t)] = t$ and $\mathrm{Var}[S(t)] = \beta t$; therefore, the random time S has a "mean rate" of one and a "variance rate" of β. There is no loss of generality in restricting the mean rate of the Gamma process S to one since, as a matter of fact, any process of the form

$$\theta_1 S_1(t) + \sigma_1 W(S_1(t)),$$

where $S_1(t)$ is an arbitrary Gamma Lévy process, $\theta_1 \in \mathbb{R}$, and $\sigma_1 > 0$, has the same law as (3.2.16) for suitably chosen θ, σ, and β. This a consequence of the *self-similarity*[2] property of the Brownian motion and the fact that β in (3.2.17) is a scale parameter.

The process X is itself a Lévy process since Gamma processes are *subordinators* (see

[2]namely, $\{W(ct)\}_{t\geq 0} \stackrel{\mathfrak{D}}{=} \{c^{1/2}W(t)\}_{t\geq 0}$, for any $c > 0$.

Theorem 30.1 of [39]). Moreover, it is not hard to check that "statistically" X is the difference of two Gamma Lévy processes (see 2.1 of [11]):

$$\{X(t)\}_{t\geq 0} \stackrel{\mathfrak{D}}{=} \{X_+(t) - X_-(t)\}_{t\geq 0}, \tag{3.2.18}$$

where $\{X_+(t)\}_{t\geq 0}$ and $\{X_-(t)\}_{t\geq 0}$ are Gamma Lévy processes with respective Lévy measures

$$\nu_\pm(dx) = \alpha \exp\left(-\frac{x}{\beta_\pm}\right) dx, \quad \text{for } x > 0.$$

Here, $\alpha = 1/\beta$ and

$$\beta_\pm = \sqrt{\frac{\theta^2\beta^2}{4} + \frac{\sigma^2\beta}{2}} \pm \frac{\theta\beta}{2}.$$

As a consequence of this decomposition, the Lévy density of X takes the form

$$s(x) = \begin{cases} \frac{\alpha}{|x|} \exp\left(-\frac{|x|}{\beta_-}\right) & \text{if } x < 0, \\ \frac{\alpha}{x} \exp\left(-\frac{x}{\beta_+}\right) & \text{if } x > 0, \end{cases}$$

where $\alpha > 0, \beta_- \geq 0$, and $\beta_+ \geq 0$ (of course, $\beta_-^2 + \beta_+^2 > 0$). As in the case of Gamma Lévy processes, α controls the overall jump activity, while β_+ and β_- take respectively charge of the intensity of large positive and negative jumps. In particular, the difference between $1/\beta_+$ and $1/\beta_-$ determines the frequency of drops relative to rises, while their sum measures the frequency of large moves relative to small ones.

The conclusion we want to draw in this part is that, from an algorithmic point of view, the estimation of this model based on projection estimation or approximate projection estimate is not different from the estimation of the Gamma process. We can simply estimate both tails of the variance Gamma process separately. However, from the point of view of maximum likelihood estimation, the problem is numerically challenging. The density function has closed form expression, but they involves Bessel functions (see [13]).

CHAPTER IV

TEMPERED STABLE DISTRIBUTIONS

The class of *tempered stable* distributions and its associated Lévy processes were recently studied by J. Rosiński [35] as a generalization of the *truncated stable* distributions introduced in the physics literature by [28], [27], and [22]. It is worthy to point out that Lévy processes with truncated stable distributions has been rediscovered and applied to mathematical finance by [11], [4], and others (see the references herein). In this part, we present a survey of his results, provide proofs when not given, and make some additional remarks.

4.1 Basic properties

There are different ways to construct tempered stable distributions from stable distributions. First, let us recall some features of the stable class (see e.g. Theorem 14.3 and 14.10 of [37]). Bellow and throughout, $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$.

Theorem 4.1.1 *Let η be a non-trivial infinitely divisible probability measure on $\mathbb{R}^d$ with generating triplet $(\Sigma, \gamma, \mathbf{b})$ and let $\mathbf{Y} = \{\mathbf{Y}(t)\}_{t\in[0,1]}$ be its associated Lévy process so that $\hat{\eta}(\mathbf{z}) = \mathbb{E}\left[e^{i\,\mathbf{z}\cdot\mathbf{Y}(1)}\right] = \exp(\psi(\mathbf{z}))$, where*

$$\psi(\mathbf{z}) = -\frac{1}{2}\,\mathbf{z}\cdot\Sigma\mathbf{z} + i\,\mathbf{z}\cdot\mathbf{b} + \int\limits_{\mathbb{R}^d_0}\left\{e^{i\,\mathbf{z}\cdot\mathbf{x}} - 1 - i\,\mathbf{z}\cdot\mathbf{x}\,I\left(\|\mathbf{x}\|\leq 1\right)\right\}\gamma(d\mathbf{x}). \tag{4.1.1}$$

For $0 < \alpha < 2$, the following statements are equivalent:

(i) η is $\alpha-$ stable;

(ii) $\Sigma = 0$ and there is a finite measure σ on the unit sphere S^{d-1} such that

$$\gamma(B) = \int_{S^{d-1}}\int_0^\infty I_B(r\mathbf{u})r^{-\alpha-1}dr\sigma(d\mathbf{u}), \tag{4.1.2}$$

for every $B \in \mathcal{B}(\mathbb{R}^d)$*;*

(iii) there exist a finite measure σ *on* S^{d-1}*, a constant* c_α *depending only on* α*, and a vector* $\mathbf{a} \in \mathbb{R}^d$ *such that*

$$\hat{\eta}(\mathbf{z}) = \exp\left\{-c_\alpha \int_{S^{d-1}} |\mathbf{z}\cdot\mathbf{u}|^\alpha \left(1 - i\ \tan\frac{\pi\alpha}{2}\mathrm{sgn}(\mathbf{z}\cdot\mathbf{u})\right)\sigma(d\mathbf{u}) + i\ \mathbf{a}\cdot\mathbf{z}\right\}, \tag{4.1.3}$$

for $\alpha \neq 1$*, and*

$$\hat{\eta}(\mathbf{z}) = \exp\left\{-c_1 \int_{S^{d-1}} \left(|\mathbf{z}\cdot\mathbf{u}| + i\ \frac{2}{\pi}(\mathbf{z}\cdot\mathbf{u})\log|\mathbf{z}\cdot\mathbf{u}|\right)\sigma(d\mathbf{u}) + i\ \mathbf{a}\cdot\mathbf{z}\right\}, \tag{4.1.4}$$

for $\alpha = 1$.

We proceed to introduce the tempered stable distribution.

Definition 4.1.2 *An infinitely divisible probability measure* μ *on* $\mathbb{R}^d$ *is called tempered stable if it does not have a Gaussian component (*$\Sigma = 0$*), and if its Lévy measure* ν *is of the form*

$$\nu(B) = \int_{\mathbb{R}^d_0}\int_0^\infty I_B(s\mathbf{x})s^{-\alpha-1}e^{-s}ds\rho(d\mathbf{x}), \tag{4.1.5}$$

where $\alpha \in (0,2)$ *and* ρ *is a* $\sigma - finite$ *Borel measure on* $\mathbb{R}^d_0 \equiv \mathbb{R}^d\backslash\{0\}$ *such that*

$$\int_{\mathbb{R}^d_0} ||\mathbf{x}||^\alpha \rho(d\mathbf{x}) < \infty. \tag{4.1.6}$$

The following remark will help us to better understand the relationship between the tempered and the standard stable distributions. In particular, it will be clear that ν above satisfies the integrability conditions of a Lévy measure.

Remark 4.1.3 *Given a Lévy measure* ν *as in (4.1.5) consider the measure*

$$\gamma(B) = \int_{\mathbb{R}^d_0}\int_0^\infty I_B(s\mathbf{x})s^{-\alpha-1}ds\rho(d\mathbf{x}), \quad B \in \mathcal{B}\left(\mathbb{R}^d_0\right). \tag{4.1.7}$$

It is not hard to see that γ *above is indeed of the form (4.1.2) with spherical part* σ *given by*

$$\sigma(C) = \int_{\mathbb{R}^d_0} I_C\left(\frac{\mathbf{x}}{||\mathbf{x}||}\right)||\mathbf{x}||^\alpha \rho(d\mathbf{x}), \tag{4.1.8}$$

for any $C \in \mathcal{B}\left(S^{d-1}\right)$. Therefore, we can associate a stable distribution with each tempered stable distribution, namely, an stable distribution with Lévy density η. In the next sections we will explore the relationship between these two distributions from the point of view of the corresponding Lévy processes. Notice also that since $\nu \le \gamma$ everywhere, the fact that γ satisfies $\int_{\mathbb{R}_0^d}(\|\mathbf{x}\|^2 \wedge 1)\gamma(d\mathbf{x}) < \infty$ implies the same condition for ν.

