Stochastic inference for discretely observed compound Poisson processes

Vasilev Nikolay

June 28, 2023

Contents

1 Introduction

Compound Poisson processes are an essential tool in probability theory, statistics, and applied mathematics. They are stochastic processes that model the behavior of random variables that are a sum of a large number of independent, identically distributed random variables. Compound Poisson processes have been extensively studied in the literature due to their wide range of applications in various fields, including finance, insurance, telecommunications, and physics.

Poisson processes in particular as Lévy processes are used in modeling some events in which a strong shock may exist. This allows to make more accurate description of events with unexpected events. Examples of such events are modeling of highly volatile financial instruments.

Moreover, in this paper we focus on the study of the situation when the values of the processes cannot be observed continuously. For example, we may receive data at

some arbitrary frequency when accessing the server, and we do not have access to the full picture of the data. Such situations for the most part occur in real life, often we cannot reconstruct the full picture of some process, so in such a case we cannot use classical methods to find model parameters. Our task is to decompose the process into increments in a certain period and try to reconstruct by mathematical methods what could have happened in each moment of time. For example if there was a strong random jump at a certain point in time, we can catch it with a frequency decomposition on the spectrum, then all big or small jumps will become more obvious and we can evaluate them.

In this paper, we aim to investigate the practical utility of compound Poisson processes. We will start by providing a brief overview of the theory of compound Poisson processes and their fundamental properties. Then, we will discuss the different ways in which compound Poisson processes have been applied in real-world scenarios. Specifically, we will examine their use in modeling insurance claim amounts, analyzing stock market returns, and predicting network traffic.

Furthermore, we will also explore some of the challenges associated with using compound Poisson processes in practical applications. For instance, we will discuss issues related to parameter estimation, model selection, and model validation. To overcome these challenges, we will present some advanced techniques and methods that have been developed to improve the accuracy and reliability of compound Poisson process models.

Overall, this paper will provide a comprehensive overview of the practical utility of compound Poisson processes and highlight their significance in various fields. We believe that our findings will be useful for researchers, practitioners, and students who are interested in understanding and utilizing compound Poisson processes in their work.

2 Theory

2.1 General definitions

Definition 2.1. First, for convenience, let us introduce the definition of probability and event. A probability space (Ω, F, P) is a triplet of a set Ω, F is a family of subsets of set Ω, and P is a mapping from F into $\mathbb{R}$ satisfying the following conditions:
(1) $\Omega \in F, \emptyset \in F$($\emptyset$ is the empty set).
(2) If $A_n \in F$, n $\in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n$ and $\bigcap_{n=1}^{\infty} A_n$ are in F.
(3) If $A \in F$, then $A^c \in F$, where A^c is set Ω without subset A, in other words A^c is the complement of A.
(4) $0 \leq P[A] \leq 1, P[\Omega] = 1$, and $P[\emptyset] = 0$.
(5) If $A_n \in F$, $n \in \mathbb{N}$ and they are disjoint (that is, $A_n \cap A_m = \emptyset$ for $n \neq m$), then $P\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} P\left[A_n\right]$.

F called a family of subsets in Ω, and if $A \in F$ then A is called an event from family F. $\mathbb{P}$ is a probability measure of specific event, then $\mathbb{P}\left[A\right]$ called a probability of event A to occur. The pair (Ω, F) is a measurable space.

As a next step, let's define a continuous stochastic process, since it is an important property of the Lévy process.

Definition 2.2. A stochastic process X_t on $\mathbb{R}^d$ called stochastically continuous or con-

tinuous in probability if the following condition is met , for every $t \geq 0$ and $\varepsilon > 0$,

$$\lim_{s \to t} P\left[|X_s - X_t| > \varepsilon\right] = 0.$$

2.2 Poisson processes

Let $Y_1, Y_2, ..., Y_N$ be independent and identically distributed random variables with a common distribution function F. Let N_t be a counting process with intensity parameter λ, and S_n is a renewal process:

$$S_n = S_{n-1} + \xi_n, \quad S_0 = 0, \quad n = 1, 2, \ldots,$$

where $\xi_1, \xi_2, \ldots$ are independent and identically distributed random variables from exponential distribution with specific intensity. Then we can define counting in each time point t as maximum k such as value of renewal process less than time t. In other words k is a last time when renewal occurred.

$$N_t := \max\{k : S_k \leq t\}.$$

As mentioned in Bingham and Kiesel [2004] Poisson process has special property called lack of memory property:

$$\mathbb{P}\{S_n > t + h | S_n > h\} = \mathbb{P}\{S_n > t\}, \quad \text{where } t > 0 \text{ and } h > 0.$$

This property is also satisfied for compound Poisson processes, which will be defined later in the paper. This property allows us to estimate more accurately the property of the mentioned processes. In simple words, regardless of the moment in time when we observe a process, its properties do not change and the probability of observing jumps and any change in the process is the same.