Example 4.1.4 *For $d = 1$, the tempered stable distribution μ has the Lévy density*

$$s(x) = \begin{cases} x^{-\alpha-1}q^{+}(x) & \text{if } x > 0, \\ |x|^{-\alpha-1}q^{-}(|x|) & \text{if } x < 0, \end{cases}$$

where

$$q^{+}(x) = \int_{(0,\infty)} e^{-x/s}s^{\alpha}\rho(ds), \text{ and } q^{-}(x) = \int_{(-\infty,0)} e^{-x/|s|}|s|^{\alpha}\rho(ds).$$

In particular, Bernstein's representation tell us that q^{+} and q^{-} are completely monotone functions such that $q^{\pm}(\infty) = 0$ and $q^{\pm}(0^{+}) < \infty$ (see XIII.4 in [15] for versions of the Bernstein's Theorem). In fact, any completely monotone function q on $(0,\infty)$ for which $q(\infty) = 0$ and $q(0^{+}) < \infty$ can be written as $q(x) = \int_{(0,\infty)} e^{-x/s}s^{\alpha}\rho(ds)$, for a suitable finite measure ρ on $(0,\infty)$. The tempered stable distribution with $\rho(ds) = w^{-}\delta_{\lambda^{-}}(ds) + w^{+}\delta_{\lambda^{+}}(ds)$, where $\lambda^{-} < 0 < \lambda^{+}$, is an important case that has been studied in financial applications by [11], [9], and [4]. Such distributions will be called truncated stable.

We establish now the analog of Theorem 4.1.1 (iii) for tempered stable distributions. Bellow, the branches of $\log(v) = \log|v| + i\arg(v)$ and $v^{\alpha} = |v|e^{i\,\alpha\arg(v)}$ are chosen such that $\arg(v) \in (-\pi,\pi]$, for any complex number v.

Theorem 4.1.5 *Let μ be a tempered stable distribution with Lévy measure (4.1.5). Then, its characteristic function $\hat{\mu}$ is given by*

$$\hat{\mu}(\mathbf{z}) = \exp\left\{k_{\alpha}\int_{\mathbb{R}_0^{d-1}} \psi_{\alpha}(\mathbf{z}\cdot\mathbf{x})\rho(d\mathbf{x}) + i\,\mathbf{a}\cdot\mathbf{z}\right\}, \tag{4.1.9}$$

where

$$\psi_\alpha(\omega) = \begin{cases} 1-(1-i\,\omega)^\alpha\,, & \text{if } 0<\alpha<1, \\ (1-i\,\omega)\log(1-i\,\omega)\,, & \text{if } \alpha=1, \\ (1-i\,\omega)^\alpha - 1 + i\,\alpha\omega, & \text{if } 1<\alpha<2, \end{cases}$$

and $k_\alpha = |\Gamma(-\alpha)|^1$*, for* $\alpha \neq 1$*, and* $k_1 = 1$.

The previous theorem allows us to study the "scaling properties" of tempered stable distributions. Let $\{\mathbf{X}(t)\}_{t\geq 0}$ be the Lévy process such that $\mu = \mathcal{L}(\mathbf{X}(1))$. Suppose that we are interested in the dynamics of the process when time is measured in small or large units. That is, we want to study $\mathbf{X}_h(t) \equiv \mathbf{X}(ht)$ for small or large $h > 0$. For instance, if originally t is measured in years, $\mathbf{X}_{1/365}(t)$ is simply the value of $\mathbf{X}$ in t days. The following result of [35] addresses the "microscalar" ($h \to 0$) and "macroscalar" ($h \to \infty$) behavior of tempered stable distributions (referred by Rosiński [35] as the short and long time behavior):

Theorem 4.1.6 *Let* μ *be a tempered stable distributions with characteristic function as in (4.1.9) with* $\mathbf{a} = \mathbf{0}$*. Let* $\{\mathbf{X}(t)\}_{t\geq 0}$ *be the Lévy process such that* $\mu = \mathcal{L}(\mathbf{X}(1))$*, and define*

$$\mathbf{X}_h(t) \equiv \mathbf{X}(ht),\ t \geq 0.$$

The limits below hold in the sense of convergence in law of the finite dimensional distributions:

(i) If $\alpha \neq 1$*, then*

$$\left\{\frac{1}{h^{1/\alpha}}\mathbf{X}_h(t)\right\}_{t\geq 0} \xrightarrow{\mathfrak{D}} \{\mathbf{Y}(t)\}_{t\geq 0}\,,\ \textit{as } h \to 0, \tag{4.1.10}$$

where $\{\mathbf{Y}(t)\}_{t\geq 0}$ *is a strictly stable process with characteristic function*

$$\mathbb{E}\left[e^{i\,\mathbf{z}\cdot\mathbf{Y}(t)}\right] = \exp\left\{-tc_\alpha \int_{\mathbb{R}_0^{d-1}} |\mathbf{z}\cdot\mathbf{x}|^\alpha \left(1 - i\,\tan\frac{\pi\alpha}{2}\operatorname{sgn}(\mathbf{z}\cdot\mathbf{x})\right)\rho(d\mathbf{x})\right\}, \tag{4.1.11}$$

and c_α *is a constant depending only on* α.

(ii) If $\alpha = 1$*, then*

$$\left\{\frac{1}{h}\mathbf{X}_h(t) - t\mathbf{a}_h\right\}_{t\geq 0} \xrightarrow{\mathfrak{D}} \{\mathbf{Y}(t)\}_{t\geq 0}\,,\ \textit{as } h \to 0, \tag{4.1.12}$$

[1] *The function Gamma is defined for negative real numbers* $x \neq -1, -2, \ldots$ *by applying recursively the property* $\Gamma(x) = \Gamma(x+1)/x$*. For instance,* $\Gamma(-\alpha) = \Gamma(2-\alpha)/(\alpha(\alpha-1))$ *if* $0 < \alpha < 2$, $\alpha \neq 1$.

where $\mathbf{a}_h \equiv \log(h) \int_{\mathbb{R}_0^d} \mathbf{x}\rho(d\mathbf{x})$ *and where* $\{\mathbf{Y}(t)\}_{t\geq 0}$ *is a strictly stable process such that*

$$\mathbb{E}\left[e^{i\,\mathbf{z}\cdot\mathbf{Y}(t)}\right] = \exp\left\{-tc_1 \int_{\mathbb{R}_0^{d-1}} \left(|\mathbf{z}\cdot\mathbf{x}| + i\,\frac{2}{\pi}(\mathbf{z}\cdot\mathbf{x})\log|\mathbf{z}\cdot\mathbf{x}|\right)\rho(d\mathbf{x})\right\}. \tag{4.1.13}$$

(iii) If $1 \leq \alpha < 2$ *and*

$$\int_{\mathbb{R}_0^d} \|\mathbf{x}\|^2 \rho(d\mathbf{x}) < \infty, \tag{4.1.14}$$

then

$$\left\{\frac{1}{h^{1/2}}\mathbf{X}_h(t)\right\}_{t\geq 0} \xrightarrow{\mathfrak{D}} \{\mathbf{B}(t)\}_{t\geq 0}\,,\ \textit{as } h \to \infty, \tag{4.1.15}$$

where $\{\mathbf{B}(t)\}_{t\geq 0}$ *is a Brownian motion with characteristic function*

$$\mathbb{E}\left[e^{i\,\mathbf{z}\cdot\mathbf{B}(t)}\right] = \exp\left\{-\frac{t}{2}\Gamma(2-\alpha)\int_{\mathbb{R}_0^{d-1}} |\mathbf{z}\cdot\mathbf{x}|^2\rho(d\mathbf{x})\right\}. \tag{4.1.16}$$

(iv) If $0 < \alpha < 1$ *and condition (4.1.14) is met, then*

$$\left\{\frac{1}{h^{1/2}}\mathbf{X}_h(t) - t\mathbf{a}_h\right\}_{t\geq 0} \xrightarrow{\mathfrak{D}} \{\mathbf{B}(t)\}_{t\geq 0}\,,\ \textit{as } h \to \infty, \tag{4.1.17}$$

where $\mathbf{B}$ *is as above and* $\mathbf{a}_h \equiv h^{1/2}\Gamma(1-\alpha)\int_{\mathbb{R}_0^d} \mathbf{x}\rho(d\mathbf{x})$*.*

4.2 Series Representations

It is clear from (4.1.5) and (4.1.7) that $\nu(B) \leq \gamma(B)$, for all $B \in \mathcal{B}\left(\mathbb{R}_0^d\right)$. We might wonder on which regions of $\mathbb{R}^d$ the Lévy measure ν is more alike to or more different from the Lévy measure γ. An answer to this is relevant in order to compare the jump dynamics of the Lévy processes associated with ν and γ. The subsequent result of [35] addresses this question and establishes roughly speaking that the tempered stable Lévy process is generated by truncating the jumps of the stable Lévy process. The truncation procedure truncates the size of the jumps, while keeping the direction of the jumps. We assume through this part, that the canonical tempered stable distribution μ of Definition 4.1.2 has characteristic function

$$\hat{\mu}(\mathbf{z}) = \exp\left(i\mathbf{z}\cdot\mathbf{b} + \int_{\mathbb{R}_0^d}\left\{e^{i\mathbf{z}\cdot\mathbf{x}} - 1 - i\mathbf{z}\cdot\mathbf{x}\,\mathrm{I}\left(\|\mathbf{x}\| \leq 1\right)\right\}\nu(d\mathbf{x})\right), \tag{4.2.1}$$

while η is a stable distribution with the same characteristic function as μ but substituting γ for ν (see Remark 4.1.3 above).