2.3 Lévy processes

Compound Poisson processes are also a class of Lévy processes, so for further work with stochastic processes of this type, we need to formulate definitions and properties of the Lévy process.

To some extent, Lévy processes are a generalization of compound Poisson processes. They have some kind of constant displacement, independent Brownian motion and random jumps can occur at a random point in time within this process, but at the same time infinitesimal jumps can occur. These processes may be useful in modeling some more complex situations as opposed to the rest, where various unpredictable shocks may occur in the process. For example, these processes are well suited for modeling highly volatile assets in financial markets whose behavior is difficult to predict. Or economic indicators of some country whose economy is highly exposed to external market shocks. Also these types of processes are widely used in engineering physics and biology, which emphasizes the importance of this type of processes.

Definition 2.3. A stochastic process X_t, where $t \in \mathbb{R}^+$ on $\mathbb{R}^d$ is a Lévy process if the following conditions are satisfied:
(1) Independent increments property: for any $n \geq 1$, where $n \in \mathbb{B}$ and $0 \leq t_0 < t_1 <$

$\cdots < t_n$, random variables $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent .
(2) $X_0 = 0$ assumed by convenience.
(3) The distribution of $X_{s+t} - X_s$ does not depend on s means that it has stationary increments.
(4) It is stochastically continuous according with definition 2.2 .
(5) There is $\Omega_0 \in \mathcal{F}$ with $P[\Omega_0] = 1$ such that, for every $\omega \in \Omega_0, X_t(\omega)$ is right-continuous in $t \geq 0$ and has left limits in $t > 0$, means that this process is almost sure cádlág.

The Levy measure is a fundamental concept in the study of Levy processes, which are stochastic processes that exhibit many of the same properties as Brownian motion. It is a measure that describes the jump behavior of a Levy process, and plays a crucial role in characterizing the behavior of these processes. In particular, it specifies the distribution of the jumps of the process, including their size and frequency. By characterizing the distribution of jumps within these processes, it provides valuable insights into their underlying dynamics and enables accurate modeling of complex phenomena in a wide range of fields. The Lévy measure describes the intensity and size of the jumps in a process, and it plays a crucial role in the Lévy-Khintchine formula, which expresses the characteristic function of a process in terms of its Lévy measure.

Definition 2.4. Lévy measure ν can be defined as expected number of non-zero jumps in the time interval $[0; 1]$:

$$\nu(B) = \mathbb{E}\left[\#\{J_t \in B \ : t \in [0;1]\}\right],$$

where J is jump in the process and B is subset $B \subset \mathbb{R}\backslash\{0\}$

$$\begin{aligned}
\nu &= \lambda \mathbb{P}\{\xi \in B\} \\
&= \lambda \left(w_1 \mathbb{P}\{\xi_{ac} \in B\} + w_2 \mathbb{P}\{\xi_d \in B\}\right) \\
&= \lambda w_1 \sum_i \delta_{C_i}(B) p_i + \lambda w_2 \mathbb{P}\{\xi_{ac} \in B\} \\
&= \sum_{i=1}^{n} q_i \delta_{C_i}(B) + \lambda \int_B p(u) du,
\end{aligned} \tag{1}$$

Lévy measure can be also defined in the followng way:

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

where $F(t)$ is jump distributions and $N(t)$ is a Lévy distribution.

$$\nu = \nu_d + \nu_{ac}, \tag{1.1.1}$$

$$\nu_d = \sum_{j \in \mathcal{J}} q_j \delta_{J_j}, \quad \text{where } \mathcal{J} \text{ is countable }, q_j \in [0; \lambda], \tag{1.1.2}$$

ν is a Lévy measure, $\mathcal{J} \in \mathbb{R}\backslash\{0\}$ ν_d and ν_{ac} are finite measures that are discrete and absolutely continuous with respect to Lebesgue's measure. A completely continuous measure is determined by the density of the distribution, while a discrete measure is determined by the delta-measure. In order to work with actual weights of the jumps we can define their influence in distribution:

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

where $q = \sum\limits_{j \in \mathcal{J}} q_j$. The intuition behind the Lévy measure is that it provides a way to quantify the severity and frequency of jumps in a process. The Lévy measure is defined as the probability of observing a jump of a certain size within a given time interval. The larger the jump and the shorter the time interval, the higher the value of the Lévy measure.

The Lévy measure is a fundamental concept in the study of Lévy processes, which are stochastic processes with stationary and independent increments. Lévy processes are widely used in mathematical finance to model the behavior of asset prices, as well as in physics to model the behavior of particles in complex systems.

One of the practical applications of the Lévy measure is in the pricing of financial derivatives, such as options. The Lévy measure is used to estimate the probability of extreme events, such as large market movements, which can have a significant impact on the value of a derivative. By incorporating the Lévy measure into the pricing model, traders and investors can better manage the risk associated with these extreme events.

The Levy-Ito decomposition is an important method in the theory of random processes, used to decompose complex processes into simpler, independent components. The decomposition method has applications in various fields, including physics, engineering, and biology. Understanding the Levy-Ito decomposition is essential for accurately modeling a wide range of phenomena in these fields. In particular, it can be useful for representing the characteristic function, a tool used for identifying and analyzing the properties of random variables. The purpose of this scientific article is to explore the Levy-Ito decomposition, its applications in different fields, and its usefulness in understanding and modeling stochastic processes.

According to Lévy-Ito decomposition Lévy proccess can be represented as combination of drift, random deviations generated by Brownian motion and Compound Poisson process.

$$X_t = \gamma t + cW_t + \sum_{i=0}^{N_t} Y_t$$

,

characteristic function of which can be written as

$$\varphi(u) = exp\left[t\left(iu\gamma - \frac{(cu)^2}{2} + \int\limits_{\mathbb{R}} \left(e^{iux} - 1\right)\lambda F(dx)\right)\right]$$

The above-mentioned components of the Lévy process and their effects on function are reflected in the Lévy-Khintchine representation for one-dimensional real values Lévy processes:

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

and also

$$\int\limits_{|x|\leq 1} x^2\nu dx < \infty, \qquad \int\limits_{|x|>1} \nu dx < \infty$$

where combination (σ, ν, b) called Lévy triplet and in case of compound Poisson process $\sigma = 0$, $b = \gamma + \int \left(1 + x\mathbb{1}_{|x|\leq 1}\right)\nu(dx)$ and ν represented by finite measure called Lévy measure. In this representation drift is denoted by b and σ^2 represents diffusion of the process.

The basic idea behind the Lévy-Khintchine decomposition is that any probability distribution can be expressed as a sum of a Gaussian component and a non-Gaussian component. The Gaussian component describes the smooth, normal part of the distribution, while the non-Gaussian component describes the rough, irregular part.