Theorem 4.2.1 *Let $\{\mathbf{X}(t)\}_{t\in[0,1]}$ and $\{\mathbf{Y}(t)\}_{t\in[0,1]}$ be Lévy processes such that $\mathcal{L}(\mathbf{X}(1)) \sim \mu$ and $\mathcal{L}(\mathbf{Y}(1)) \sim \eta$. On a common probability space, define independent sequences as follows:*

- $\{U_i\}_{i\geq 1}$, $\{T_i\}_{i\geq 1}$ *are i.i.d. uniform on* $[0,1]$ *random variables;*
- $\{E_i\}_{i\geq 1}$, $\{W_i\}_{i\geq 1}$ *are i.i.d. exponential random variables with mean* 1*;*
- $\{\mathbf{V}_i\}_{i\geq 1}$ *are i.i.d. random vectors in* $\mathbb{R}^d_0$ *with common distribution*

$$\rho_1(d\mathbf{x}) = \frac{1}{m(\rho)^\alpha}\|\mathbf{x}\|^\alpha \rho(d\mathbf{x}), \tag{4.2.2}$$

where $m(\rho)^\alpha$ is the normalizing constant $\int_{\mathbb{R}^d_0}\|\mathbf{x}\|^\alpha\rho(d\mathbf{x})$.

Then, if $\alpha \in (0,1)$ or if ρ is symmetric, the following series representations hold

$$\mathbf{Y}(t) \overset{\mathfrak{D}}{=} \sum_{i=1}^{\infty} m(\rho)\,(\alpha\Gamma_i)^{-1/\alpha}\,\frac{\mathbf{V}_i}{\|\mathbf{V}_i\|} I\,(T_i \leq t) + \mathbf{b}_1 t, \tag{4.2.3}$$

while

$$\mathbf{X}(t) \overset{\mathfrak{D}}{=} \sum_{i=1}^{\infty} \left\{\left(m(\rho)\,(\alpha\Gamma_i)^{-1/\alpha}\right) \wedge \left(\|\mathbf{V}_i\| E_i U_i^{1/\alpha}\right)\right\} \frac{\mathbf{V}_i}{\|\mathbf{V}_i\|} I\,(T_i \leq t) + \mathbf{b}_2 t, \tag{4.2.4}$$

for suitable vectors $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^d$. Here, the sequence $\{\Gamma_i\}_{i\geq 1}$ is the Poisson process $\Gamma_i = \sum_{k=1}^i W_k$. Moreover, the representations in (4.2.3) and (4.2.4) are in the sense of finite dimensional distributions and the series on the right hand sides converge uniformly in $t \in [0,1]$, with probability 1.

In the next section we show another method to construct series representations for tempered stable Lévy processes from Lévy stable processes. The procedure consists in *thinning* the "big" jumps of the stable process by a suitable *rejection* criterion (see Remark 4.3.3).

4.3 Spectral Decomposition

The following result gives the *spectral representation* of the tempered stable distributions. If we view an infinitely divisible random variable as a superposition of infinitely many jumps, a spectral representation of its distribution decomposes the content information of the jumps into a spherical part and a radial part. Generally speaking the former controls the direction of the jumps, while the latter determines the size of the jumps, although some compensation may be necessary.

Theorem 4.3.1 *The measure ν is the Lévy measure of a tempered stable distribution on $\mathbb{R}^d$ if and only if for $B \in \mathcal{B}(\mathbb{R}^d_0)$*

$$\nu(B) = \int_{S^{d-1}} \int_0^\infty I_B(r\mathbf{u}) r^{-\alpha-1} q(r, \mathbf{u}) dr \sigma(d\mathbf{u}), \tag{4.3.1}$$

where σ is a probability measure on S^{d-1} and $q : (0, \infty) \times S^{d-1} \to (0, \infty)$ is a Borel function such that $q(\cdot, \mathbf{u})$ is completely monotone with $q(\infty, \mathbf{u}) = 0$, for every $\mathbf{u} \in S^{d-1}$, and such that

$$\int_{S^{d-1}} q(0^+, \mathbf{u}) \sigma(d\mathbf{u}) < \infty. \tag{4.3.2}$$

Remark 4.3.2 *Let us briefly digress on a possible probabilistic interpretation of the spectral representation (4.3.1). Consider the more general setting where $\mathbf{X}$ is an infinitely divisible random vector without Gaussian component and with Lévy measure ν of the form:*

$$\nu(B) = \int_{S^{d-1}} \int_0^\infty I_B(r\mathbf{u}) \pi(\mathbf{u}, dr) \sigma(d\mathbf{u}), \tag{4.3.3}$$

where σ is a finite measure on S^{d-1} and $\pi : S^{d-1} \times \mathcal{B}((0, \infty)) \to [0, \infty]$ is a transition kernel. In the special case where σ is a probability measure and π is a probability kernel (a probability measure for each $\mathbf{u}$), the probabilistic nature of $\mathbf{X}$ is given by

$$\mathbf{X} \stackrel{\mathfrak{D}}{=} \sum_{i=1}^{N} R_i \mathbf{U}_i + \mathbf{b}, \tag{4.3.4}$$

where the $\mathbf{U}_i$'s, $i \geq 1$, are S^{d-1}-valued independent random elements with common distribution σ, and $\{R_i\}_{i\geq 1}$ is a sequence of conditionally independent random variables given

$\{\mathbf{U}_i\}_{i\geq 1}$ *such that*

$$\pi(\mathbf{U}_i, dr) = \Pr[R_i \in dr|\mathbf{U}_i] = \Pr[R_i \in dr|\{\mathbf{U}_j\}_{j\geq 1}, \{R_j\}_{j\neq i}]. \tag{4.3.5}$$

In (4.3.4), N is a Poisson random variable with mean 1 independent of $\{\mathbf{U}\}_{i\geq 1}$ and $\{R\}_{i\geq 1}$. This fact follows from Propositions 3.7 and 3.8 of [32]. If π is not a probability kernel, but $m \equiv \int_{S^{d-1}} \pi(\mathbf{u}, \mathbb{R}_0^+)\sigma(d\mathbf{u}) < \infty$, we can normalize π by $\pi(\mathbf{U}_i, \mathbb{R}_0^+)$ in (4.3.5) so that (4.3.4) holds with $N \sim Poisson(m)$ and

$$\mathbf{U}_i \sim \frac{1}{m}\pi(\mathbf{u}, \mathbb{R}_0^+)\sigma(d\mathbf{u}).$$

In the most general scenario, the jumps of small size need to be compensated to generate X as a limit in distribution of processes of the form

$$X \stackrel{\mathfrak{D}}{=} \lim_{\varepsilon\downarrow 0}\left(\sum_{i=1}^{N(\varepsilon,1)} R_{i,\varepsilon}\mathbf{U}_{i,\varepsilon} - \int_{S^{d-1}}\int_{\varepsilon}^{1} r\mathbf{u}\,\pi(\mathbf{u}, dr)\sigma(d\mathbf{u})\right) + \sum_{i=1}^{N(1,\infty)} R_i\mathbf{U}_i + \mathbf{b},$$

where

- *$N(a,b) \sim Poisson(m(a,b))$,with $m(a,b) \equiv \int_{S^{d-1}} \pi(\mathbf{u}, (a,b))\sigma(d\mathbf{u})$, for $0 < a < b < \infty$;*

- *For each $\varepsilon > 0$, the $\mathbf{U}_{i,\varepsilon}$, $i \geq 1$, are S^{d-1}-valued random elements with common distribution $\frac{1}{m(\varepsilon,1)}\pi(\mathbf{u}, (\varepsilon, 1))\sigma(d\mathbf{u})$.*

- *For each $\varepsilon > 0$, $\{R_{i,\varepsilon}\}_{i\geq 1}$ are conditionally independent given $\{\mathbf{U}_{i,\varepsilon}\}_{i\geq 1}$ such that*

$$\Pr[R_{i,\varepsilon} \in dr|\{\mathbf{U}_{j,\varepsilon}\}_{j\geq 1}, \{R_{j,\varepsilon}\}_{j\neq i}] = \frac{1}{\pi(\mathbf{U}_{i,\varepsilon}, (\varepsilon, 1))}\pi(\mathbf{U}_{i,\varepsilon}, dr \cap (\varepsilon, 1)).$$

- *Similar definitions for $\{R_i\}$ and $\{\mathbf{U}_i\}$ holds but substituting $(\varepsilon, 1)$ by $(1, \infty)$.*

- *$N(\varepsilon, 1)$, $N(1,\infty)$, $\{\mathbf{U}_{i,\varepsilon}\}_{i\geq 1}$ and $\{\mathbf{U}_i\}_{i\geq 1}$ are mutually independent. $\{R_{i,\varepsilon}\}_{i\geq 1}$ is independent of $N(\varepsilon, 1)$, $N(1,\infty)$, and $\{\mathbf{U}_i\}_i$. A similar statement holds for $\{R_i\}_{i\geq 1}$.*

Remark 4.3.3 *Notice that (4.3.1) can be written as*

$$\nu(B) = \int_{S^{d-1}} \int_0^\infty I_B(r\mathbf{u}) r^{-\alpha-1} \tilde{q}(r, \mathbf{u}) dr \tilde{\sigma}(d\mathbf{u}), \tag{4.3.6}$$

where $\tilde{\sigma}(d\mathbf{u}) = q(0^+, \mathbf{u})\sigma(d\mathbf{u})$ *and*

$$\tilde{q}(r, \mathbf{u}) = \begin{cases} \frac{q(r,\mathbf{u})}{q(0^+,\mathbf{u})} & q(0^+, \mathbf{u}) > 0, \\ 1 & \text{elsewhere.} \end{cases} \tag{4.3.7}$$