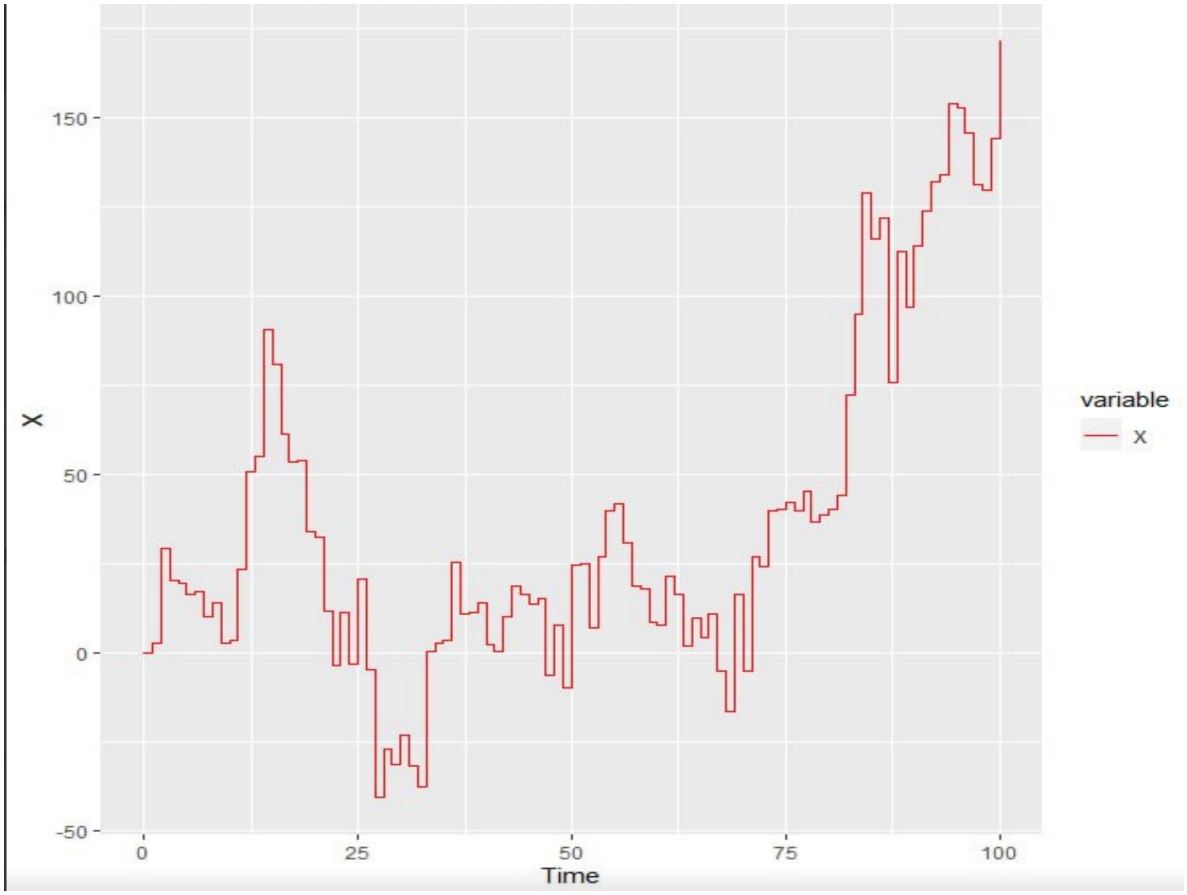
The non-Gaussian component can be further broken down into three parts: a drift term, a diffusion term, and a jump term. The drift term describes the average tendency of the process to move in a particular direction. The diffusion term describes the random fluctuations or noise in the process. The jump term describes the sudden, discontinuous changes that can occur in the process.

The intuition behind the Lévy-Khintchine decomposition is that it provides a way to understand the underlying structure of a stochastic process. By breaking the process down into its component parts, we can analyze each part separately and gain insight into how the process behaves.

2.4 Unifying the theory

A compound Poisson process is a stochastic process that is used to model the total value of a random number of independent and identically distributed random variables, each with a certain probability distribution, that are summed together.

$$X_t = \gamma t + \sum_{k=1}^{N_t} Y_k \tag{1.1}$$



Consider a compound Poisson process which is characterized by its tendency to increase or decrease, represented by the drift factor, and the intensity of an underlying counting process generating values for the compound Poisson process from an unknown distribution. To express this notion mathematically, we introduce a Poisson process N_t with intensity parameter $\lambda > 0$, a drift factor $\gamma \in \mathbb{R}$ indicating the increasing or decreasing tendency over time, and a sequence of independent and identically distributed random variables $Y_1, Y_2, \ldots, Y_n$ drawn from an unknown distribution F.

Taking into account we can present distribution F as a mixture of discrete and absolutely continuous jumps, where a discrete jump occurs with probability p and an absolutely continuous jump occurs with probability $1-p$. In this paper we assume the following definition of Poisson processes:

$$\begin{cases} X_t = \gamma t + \sum_{k=1}^{N_t} Y_k; \\ Y_k \stackrel{d}{=} \tau\eta_d + (1-\tau)\eta_{ac}. \end{cases}, \tag{1.2}$$

where τ is a Bernoulli random variable with probability of success p, and η_d and η_{ac} represent the discrete and absolutely continuous components of F, respectively.