We can use (4.3.6) to obtain series representations of a Lévy tempered stable process $\mathbf{X} = \{\mathbf{X}(t)\}_{t\in[0,1]}$*, with Lévy measure* ν*. Let us consider for simplicity the case of symmetric* σ *or* $0 < \alpha < 1$ *(otherwise, compensating constants will be needed in the terms of the series). The method, a straightforward application of the* ***rejection method*** *of [34], consists in "thinning" the jumps of the stable Lévy process* $\mathbf{Y} = \{\mathbf{Y}(t)\}_{t\in[0,1]}$ *with Lévy measure*

$$\gamma(B) = \int_{S^{d-1}} \int_0^\infty I_B(r\mathbf{u}) r^{-\alpha-1} dr \tilde{\sigma}(d\mathbf{u}). \tag{4.3.8}$$

It was shown in Theorem 4.2.1 that for a random sample $\{T_i\}_{i\geq 1}$ *of uniform* $[0, 1]$ *random variables, a homogeneous Poisson process* $\{\Gamma_i\}_{i\geq 1}$ *on* $\mathbb{R}_+$ *with unit rate, and a constant vector* $\mathbf{b}$*, the following representation holds:*

$$\mathbf{Y}(t) \stackrel{\mathfrak{D}}{=} \sum_{i=1}^{\infty} R_i \mathbf{U}_i I(T_i \leq t) + \mathbf{b}t, \tag{4.3.9}$$

where $\mathfrak{D}$ *is in the sense of finite dimensional distributions. Here,*

$$R_i \equiv \left(\frac{\alpha\Gamma_i}{m(\tilde{\sigma})}\right)^{-1/\alpha},$$

$m(\tilde{\sigma}) \equiv \tilde{\sigma}(S^{d-1})$*, and* $\{\mathbf{U}_i\}_{i\geq 1}$ *is an independent sequence of i.i.d.* S^{d-1}*-valued vectors with common distribution* $\tilde{\sigma}(d\mathbf{u})/m(\tilde{\sigma})$*. Since*

$$\frac{d\nu}{d\gamma}(\mathbf{x}) = \tilde{q}\left(||\mathbf{x}||, \frac{\mathbf{x}}{||\mathbf{x}||}\right),$$

and $\tilde{q}(r, \mathbf{u}) \leq 1$*, the rejection method implies that the random measure* $N_{\mathbf{X}}$ *associated with the jumps on* $\mathbf{X}$ *has the same law as the Poisson process*

$$\sum_{i=1}^{\infty} \delta_{(T_i, R_i \mathbf{U}_i I[\tilde{q}(R_i, \mathbf{U}_i) \geq W_i])},$$

where $\{W_i\}_{i\geq 1}$ *is a sequence of i.i.d. random variables uniformly distributed on* $[0,1]$ *and independent of all the other sequences. Moreover, the tempered stable Lévy process* $\mathbf{X}$ *is such that*

$$\mathbf{X}(t) \stackrel{\mathfrak{D}}{=} \sum_{i=1}^{\infty} R_i \mathbf{U}_i I\left(\tilde{q}(R_i, \mathbf{U}_i) \geq W_i\right) I\left(T_i \leq t\right) + \mathbf{b}_1 t, \tag{4.3.10}$$

for a suitable vector $\mathbf{b}_1$ *(see the proof of Theorem 4.2.1 for more details). Notice that the representation (4.3.10) is different from (4.2.4) even when both series are generated from the same stable Lévy process. In some sense, the method above filters the jumps of* $\mathbf{Y}$ *when they are too big, while in (4.2.4) the jumps are truncated .*

4.4 Absolutely continuity with respect to stable processes

From the point of view of the corresponding Lévy Processes, there is yet another relationship between stable and tempered stable distributions. It is shown below that the distribution of a tempered stable Lévy process is (locally) absolutely continuous with respect to the distribution of its associated Lévy stable process. This implies the existence of a new probability measure, equivalent to the primary measure, such that under this measure the tempered stable process has the same statistical behavior as the associated stable process. Necessary and sufficient conditions for (locally) absolutely continuity between Lévy processes are well-known in the literature (see Section 33 of [39]), and we only need to apply these results in the context of tempered stable and stable distributions. We assume below that μ is a tempered stable distribution having generating triple $(0,\nu,\mathbf{b})$ with Lévy measure ν of the form (4.3.1), while η is a stable distribution with generating triple $(0,\gamma,\mathbf{c})$, where γ is given in (4.3.8) (see Remark 4.3.3). Recall that $D[0,T]$ stands for the space of functions $f : [0,T] \to \mathbb{R}^d$, that are right-continuous on $[0,T)$ and have left limits on $(0,T]$ (cádlág).

Theorem 4.4.1 *Let* $\{\mathbf{X}(t)\}_{t\geq 0}$ *be a Lévy process defined on a probability space* $(\Omega,\mathcal{F},\mathbb{P})$ *such that* $\mathbf{X}(1) \sim \mu$ *under* $\mathbb{P}$*. Let* $\{\mathbf{Y}(t)\}_{t\geq 0}$ *be another Lévy processes defined on a probability* *space* $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \mathbb{Q})$ *such that* $\mathbf{Y}(1) \sim \eta$ *under* $\mathbb{Q}$*. Let*

$$\int_{S^{d-1}} (\tilde{q}'(0^+, \mathbf{u}))^2 \, \tilde{\sigma}(d\mathbf{u}) < \infty, \tag{4.4.1}$$

and let

$$\mathbf{c} = \mathbf{b} + \int_{S^{d-1}} \mathbf{u} \int_0^1 (\tilde{q}\,(r, \mathbf{u}) - 1)\, r^{-\alpha} dr \tilde{\sigma}(d\mathbf{u}), \tag{4.4.2}$$

with $\tilde{q}$ *given by (4.3.7). Then, for each* $T > 0$*, the distribution of* $\{\mathbf{X}(t)\}_{0 \le t \le T}$ *on* $D[0, T]$ *is absolutely continuous with respect to the distribution of* $\{\mathbf{Y}(t)\}_{0 \le t \le T}$*. Moreover, let* U_T *be given by*

$$U_T \equiv \lim_{\varepsilon \to 0} \left(\int_0^T \int_{\{\|\mathbf{x}\| > \varepsilon\}} \ln g(\mathbf{x}) \mathcal{J}_{\mathbf{Y}}(dt, d\mathbf{x}) - T \int_{\{\|\mathbf{x}\| > \varepsilon\}} (g(\mathbf{x}) - 1)\, \gamma(d\mathbf{x}) \right), \tag{4.4.3}$$

where

$$g(\mathbf{x}) = \tilde{q} \left(\|\mathbf{x}\|, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right),$$

and where $\mathcal{J}_{\mathbf{Y}}$ *is the random measure associated with the jumps of* $\mathbf{Y}$*. Then, for any* $A \in \mathcal{F}_T$*,*

$$\mathbb{P}\left[\{\omega \in \Omega : \mathbf{X}(\cdot; \omega) \in A\}\right] = \mathbb{E}^{\mathbb{Q}} \left[e^{U_T} I \left[\{\omega \in \widetilde{\Omega} : \mathbf{Y}(\cdot; \omega) \in A\} \right] \right], \tag{4.4.4}$$

where $\mathbf{X}(\cdot; \omega)$ *(respectively,* $\mathbf{Y}(\cdot; \omega)$*) is the function in* $D[0, T]$ *defined by the mapping* $t \to \mathbf{X}(t; \omega)$ *(similarly definition for* $\mathbf{Y}(\cdot; \omega)$*). Above,* $\mathbb{E}^{\mathbb{Q}}$ *refers to the expectation with respect to the probability measure* $\mathbb{Q}$ *and* $\mathcal{F}_T$ *is the* σ*-field on* $D[0, T]$ *generated by the family of marginal projections* $\{\pi_t\}_{t \in [0,T]}$*, where* $\pi_t(\xi) \equiv \xi(t)$*, for* $\xi \in D[0, T]$*.*

Remark 4.4.2 *We are not assuming that* $\mathbf{X}$ *and* $\mathbf{Y}$ *are defined on the same probability space. Note as well that the expectation on the right hand side defines an equivalent probability measure such that the process* $\{\mathbf{Y}(t)\}_{0 \le t \le T}$ *under this probability measure has the same distribution as the process* $\{\mathbf{X}(t)\}_{0 \le t \le T}$*. Indeed, defining*

$$\widetilde{\mathbb{Q}}\,[C] = \mathbb{E}^{\mathbb{Q}} \left[e^{U_T} I\,[C] \right], \tag{4.4.5}$$

for $C \in \sigma\,(\mathbf{Y}_t : 0 \le t \le T)$*, it immediately follows that*

$$\mathbb{P}\left[\{\omega \in \Omega : \mathbf{X}(\cdot; \omega) \in A\}\right] = \widetilde{\mathbb{Q}} \left[\{\omega \in \widetilde{\Omega} : \mathbf{Y}(\cdot; \omega) \in A\} \right],$$

for any $A \in \mathcal{F}_T$*.*

Remark 4.4.3 *The conditions (4.4.1) and (4.4.2) are sufficient for the (local) absolutely continuity of the processes* $\mathbf{X}$ *and* $\mathbf{Y}$. *As it is indicated in the proof of Theorem 4.4.1, (4.4.2) combined with*

$$\int_{S^{d-1}} \int_0^\infty \left(\frac{\tilde{q}\,(r, \mathbf{u}) - 1}{r} \right)^2 r^{1-\alpha} dr \tilde{\sigma}(d\mathbf{u}) < \infty. \tag{4.4.6}$$

are both necessary and sufficient for the conclusion of Theorem 4.4.1. Now, suppose that $\tilde{q}$ *has Bernstein's representation*

$$\tilde{q}(r, \mathbf{u}) = \int_{[0,\infty)} e^{-rs} F(ds; \mathbf{u}),$$

where $\{F(ds; \mathbf{u})\}_{\mathbf{u} \in S^{d-1}}$ *is a measurable family of probability measures on* $[0, \infty)$. *Then, by the property (iii) of Section XIII.2 of [15], the condition (4.4.6) is equivalent to*

$$\int_{S^{d-1}} \int_0^\infty \int_0^\infty \frac{1}{(s_1 + s_2)^{2-\alpha}} F([s_1, \infty); \mathbf{u}) F([s_2, \infty); \mathbf{u}) ds_1 ds_2 \tilde{\sigma}(d\mathbf{u}) < \infty. \tag{4.4.7}$$

In particular, the above integral is finite if

$$\int_{S^{d-1}} (\mathbb{E}\,[R_{\mathbf{u}}])^2\, \tilde{\sigma}(d\mathbf{u}) < \infty,$$

where $R_{\mathbf{u}}$ *is a random variable with distribution* $F(\cdot; \mathbf{u})$. *The condition below is also sufficient for (4.4.7) to hold:*

$$\int_{S^{d-1}} \mathbb{E}\,[R_{\mathbf{u}}^\alpha]\, \tilde{\sigma}(d\mathbf{u}) < \infty.$$

4.5 Proofs of the main results.

Proof of Theorem 4.1.5: The Lévy-Khintchine representation for μ takes the form:

$$\hat{\mu}(\mathbf{z}) = \exp\left\{ i\,\mathbf{z} \cdot \mathbf{b} + \int_{\mathbb{R}_0^d} \left[e^{i\,\mathbf{z} \cdot \mathbf{x}} - 1 - i\,\mathbf{z} \cdot \mathbf{x}\, I\,(\|\mathbf{x}\| \le 1) \right] \nu(d\mathbf{x}) \right\}.$$

Using standard arguments of integration that goes from simple functions to integrable measurable functions, the integral in the exponent above is equal to

$$\int_{\mathbb{R}_0^d} \int_0^\infty \left[e^{is(\mathbf{z} \cdot \mathbf{x})} - 1 - is(\mathbf{z} \cdot \mathbf{x}) I\,(s\|\mathbf{x}\| \le 1) \right] s^{-\alpha-1} e^{-s} ds \rho(d\mathbf{x}). \tag{4.5.1}$$

Since for $0 < \alpha < 1$

$$\mathbf{c} \equiv \int_{\mathbb{R}_0^d} \mathbf{x} \int_0^\infty I\left(s\|\mathbf{x}\| \leq 1\right) s^{-\alpha} e^{-s} ds \rho(d\mathbf{x}) < \infty,$$

the integral (4.5.1) can be broken up into the two terms $\int_{\mathbb{R}_0^d} \int_0^\infty \left[e^{is(\mathbf{z}\cdot\mathbf{x})} - 1\right] s^{-\alpha-1} e^{-s} ds \rho(d\mathbf{x}) - i\, z \cdot \mathbf{c}$. Then, (4.1.9) will be true with $\mathbf{a} \equiv \mathbf{b} - \mathbf{c}$ since

$$\int_0^\infty \left[e^{is\omega} - 1\right] s^{-\alpha-1} e^{-s} ds = \frac{\Gamma(1-\alpha)}{\alpha} \left[1 - (1 - i\omega)^\alpha\right]. \tag{4.5.2}$$

Indeed, we can write the left hand side of (4.5.2) as

$$\begin{aligned} i \int_0^\infty \int_0^\omega e^{isv} dv s^{-\alpha} e^{-s} ds &= i \int_0^\omega \int_0^\infty e^{-s(1-iv)} s^{-\alpha} ds dv \\ &= \Gamma(1-\alpha) i \int_0^\omega (1 - iv)^{\alpha-1} dv. \end{aligned}$$

Note that the second equality above follows from the form of the characteristic function of the Gamma distributions. For $1 < \alpha < 2$, we take instead

$$\mathbf{c} \equiv -\int_{\mathbb{R}_0^d} \mathbf{x} \int_0^\infty I\left(s\|\mathbf{x}\| > 1\right) s^{-\alpha} e^{-s} ds \rho(d\mathbf{x}) < \infty,$$

so that (4.5.1) can written as $\int_{\mathbb{R}_0^d} \int_0^\infty \left[e^{is(\mathbf{z}\cdot\mathbf{x})} - 1 - is(\mathbf{z}\cdot\mathbf{x})\right] s^{-\alpha-1} e^{-s} ds \rho(d\mathbf{x}) - i\, z \cdot \mathbf{c}$. It suffices to show that

$$\int_0^\infty \left[e^{is\omega} - 1 - is\omega\right] s^{-\alpha-1} e^{-s} ds = \frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)} \left[(1 - i\omega)^\alpha - 1 + i\alpha\omega\right], \tag{4.5.3}$$

for any real number ω and any $0 < \alpha < 2$ $(\alpha \neq 1)$. Integrating by parts the left hand side in (4.5.3) and applying (4.5.2):

$$\begin{aligned} -\frac{1}{\alpha} \int_0^\infty \left[e^{is\omega} - 1 - is\omega\right] e^{-s} d(s^{-\alpha}) &= \frac{1}{\alpha} \int_0^\infty \left[(i\omega - 1)\left(e^{is\omega} - 1\right) + is\omega\right] s^{-\alpha} e^{-s} ds \\ &= \frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)} \left[(1 - i\omega)^\alpha - 1 + i\alpha\omega\right]. \end{aligned}$$

Now, if $\alpha = 1$, (4.5.1) can be written as

$$\int_{\mathbb{R}_0^d} \int_0^\infty \left[e^{is(\mathbf{z}\cdot\mathbf{x})} - 1 - is(\mathbf{z}\cdot\mathbf{x}) I\left(s \leq 1\right)\right] s^{-2} e^{-s} ds \rho(d\mathbf{x}) + i\, \mathbf{c} \cdot \mathbf{z},$$

where $\mathbf{c} \equiv \int_{\mathbb{R}_0^d} \mathbf{x} \int_0^\infty [I(s \le 1) - I(s\|x\| \le 1)]\, s^{-1} e^{-s} ds \rho(d\mathbf{x}) < \infty$. So, we need to evaluate integrals of the form

$$\int_0^\infty \left[e^{is\omega} - 1 - is\omega I(s \le 1)\right] s^{-2} e^{-s} ds. \tag{4.5.4}$$

Then, (4.5.4) can be written as

$$\Psi(1 - i\omega) - \Psi(1) - i\omega \int_0^1 [e^{-s} - 1]\, s^{-1} ds, \tag{4.5.5}$$

where

$$\Psi(y) \equiv \int_0^\infty [e^{-sy} - 1 + syI(s \le 1)]\, s^{-2} ds, \tag{4.5.6}$$

provided that Ψ is well-defined. By Theorem 25.17 of [39], Ψ is definable for any for $y \in \mathbb{C}$ with $\mathrm{Re}(y) \ge 0$. For any positive real y, we have that

$$\begin{aligned}\Psi(y) &= \int_0^{1/y} [e^{-sy} - 1 + sy]\, s^{-2} ds + \int_{1/y}^\infty [e^{-sy} - 1]\, s^{-2} ds + y \int_{1/y}^1 s^{-1} ds \\ &= cy + y \log(y),\end{aligned}$$

where $c = \int_0^\infty [e^{-v} - 1 + vI(v \le 1)]\, v^{-2} dv$. Since the function $\widetilde{\Psi}(y) = cy + y\log(y)$ is analytic on $D = \{c \in \mathbb{C} : \arg(c) \in (-\pi, \pi)\}$, by analyticity inside and continuity to the boundary, $\Psi(y) = \widetilde{\Psi}(y)$ for any $y \in \mathbb{C}$ with $\mathrm{Re}(y) > 0$ (see for instance p. 51 of [37] and references here in). Evaluating $\widetilde{\Psi}$ at $1 - i\omega$ and -1, (4.5.4) is equal to

$$(1 - i\omega)\log(1 - i\omega) - i\omega\left(\int_0^1 [e^{-s} - 1]\, s^{-1} ds + c\right).$$

We can easily infer (4.1.9) for the case $\alpha = 1$. □

Proof of Theorem 4.1.6: Since the processes involved in the limits are Lévy processes, we only need to prove the weak convergence of the marginal distributions at $t = 1$. Below, we will need the following expansion valid for any real number u and $n \in \mathbb{N}$:

$$e^{iu} = \sum_{k=0}^{n-1} \frac{(iu)^k}{k!} + \theta \frac{|u|^n}{n!}, \tag{4.5.7}$$

where $\theta \in \mathbb{C}$ satisfies $|\theta| \le 1$ and depends on u and n (see Lemma 8.6 of [39]).