When X_t is observed continuously, unknown parameters of the compound Poisson process λ, γ, F can be estimated using basic statistical inference techniques. However, if we can only observe X_t at discrete time points $X_\Delta, X_{2\Delta}, X_{k\Delta}$ where $k = \frac{T}{\Delta}$, more complex estimation functions are required to derive the unknown parameters. To do this, we begin by deriving the characteristic function of the discrete increment in $X_{k\Delta}$:

$$Z_k = X_{k\Delta} - X_{(k-1)\Delta} = \gamma\Delta + \sum_{j=N_{(k-1)\Delta}+1}^{N_{k\Delta}} Y_j. \tag{1.3}$$

In order to estimate the unknown parameters for discrete observations, we need to derive the characteristic function of the observed discrete increments of Z_k (1.2):

$$\varphi(u) = \mathbb{E}[e^{iuZ_t}] = e^{\Delta(iu\gamma + \mathcal{F}\nu(u) - \lambda)}. \tag{1.4}$$

In order to get any real values of the estimates we need to use the estimated values of the characteristic function. By definition, it is represented as the mathematical expectation of the characteristic exponent from the function, so we can use its estimate as the mean value of the characteristic exponent from a given process increment function.Furthermore, the characteristic function can be estimated using the sample characteristic function:

$$\hat{\varphi}_n(u) = \frac{1}{n}\sum_{k=1}^{n} e^{iuZ_k}, \tag{1.6}$$

where n is the number of observations.

Based on the characteristic function of the discrete increment in $X_{k\Delta}$, we can now construct an estimation function for the unknown parameters of the compound Poisson process, namely λ, γ, and F. We use Fourier analysis to take the inverse Fourier transform of logarithm of the characteristic function of Z_t:

$$\frac{1}{\Delta}\int_{\mathbb{R}} f(u)\mathcal{F}^{-1}[log(\varphi(u)]du =$$

$$= \frac{1}{2\pi\Delta}\int_{\mathbb{R}}\int f(u)e^{-iux}(iu\gamma + \mathcal{F}_{\nu(u)} - \lambda)dxdu =$$

$$= \gamma f'(0) - \lambda f(0) + \sum_{j\in\mathbb{Z}\setminus\{0\}} q_j f(J_j) + \int_{\mathbb{R}} f\nu_{ac}(dx), \tag{1.7}$$

where $f : \mathbb{R} \to \mathbb{R}$ is differentiable at the origin and uniformly bounded function. And Fourier and inverse Fourier transform denoted by $\mathcal{F}$ and $\mathcal{F}^{-1}$ respectively:

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

$$\mathcal{F}^{-1}f(u) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iu} f(u)du, \quad u \in \mathbb{R}.$$

This spectral analysis of the characteristic function allows us to isolate the influence of each parameter, and selecting the appropriate functions allows us to evaluate each parameter based on an empirical characteristic function.

Note that $\varphi(u)$ is the unique function that satisfies $e^{\ln(\varphi(u))} = \varphi(u)$. Although we do not know the summation terms on the right-hand side of equation (1.6), we can estimate $\varphi(\cdot)$ with $\hat{\varphi}_n(\cdot)$ (1.4) on the left-hand side and choose $f(u)$ so that it vanishes

other parts of the summation terms on the right-hand side of the equation. Using this, we can then write estimation functions for the unknown parameters:

$$\gamma = \lim_{n\to\infty} \frac{1}{\Delta} \int_{\mathbb{R}} x\mathbb{1}[|x| < \zeta_n]\mathcal{F}^{-1}(\ln(\varphi(x))dx, \tag{1.8.1}$$

$$\lambda = -\lim_{n\to\infty} \frac{1}{\Delta} \int_{\mathbb{R}} \mathbb{1}[|x| < \zeta_n]\mathcal{F}^{-1}(\ln(\varphi(x))dx, \tag{1.8.2}$$

$$q_j = \lim_{n\to\infty} \frac{1}{\Delta} \int_{\mathbb{R}} \mathbb{1}[|x - J_j| < \zeta_n]\mathcal{F}^{-1}(\ln(\varphi(x))dx, \tag{1.8.3}$$

$$N(t) = \lim_{n\to\infty} \frac{1}{\Delta} \int_{-\infty}^{t} \mathbb{1}[|x| \geq \zeta_n]]\mathcal{F}^{-1}(\ln(\varphi(x))dx. \tag{1.8.4}$$

Here, ζ_n is a sequence of positive real numbers such that $\zeta_n \to 0$ as $n \to \infty$, and $\mathbb{1}[|x| < \zeta_n]$ is the indicator function. The functions $\mathcal{F}^{-1}$ is the inverse Fourier transform. The limit operations in the above expressions mean that we take the limit as $n \to \infty$ after the integrals have been evaluated. If we can estimate these 3 parameters then We can reconstruct the entire distribution of the Levy process, through obtaining estimates of the distribution of process jumps and the weights of discrete jumps.