(i) By Theorem 4.1.5, the characteristic function of $h^{-1/\alpha}\mathbf{X}_h(1) - \mathbf{a}_h$, for a constant vector

$\mathbf{a}_h$ to be determined, is

$$\exp\left\{k_\alpha \int_{\mathbb{R}_0^{d-1}} h\psi_\alpha\left(h^{-1/\alpha}(\mathbf{z}\cdot\mathbf{x})\right)\rho(d\mathbf{x}) - i\,\mathbf{a}_h\cdot\mathbf{z}\right\}. \tag{4.5.8}$$

Notice that

$$h\psi_\alpha(h^{-1/\alpha}\omega) = \begin{cases} h - \left(h^{1/\alpha} - i\,\omega\right)^\alpha, & \text{if } 0<\alpha<1,\\ \left(h^{1/\alpha} - i\,\omega\right)^\alpha - h - i\,\omega h^{1-1/\alpha}, & \text{if } 1<\alpha<2,\end{cases}$$

converges to $-(-i\omega)^\alpha$ as $h \to 0$. By (4.5.2) or (4.5.3), there exists a constant C_α depending only on α such that $|\psi_\alpha(\omega)| \le C_\alpha|\omega|^\alpha$ (indeed, make the change of variables $u = s\omega$, upper bound $e^{-u/\omega}$ by 1, and apply (4.5.7)). Therefore,

$$\left|h\psi_\alpha\left(h^{-1/\alpha}(\mathbf{z}\cdot\mathbf{x})\right)\right| \le C_\alpha|\mathbf{z}\cdot\mathbf{x}|^\alpha \le C_\alpha\|\mathbf{z}\|^\alpha\|\mathbf{x}\|^\alpha,$$

and thus the limit and the integral in (4.5.8) can be exchanged. Fixing $\mathbf{a}_h = \mathbf{0}$, the limit yields

$$\exp\left\{-k_\alpha \int_{\mathbb{R}_0^{d-1}} (i\mathbf{z}\cdot\mathbf{x})^\alpha\rho(d\mathbf{x})\right\}, \tag{4.5.9}$$

which is equal to (4.1.11) with $c_\alpha = k_\alpha\cos(\pi\alpha/2)$ (remember that $\arg(v) \in (-\pi,\pi]$ in $v^\alpha = |v|e^{i\,\alpha\arg(v)}$).

(ii) If $\alpha = 1$ and $\mathbf{a}_h$ is chosen to be $\log(h)\int_{\mathbb{R}_0^d}\mathbf{x}\rho(d\mathbf{x})$, (4.5.8) becomes

$$\exp\left\{\int_{\mathbb{R}_0^{d-1}} \left((h - i\,\mathbf{z}\cdot\mathbf{x})\log\left(h - i\,\mathbf{z}\cdot\mathbf{x}\right) - h\log(h)\right)\rho(d\mathbf{x})\right\}.$$

As $h \to 0$, the expression above converges to $\exp\left\{\int_{\mathbb{R}_0^{d-1}}(-i\,\mathbf{z}\cdot\mathbf{x})\log\left(-i\,\mathbf{z}\cdot\mathbf{x}\right)\rho(d\mathbf{x})\right\}$, which is itself equal to (4.1.13) because $\log(-i\mathbf{z}\cdot\mathbf{x}) = \log|\mathbf{z}\cdot\mathbf{x}| - i(\pi/2)\mathrm{sgn}(\mathbf{z}\cdot\mathbf{x})$.

(iii) Let $1 < \alpha < 2$. By (4.5.3),

$$h\psi_\alpha(h^{-1/2}\omega) = \frac{\alpha(\alpha-1)}{\Gamma(2-\alpha)}\int_0^\infty h\left[e^{ish^{-1/2}\omega} - 1 - ish^{-1/2}\omega\right]s^{-\alpha-1}e^{-s}ds.$$

Then, (4.5.7) with $n = 2$ implies that $|h\psi_\alpha(h^{-1/2}\omega)| \le \alpha(\alpha-1)\omega^2/2$, and from the bounded convergence theorem,

$$\lim_{h\to\infty}\int_{\mathbb{R}_0^{d-1}} h\psi_\alpha(h^{-1/2}(\mathbf{z}\cdot\mathbf{x}))\rho(d\mathbf{x}) = \frac{\alpha(\alpha-1)}{2}\int_{\mathbb{R}_0^{d-1}}|\mathbf{z}\cdot\mathbf{x}|^2\rho(d\mathbf{x}).$$

Above, we apply (4.5.7) with $n = 3$ to evaluate the limit. This proves (4.1.15) because according to Theorem 4.1.5 the characteristic function of $h^{-1/2}\mathbf{X}_h(1)$ is given by

$$\exp\left\{k_\alpha \int_{\mathbb{R}_0^{d-1}} h\psi_\alpha\left(h^{-1/\alpha}(\mathbf{z}\cdot\mathbf{x})\right)\rho(d\mathbf{x})\right\}.$$

where $k_\alpha = \Gamma(2-\alpha)/(\alpha(1-\alpha))$.

(iv) Take $0 < \alpha < 1$ and $\mathbf{a}_h \equiv h^{1/2}\Gamma(1-\alpha)\int_{\mathbb{R}_0^d} \mathbf{x}\rho(d\mathbf{x})$. The characteristic function of $h^{-1/2}\mathbf{X}_h(1) - \mathbf{a}_h$ is

$$\exp\left\{\frac{\Gamma(1-\alpha)}{\alpha}\int_{\mathbb{R}_0^{d-1}} h\left[1-\left(1-ih^{-1/2}\mathbf{z}\cdot\mathbf{x}\right)^\alpha - i\alpha h^{-1/2}\mathbf{z}\cdot\mathbf{x}\right]\rho(d\mathbf{x})\right\}.$$

Then, we can proceed as in the case (iii). □

Proof of Theorem 4.2.1: The series representation (4.2.3) is well known in the literature. It has a long history from Gnedenko to LePage and beyond (see [34] and the references therein). We present below another method of proof in the context of the *Shot Noise Method* of [34] because of its relevance for the case of tempered stables laws (for a review of this method see Section 3.1.2). Consider the marked Poisson process $N(\cdot) = \sum_{i=1}^\infty \delta_{(\Gamma_i,\mathbf{V}_i)}(\cdot)$ on $\mathbb{R}_+ \times \mathbb{R}_0^d$ with mean measure $\rho(dt, d\mathbf{x}) \equiv dt\rho_1(d\mathbf{x})$ (see Proposition 3.8. of [32]), and take the transformation $\mathbf{H} : \mathbb{R}_+ \times \mathbb{R}_0^d \to \mathbb{R}_0^d$ defined by

$$\mathbf{H}(t, \mathbf{v}) \equiv \left(m(\rho)\,(\alpha t)^{-1/\alpha}\right)\frac{\mathbf{v}}{||\mathbf{v}||}.$$

Then, $N\circ\mathbf{H}^{-1}(\cdot) = \sum_{i=1}^\infty \delta_{\mathbf{H}(\Gamma_i,\mathbf{V}_i)}(\cdot)$ is a Poisson process on $\mathbb{R}_0^d$, with mean measure $\rho\circ\mathbf{H}^{-1}(\cdot)$. A straightforward evaluation of (4.1.7) and of $\rho\circ\mathbf{H}^{-1}(\cdot)$ for sets of the form $(a,\infty)C = \left\{\mathbf{x}\in\mathbb{R}_0^d : \mathbf{x}/||\mathbf{x}|| \in C, ||\mathbf{x}|| > a\right\}$ shows the equality between these two measures. Indeed,

$$\begin{aligned}
\rho\circ\mathbf{H}^{-1}((a,\infty)C) &= \int_{\mathbb{R}_+\times\mathbb{R}_0^d} I\left(\mathbf{H}(t,\mathbf{x})\in(a,\infty)C\right)dt\rho_1(d\mathbf{x}) \\
&= \int_{\mathbb{R}_0^d} \frac{1}{\alpha}\left(\frac{m(\rho)}{a}\right)^\alpha I_C\left(\frac{\mathbf{x}}{||\mathbf{x}||}\right)||\mathbf{x}||^\alpha\rho_1(d\mathbf{x}) \\
&= \int_{\mathbb{R}_0^d}\int_a^\infty r^{-\alpha-1}dr I_C\left(\frac{\mathbf{x}}{||\mathbf{x}||}\right)||\mathbf{x}||^\alpha\rho(d\mathbf{x}) \\
&= \gamma((a,\infty)C).
\end{aligned}$$

Therefore, the marked Poisson process $N(\cdot) = \sum_{i=1}^{\infty} \delta_{(T_i, \mathbf{H}(\Gamma_i, \mathbf{V}_i))}(\cdot)$ has the same law as the random measure $\mathcal{J}_Y$ associated with the jumps of the Lévy process $\mathbf{Y}$. From the arguments of [34] Section 5, if the function

$$\mathbf{A}(s) \equiv \int_0^s \int_{\mathbb{R}_0^d} \mathbf{H}(r, \mathbf{v}) I\left(\|\mathbf{H}(r, \mathbf{v})\| \leq 1\right) \rho_1(d\mathbf{v}) dr, \tag{4.5.10}$$

is such that, a.s.

$$\lim_{n \to \infty} \left(\mathbf{A}(\Gamma_n) - \mathbf{A}(n)\right) = 0, \tag{4.5.11}$$

then the series

$$\tilde{\mathbf{Y}}(t) \equiv \sum_{i=1}^{\infty} \left(\mathbf{H}(\Gamma_i, \mathbf{V}_i) I\left(T_i \leq t\right) - t\left(\mathbf{A}(i) - \mathbf{A}(i-1)\right)\right), \tag{4.5.12}$$

converges (uniformly in $t \in [0, 1]$) a.s. Moreover, the process $\left\{\tilde{\mathbf{Y}}(t)\right\}_{t \in [0,1]}$ is a pure jump Lévy process with the same Lévy measure as the process $\mathbf{Y}$. Hence, the two processes have the same law up to a term of the form $\tilde{\mathbf{b}}_1 t$. The relation (4.5.11) holds for any $\alpha \in (0, 2)$, because for each $\mathbf{v} \in \mathbb{R}_0^d$, $r \to \|\mathbf{H}(r, \mathbf{v})\|$ is non-decreasing (see p. 409 of [34]). When $0 < \alpha < 1$, then $\lim_{s \to \infty} \mathbf{A}(s)$ converges to $\hat{\mathbf{b}}_1 \equiv \int_{\|x\| \leq 1} \mathbf{x} \gamma(d\mathbf{x})$ and

$$\tilde{\mathbf{Y}}(t) = -\hat{\mathbf{b}}_1 t + \sum_{i=1}^{\infty} \left(\mathbf{H}(\Gamma_i, \mathbf{V}_i) I\left(T_i \leq t\right)\right), \tag{4.5.13}$$

in the sense of convergence of the finite dimensional distributions. Then, the representation (4.2.3) follows by taking $\mathbf{b}_1 = \hat{\mathbf{b}}_1 + \mathbf{b}$, where $\mathbf{b}$ is as in (4.2.1). A similar argument works when ρ is a symmetric measure since $\mathbf{A}(\cdot)$ will be identically equal to zero.