$$\hat{F}_n = \frac{\hat{N}_n}{\hat{\lambda}_n}, \quad \hat{p}_{j,n} = \frac{\hat{q}_{j,n}}{\hat{\lambda}_n}, \text{ and } \hat{p}_n = \frac{\hat{q}_n}{\hat{\lambda}_n}$$

3 Simulation methodology

Now let us define process of simulation CCP according to definition provided in this work in order to provide simulations

1. Generate ξ from exponential distribution unless S_t exceeds time horizon of the process T.
2. We need to model Y_i using information about τ, η_d, and η_{ac}:
   (a) Simulate τ of length N_T from a Bernoulli random variable with probability of success p.
   (b) Simulate η_d and η_{ac} to have the same length as τ, and distribute them from discrete and absolutely continuous distributions respectively.
   (c) Calculate Y_i using the formula: $\tau_i\eta_{di} + (1 - \tau_i)\eta_{aci}$.
3. Finally, we can create our process as $X_t = \gamma t + \sum_{k=1}^{N_t} Y_k$, with $X_0 = 0$.
4. We then calculate the increments of X_t as $Z_k = X_{k\Delta} - X_{(k-1)\Delta}$ for the following estimation of parameters.

Now let's provide simulation on particular example. **Step 1: Simulating samples for underlying counting process**

The simulation process starts by setting a time horizon T and selecting a time step size Δ. The Poisson process is initialized by setting the counter k to zero and the first arrival time N(0) to zero by convention.

To simulate the Poisson process, we generate random samples of the process until the cumulative time of arrivals reaches T. At each step, we generate a random sample U from an exponential distribution with intensity parameter λ, which represents the time interval until the next arrival. We add this time to the current time to obtain the next arrival time, increment the counter k, and update the arrival time N(k) accordingly. We repeat this process until the cumulative time of arrivals reaches T, at which point we have simulated a sample path of the Poisson process.

To ensure the precision of our calculations, we repeat the simulation process multiple times for each time period T. As T increases and Δ decreases, our estimations become more accurate. In this step, we simulate a vector of length T for the Poisson counting process N_t with $\lambda = 5$ as the intensity parameter. This generates a full trajectory of N_t simulated underlying renewal process.

Step 2: Modeling Y_i using information about ξ, ν_d, and ν_{ac}

Now that we have simulated the Poisson counting process, we use it to simulate a compound Poisson process. To do this, we need to model Y_t for each time point t using information about the random variables τ, ν_d, and ν_{ac}.

We begin by simulating τ, which is a sequence of zeros and ones generated from a Bernoulli random variable with a probability of success $p = 0.01$. A one indicates a discrete jump in the compound Poisson process, and a zero indicates no jump.

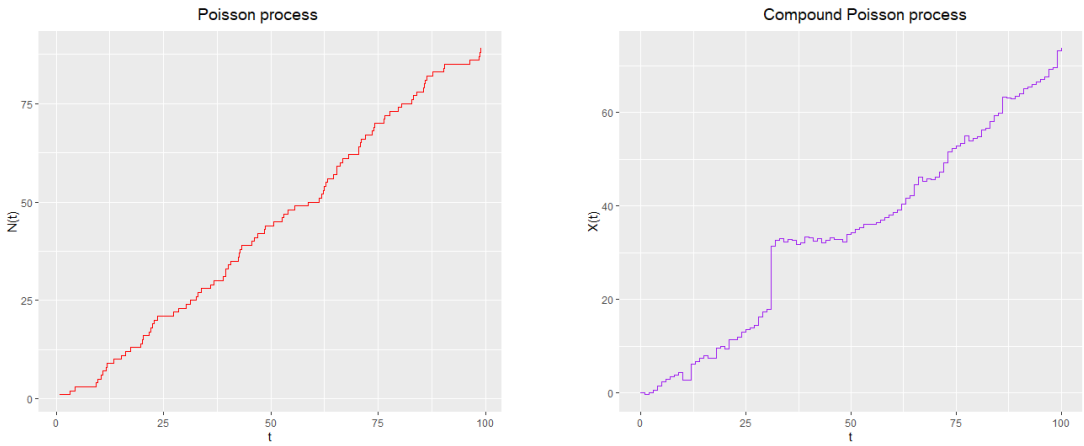
Next, we simulate ν_d and ν_{ac} to have the same length as τ. ν_d is a random variable representing the jump sizes of the compound Poisson process, while ν_{ac} is a random variable representing the continuous part of the process. We define ν_d as a sequence of independent random variables from a zero-truncated Poisson process with intensity parameter $\lambda_d = 10$. ν_{ac} is a sequence of independent random variables from a normal distribution with mean 0 and variance 1.