To prove the representation (4.2.4), we follow the same technique starting now from the Poisson process $M(\cdot) \equiv \sum_{i=1}^{\infty} \delta_{(\Gamma_i, \mathbf{V}_i, E_i, U_i)}(\cdot)$ and applying the transformation

$$\mathbf{H}_1(t, \mathbf{v}, e, u) = \left\{\left(m(\rho)\left(\alpha t\right)^{-1/\alpha}\right) \wedge \left(\|\mathbf{v}\| e u^{1/\alpha}\right)\right\} \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

The mean measure of M is $\rho_1(dt, d\mathbf{x}, ds, du) \equiv dt \rho_1(d\mathbf{x}) e^{-s} ds du$ on $\mathbb{R}_+ \times \mathbb{R}_0^d \times \mathbb{R}_+ \times [0, 1]$, and the transformed Poisson process $\sum_{i=1}^{\infty} \delta_{\mathbf{H}(\Gamma_i, \mathbf{V}_i, E_i, U_i)}(\cdot)$ has mean measure determined as follows:

$$
\begin{aligned}
\rho \circ \mathbf{H}_1^{-1}((a,\infty)C) &= \int_0^\infty \int_{\mathbb{R}_0^d} \int_0^\infty \int_0^1 I\left(\mathbf{H}_1(t,\mathbf{x},s,u) \in (a,\infty)C\right) e^{-s} du ds \rho_1(d\mathbf{x}) dt \\
&= \int_{\mathbb{R}_0^d} \frac{1}{\alpha} a^{-\alpha} \int_0^\infty \int_0^1 I_{(a,\infty)}\left(s u^{\frac{1}{\alpha}} \|\mathbf{x}\|\right) du e^{-s} ds I_C\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \|\mathbf{x}\|^\alpha \rho(d\mathbf{x}) \\
&= \int_{\mathbb{R}_0^d} \int_0^\infty I_{(a,\infty)}\left(s\|\mathbf{x}\|\right) \frac{a^{-\alpha} - (s\|\mathbf{x}\|)^{-\alpha}}{\alpha} e^{-s} ds I_C\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \|\mathbf{x}\|^\alpha \rho(d\mathbf{x}) \\
&= \int_{\mathbb{R}_0^d} \int_0^\infty I_{(a,\infty)}(s) e^{-s/\|\mathbf{x}\|} s^{-\alpha-1} ds I_C\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \|\mathbf{x}\|^\alpha \rho(d\mathbf{x}) \\
&= \nu((a,\infty)C).
\end{aligned}
$$

Since $r \to \|\mathbf{H}_1(r,\mathbf{v},e,u)\|$ is nondecreasing, the series

$$
\tilde{\mathbf{Y}}(t) \equiv \sum_{i=1}^\infty \left(\mathbf{H}_1(\Gamma_i, \mathbf{V}_i, E_i, U_i) I\left(T_i \le t\right) - t\left(\mathbf{A}_1(i) - \mathbf{A}_1(i-1)\right)\right), \tag{4.5.14}
$$

converges to a pure jump Lévy process with Lévy measure ν, where $\mathbf{A}_1$ is given by

$$
\mathbf{A}_1(s) \equiv \int_0^s \int_{\mathbb{R}_0^d} \int_0^\infty \int_0^1 \mathbf{H}_1(r,\mathbf{x},s,u) I\left(\|\mathbf{H}_1(r,\mathbf{x},s,u)\| \le 1\right) du e^{-s} ds \rho_1(d\mathbf{x}) dr. \tag{4.5.15}
$$

Clearly, $\mathbf{A}_1(s)$ is identically 0 if ρ is symmetric, while $\lim_{s\to\infty} \mathbf{A}_1(s) = \int_{\|x\|\le 1} \mathbf{x}\nu(d\mathbf{x})$ if $0 < \alpha < 1$. In any case, the representation (4.2.4) follows. □

Proof of Theorem 4.3.1: Consider a *spectral decomposition* of the measure ρ of the form

$$
\rho(A) = \int_{S^{d-1}} \int_0^\infty I_A(r\mathbf{u}) F(\mathbf{u}, dr) \sigma(d\mathbf{u}), \tag{4.5.16}
$$

where σ is a probability measure on S^{d-1} and $\{F(\mathbf{u},\cdot)\}_{\mathbf{u}\in S}$ is a measurable family of measures on $\mathcal{B}((0,\infty))$ (or transition kernel from S^{d-1} to $(0,\infty)$) such that $F(\mathbf{u},\{0\}) = 0$. Such a representation can be deduced from disintegration results[2] like Theorems 5.3 and 5.4. of [20]. It is enough to prove (4.3.1) for sets of the form

$$
BC = \{\mathbf{x} : \|\mathbf{x}\| \in B, \mathbf{x}/\|\mathbf{x}\| \in C\},
$$

[2] Concretely, we can take σ as $\frac{1}{m}\int_{\mathbb{R}_0^d} \|\mathbf{x}\|^\alpha I_C\left(\mathbf{x}/\|\mathbf{x}\|\right) \rho(d\mathbf{x})$, for a suitable normalizing constant m, and $F(\mathbf{u}, dr) = m r^{-\alpha} \pi(\mathbf{u}, dr)$, where $\{\pi(\mathbf{u},\cdot)\}_{\mathbf{u}\in S}$ is the regular version of $\mathbb{P}[R \in \cdot | \mathbf{U} = \mathbf{u}]$. Here, $\mathbb{P}$ is a probability measure on $\Omega = \mathbb{R}_0^d$ defined by $\mathbb{P}(A) = \frac{1}{m}\int_A \|\mathbf{x}\|^\alpha \rho(d\mathbf{x})$, while $R(\omega) = \|\omega\|$ and $\mathbf{U} = \omega/\|\omega\|$, for $\omega \in \Omega$.

where B and C are Borel subsets of $(0,\infty)$ and S^{d-1}, respectively. Then, the measure ν is of the form (4.1.5) if and only if

$$\begin{aligned}
\nu(BC) &= \int_{S^{d-1}}\int_0^\infty\int_0^\infty I_{BC}(sr\mathbf{u})s^{-\alpha-1}e^{-s}dsF(\mathbf{u},dr)\sigma(d\mathbf{u})\\
&= \int_C\int_0^\infty\int_0^\infty I_B(sr)s^{-\alpha-1}e^{-s}dsF(\mathbf{u},dr)\sigma(d\mathbf{u})\\
&= \int_C\int_0^\infty\int_0^\infty I_B(v)v^{-\alpha-1}e^{-v/r}dvr^\alpha F(\mathbf{u},dr)\sigma(d\mathbf{u})\\
&= \int_{S^{d-1}}\int_0^\infty I_{BC}(v\mathbf{u})v^{-\alpha-1}q(\mathbf{u},v)dv\sigma(d\mathbf{u}),
\end{aligned}$$

where

$$q(\mathbf{u},v) = \int_0^\infty e^{-v/r}r^\alpha F(\mathbf{u},dr). \tag{4.5.17}$$

We thus proved that ν is of the form (4.1.5) if and only if ν is of the form (4.3.1) with $q(\mathbf{u},\cdot)$ given by (4.5.17) for each $\mathbf{u}\in S^{d-1}$. By Bernstein's Theorem and the change of variables $r\to 1/r$, $q(\mathbf{u},\cdot)$ is as in (4.5.17), for $F(\mathbf{u},\cdot)$ such that $F(\mathbf{u},\{0\})=0$, if and only if it is completely monotone with $q(\mathbf{u},\infty)=0$. Further, by the monotone convergence theorem, $q(\mathbf{u},0^+)=\int_0^\infty r^\alpha F(\mathbf{u},dr)$, and the condition

$$\int_{\mathbb{R}_0^d}||\mathbf{x}||^\alpha\rho(d\mathbf{x}) = \int_{S^{d-1}}\int_0^\infty r^\alpha F(\mathbf{u},dr)\sigma(d\mathbf{u}) < \infty,$$

is equivalent to $\int_{S^{d-1}} q(\mathbf{u},0^+)\sigma(d\mathbf{u})<\infty$. □

Proof of Theorem 4.4.1: Clearly, for any measurable nonnegative function h

$$\int_{\mathbb{R}_0^d} h(\mathbf{x})\gamma(d\mathbf{x}) = \int_{S^{d-1}}\int_0^\infty h(r\mathbf{u})r^{-\alpha-1}dr\tilde{\sigma}(d\mathbf{u}).$$

Then,

$$g(\mathbf{x}) \equiv \frac{d\nu}{d\gamma}(\mathbf{x}) = \tilde{q}\left(||\mathbf{x}||,\frac{\mathbf{x}}{||\mathbf{x}||}\right).$$