Finally, we calculate Y_i using the formula: $Y_{ki} \stackrel{d}{=} \tau_i \eta_{di} + (1 - \tau_i)\eta_{aci}$

This formula combines the discrete and continuous parts of the process at time point i to generate a sample of the compound Poisson process at that time point. We repeat this process for each time point i in the Poisson process to simulate a sample path of the compound Poisson process.

Step 3: Generating the compound Poisson process using previous steps

Using these simulated values of N_t and Y_i, we construct the compound Poisson process $X_t = \gamma t + \sum_{k=1}^{N_t} Y_k$, by assuming $\gamma = 0.5$, with initial value $X_0 = 0$. The increments of X_t are then calculated as $Z_k = X_{k\Delta} - X_{(k-1)\Delta}$.



To assess the accuracy of our estimations, we will repeat this simulation process $n = 100$ times for each value of $T = (100, 1000, 10000)$, and compare the results against the true values of the parameters. By gradually increasing the value of Δ and repeating the simulations, we can determine the optimal value of Δ that gives us the most accurate estimations for the unknown parameters.

4 Stochastic inference in intensity parameter λ

4.1 Estimation

To start with, we will try to make inference for λ. Taking as a reference theoretical estimation of λ (1.8.2) we can derive empirical estimation of intensity parameter of Poisson process:

$$\hat{\lambda}_n = -\frac{1}{\Delta}\int_{-\zeta_n}^{\zeta_n} \mathcal{F}^{-1} log(\hat{\varphi}_n(u))dx = -\frac{1}{\Delta}\int_{-\zeta_n}^{\zeta_n} \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iu} log(\frac{1}{n}\sum_{k=1}^{n} e^{iuZ_k})dudx$$

$$\hat{\lambda}_n = -\frac{1}{2\pi\Delta}\int_{-\zeta_n}^{\zeta_n}\int_{\mathbb{R}} e^{-iu} log(\frac{1}{n}\sum_{k=1}^{n} e^{iuZ_k})dudx,$$

where $\zeta_n \to 0$. To find $\hat{n}$ we need to take real part of this complex integral.

Now we will provide some experiment which consists of several steps. We will make our simulation more complex with the following step. We need to make sure that estimation derived in () works correctly with different combinations of parameters in compound Poisson process.

This estimation formula has calculation difficulties because of unlimited integral, in practise it has difficult landscape of the optimized function, and the optimal and correct value cannot be reached. Thus Coco [] suggest to use kernel function which is symmetric function and also satisfies following conditions:

$$\int_{\mathbb{R}} K(X)dx = 1 \text{ and } supp(\mathcal{F}K) \subseteq [-1, 1] \text{ and } |K(x)| <_\sim (1 + |x|)^{-\eta} \text{ for some } \eta > 2$$

So now we can reformulate our intensity estimator derived in (1.8.2) using Kernel function:

$$\lambda = \lim_{n\to\infty} \frac{1}{\Delta} \int_{\mathbb{R}} \mathbb{1}[|x| \geq \zeta_n] \mathcal{F}^{-1}(\ln(\varphi(x))dx,$$

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

,

Coca come up with kernel function $\mathcal{F}K(u) = \left(1-u^2\right)^2 \mathbb{1}_{[-1,1]}(u), u \in \mathbb{R}$ and it is mentioned that specification is not the most import factor, it only simplifies calculation of estimation, which coincides with opinion of Wand an Es[]. Also Coca could get specific values and dependencies of hyperparameters.

$$\eta = 5^3/2^{10} \approx 0.1221, \quad M = (\eta h_n)^{-1}, \quad \varepsilon_n = 2\pi h_n M/(M+1) \approx 2\pi h_n, \quad H_n = \varepsilon_n M,$$

where $\varepsilon_n \to 0$ and $h_n = o(\varepsilon_n)$. And $h_n \in [0.0005, 0.001]$, where upper bound used for small λ and n and lower bound in other cases.