By Theorem 33.1 of [39], for $\mathbf{X}$ and $\mathbf{Y}$ to be locally absolutely equivalent, it is necessary and sufficient that (4.4.2) holds and the integration condition below holds

$$\int_{S^{d-1}}\int_0^\infty \left(\sqrt{\tilde{q}\,(r,\mathbf{u})}-1\right)^2 r^{-\alpha-1}dr\tilde{\sigma}(d\mathbf{u}) < \infty. \tag{4.5.18}$$

Since the $0 < \tilde{q}(\cdot, \mathbf{u}) \leq 1$, the above inequality is equivalent to

$$\int_{S^{d-1}} \int_0^1 \left(\frac{\tilde{q}\,(r, \mathbf{u}) - 1}{r} \right)^2 r^{1-\alpha} dr \tilde{\sigma}(d\mathbf{u}) < \infty.$$

Then, (4.5.18) will follow from (4.4.1), since $0 < \alpha < 2$ and

$$\left(\frac{\tilde{q}\,(r, \mathbf{u}) - 1}{r} \right)^2 \leq \left(\tilde{q}'(0^+, \mathbf{u})\right)^2,$$

when $\tilde{q}''(r, \mathbf{u}) > 0$. By Theorem 33.2 of [39], the equation (4.4.4) is satisfied when the process U_T is defined by

$$U_T = \lim_{\varepsilon \to 0} \left(\int_{[0,T] \times \{\|\mathbf{x}\| > \varepsilon\}} \ln g(\mathbf{x}) \mathcal{J}_{\mathbf{Y}}(dt, d\mathbf{x}) - T \int_{\{\|\mathbf{x}\| > \varepsilon\}} (g(\mathbf{x}) - 1)\, \gamma(d\mathbf{x}) \right),$$

(see (33.7) and (33.9) in [39]). □

REFERENCES

[1] Akaike, H., “Information theory and an extension of maximum likelihood principle,” *In Proceeding 2nd international symposium on information theory by P.N. Petrov & F. Csaki, eds*, pp. 267–281, 1973.

[2] Asmussen, S. and Rosiński, J., “ Approximations of small jumps of Lévy processes with a view towards simulation,” *Journal of Applied Probability*, vol. 38, pp. 482–493, 2001.

[3] Barndorff-Nielsen, O. and Shephard, N., “Modelling by Lévy processess for financial economics,” *Lévy Processes. Theory and Applications, by O.E. Barndorff-Nielsen, T. Mikosch, and S.I. Resnick*, pp. 283–318, 2001.

[4] Barndorff-Nielsen, O. and Shephard, N., “Normal modified stable processes,” *Theory of Probability and Mathematical Statistics*, vol. 65, pp. 1–19, 2001.

[5] Bertoin, J., *Lévy processes*. Cambridge University Press, 1996.

[6] Bibby, B. and Sorensen, M., “Hyperbolic Processes in Finance,” *Handbook of heavy tailed distributions in finance, by S. Rachev (ed.)*, pp. 211–248, 2003.

[7] Birgé, L. and Massart, P., “From model selection to adaptive estimation,” *In Festschrift for Lucien Le Cam*, pp. 55–87, 1997.

[8] Bondesson, L., *Generalized Gamma Convolutions and Related Classes of Disributions*. Springer, New York, 1992.

[9] Boyarchenko, S. and Levendorskii, S., “Option pricing for truncated Lévy processes,” *International journal of theoretical and applied finance*, vol. 3, pp. 549–552, 2000.

[10] Campbell, J., Lo, A., and MacKinlay, A., *The econometrics of Financial Markets*. Princeton, 1997.

[11] Carr, P., Geman, H., Madan, D., and Yor, M., “The fine structure of asset returns: An empirical investigation,” *Journal of Business*, Apr. 2002.

[12] Carr, P., Geman, H., Madan, D., and Yor, M., “Stochastic volatility for Lévy processes,” *Mathematical Finance*, vol. 13, pp. 345–, July 2003.

[13] Carr, P., Madan, D., and Chang, E., “The variance Gamma process and option pricing,” *European Finance Review*, vol. 2, pp. 79–105, 1998.

[14] Eberlein, E., “Application of Generalized Hyperbolic Lévy Motions to Finance,” *Lévy Processes. Theory and Applications, by O.E. Barndorff-Nielsen, T. Mikosch, and S.I. Resnick*, pp. 319–336, 2001.

[15] Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. 2. John Wiley & Sons, 1971.

[16] Gnedenko, B. and Kolmogorov, A., *Foundations of Modern Probability*. Springer-Verlag, Berlin, New York, Heidelberg, 1997.

[17] Hautsch, N. and Pohlmeier, W., "Econometric Analysis of financial transaction data: pitfalls and opportunities," Tech. Rep. 01-05, Center of finance and econometrics, 2001. Also available in http://econpapers.hhs.se/paper/.

[18] Houdré, C. and Privault, N., "Concentration and deviation inequalities in infinite dimensions via covariance representations," *Bernoulli*, 2002.

[19] Jiang, W., *Some simulation-based models towards mathematical finance*. PhD thesis, University of Aarhus, Apr. 2000.

[20] Kallenberg, O., *Foundations of Modern Probability*. Springer-Verlag,Berlin, New York, Heidelberg, 1997.

[21] Karr, A., *Point processes and their statistical inference*. Marcel Dekker Inc., 1991.

[22] Koponen, I., "Analytic approach to the problem of convergence of truncated Lévy flights toward the Gaussian stochastic process," *Phys. Rev. E*, vol. 52, pp. 1197–1199, 1995.

[23] Kutoyants, Y., *Statistical Inference for Spatial Poisson Processes*. Springer, 1998.

[24] Madan, D. and Seneta, E., "The variance Gamma model for share market returns," *Journal of Business*, vol. 63, pp. 511–524, 1990.

[25] Mallows, C., "Some comments on C_p," *Technometrics*, vol. 15, pp. 661–675, 1973.

[26] Mandelbrot, B., "The variation of certain speculative prices," *The Journal of Business*, vol. 36, pp. 394–419, 1963.

[27] Mantegna, R. and Stanley, H., "Stochastic process with ultraslow convergence to a Gaussian: The truncated Lévy flight," *Phys. Rev. E*, 1995.

[28] Novikov, E., "Infinitely divisible distributions in tubulence," *Phys. Rev. E, 50:R3303-R3305*, 1989.

[29] Pollard, D., *Convergence of Stochastic Processes*. Springer-Verlag, 1984.

[30] Press, S., "a compound event model for security prices," *The Journal of Business*, pp. 317–335, 1967.

[31] Raible, S., *Lévy processes in Finance: Theory, Numerics, and Empirical Facts*. PhD thesis, Albert-Ludwigs-Universitat Freiburg, Jan. 2000.

[32] Resnick, S., *Extreme Values, Regular variation, and Point Processes*. Springer-Verlag,Berlin, New York, Heidelberg, 1987.

[33] Reynaud-Bouret, P., “Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities,” Tech. Rep. 1102-001, School of mathematics, Georgia Institute of Technology, 2002.

[34] Rosiński, J., “Series representations of Lévy processes from the perspective of point processes,” *In Lévy processes-Theory and Applications*, pp. 401–415, 2001.

[35] Rosiński, J., “Tempered stable processes,” tech. rep., University of Tennessee, 2002.

[36] Rüschendorf, L. and Woerner, J., “Expansion of transition distributions of Lévy processes in small time,” *Bernoulli*, vol. 8, pp. 81–96, 2002.

[37] Samorodnitsky, G. and Taqqu, M., *Stable non-Gaussian random processes - Stochastic models with infinity variance*. Chapman & Hall, 1994.

[38] Samuelson, P., “Rational theory of warrant pricing,” *Industrial management review*, no. 6, pp. 13–32, 1965.

[39] Sato, K., *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999.

[40] Schwarz, G., “Estimating the dimension of a model,” *Annals of Statistics*, vol. 6, pp. 461–464, 1978.

INDEX

VITA

José Enrique Figueroa-López, the son of Alicia López Garcia and Enrique Figueroa Romero, was born on July 9, 1973 in a small city of Mexico: Tepic, Nayarit. At the age of seventeen he was selected for the Olympiads in mathematics among the best fifteen national contestants, and he received instruction in number theory, geometry, and general problem-solving techniques.

Following a familiar tradition, he left his hometown and started to live by himself (under the close watch of his parents and his sister) pursuing a Bachelor of Science in Mathematics at the Universidad Autonoma Metropolitana, Mexico City. After earning his Bachelor in 1995, he moved to the beautiful city of Guanajuato, and enrolled at the Universidad de Guanajuato in the Master program of Statistics. Under the direction of Professor Víctor Pérez-Abreu, he completed a thesis on construction of self-similar processes with finite variance. He earned his master in Statistics in the summer of 1998.

He began his doctoral studies at the Georgia Institute of Technology in the fall of 1998. During his first two years at Georgia Tech, he was interested in Information Theory and Coding Theory. After taking some classes in Finance, he decided to concentrate on mathematical finance as this area suited better his applied and theoretical background on Probability, Stochastic Processes, Statistics and computational methods. He completed a Master in Quantitative Computational Finance in the spring of 2002. During all this time, he was working under the direction of Professor Christian Houdré on both fields: information theory and mathematical finance. He successfully defended his thesis, "Nonparametric estimation of Lévy processes with a view towards mathematical finance", in November 2003.