Taking into account all provided information we can rewrite estimation of λ for practical and real computation (for more specific decomposition of this equation see eq.1-2 in Additional information):

$$\begin{aligned} \hat{\lambda}_n = &\frac{1}{2\pi\Delta} \int_{-H_n}^{-\varepsilon_n} \int_{-1}^{1} Re\left[e^{-iu} \log\left(\frac{1}{n}\sum_{k=1}^{n} e^{iuZ_k}\right)\right] \left(1-u^2\right)^2 dudx \\ &+ \frac{1}{2\pi\Delta} \int_{\varepsilon_n}^{H_n} \int_{-1}^{1} Re\left[e^{-iu} \log\left(\frac{1}{n}\sum_{k=1}^{n} e^{iuZ_k}\right)\right] \left(1-u^2\right)^2 dudx. \end{aligned} \tag{2}$$

This formula fully duplicates computational algorithm used in this research. As a first step make under-integral equation and then derive its real part multiplying result with kernel function. In the next step we calculate inner integral with limits $[-1,-1]$ and then outer integral with limits $[-H_n, -\varepsilon_n]$ for the first component of estimation and $[\varepsilon_n, H_n]$ respectively. In the following part we will try to test correctness of the provided estimation in simulated and real data.

4.2 Validation

In this part, we will determine the methods for checking the correctness of the estimates made in all subsequent work. To do this, we will try to evaluate not only the accuracy of the simulated data, but also consider whether the stated convergence of the estimates given in the original paper holds.

We will use the Mean Average Percentage Error, which is considered as a percentage deviation from the true value of the parameter, as a measure of the accuracy of the estimates.

$$MAPE(\hat{\theta}_n) = \left|\frac{\theta - \hat{\theta}_n}{\theta}\right| 100\%$$

Variance of the estimated parameters *theta* can written as for all $\theta \in \{\lambda, p\} \cup \{p_j : j \in \mathbb{Z}\backslash\{0\}\}$

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

which can estimated with

$$\hat{\sigma}_{\theta,n}^2 = \frac{1}{\Delta^2} \frac{1}{n} \sum_{k=1}^{n} \left(f^\theta \mathcal{F}^{-1} \left[\varphi^{-1}(-) \, \mathcal{F}_{h_n} \right](Z_k) \right)^2$$

$$f_n^{(\lambda)}(x) := \mathbb{1}_{\varepsilon_n \le |x| \le H_n}, \quad \varphi^{-1} = \left(\frac{1}{n} \sum_{k=1}^{n} e^{iuZ_k} \right)^{-1}$$

It is also shown that as $n \to \infty$, $\hat{\sigma}_{\theta,n}^2 \xrightarrow{\mathbb{P}} \sigma_\theta^2$ means that this estimation converges in probabilistic sense to real value with increasing number of observation, consequently we can you estimation of variance instead.

Step 0. Before verifying the accuracy of the estimate, make sure that the estimates have the stated convergence in the paper. To do this, we will start checking with the simplest case, when there is no drift and discrete jumps in the model, then we will gradually complicate our model using more parameters and more complex distributions.

Proposition 4.1. Estimation of lambda has the following convergence property:

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda \right) \xrightarrow{d} \mathcal{N}(0, \sigma_\lambda^2),$$

where

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

[* ,] For λ we can write its estimation of variance with formula derived before in the beginning of the section:

$$\hat{\sigma}_{\lambda,n}^2 = \frac{1}{\Delta^2} \frac{1}{n} \sum_{k=1}^{n} \left(\mathbb{1}_{[-H_n;H_n]\backslash(-\epsilon_n,\epsilon_n)} \mathcal{F}^{-1} \left[\varphi^{-1}(-) \, \mathcal{F}_{h_n} \right](Z_k) \right)^2$$

Let us start with the simpliest case of simulated compound poisson process

Step 1. In the first step we will exclude almost all parameters in the model. Assume that $\tau \sim Bern(p = 0)$, $\gamma = 0$. That means that only absolutely continuous jump present in the model. We expect that in this simple case we can achieve high precision of our estimation.

For the begging let us test quality of the provided estimation by taking default symmetric distribution of absolutely continuous jumps. In the next step we implement this estimation also for distributions with heavy tails.

Let us construct the experiment described in the previous paragraph. For this particular case assume $\nu_{ac} \sim \mathbb{N}(0, 1)$ counting process has intensity $\lambda = 3$ and take recommended bandwidth for kernel function $h_n = 0.0005$

Figure 1: Illustration of proposition 4.1 for no drift and no discrete jump process with $\lambda = 3$. $n = 1000$

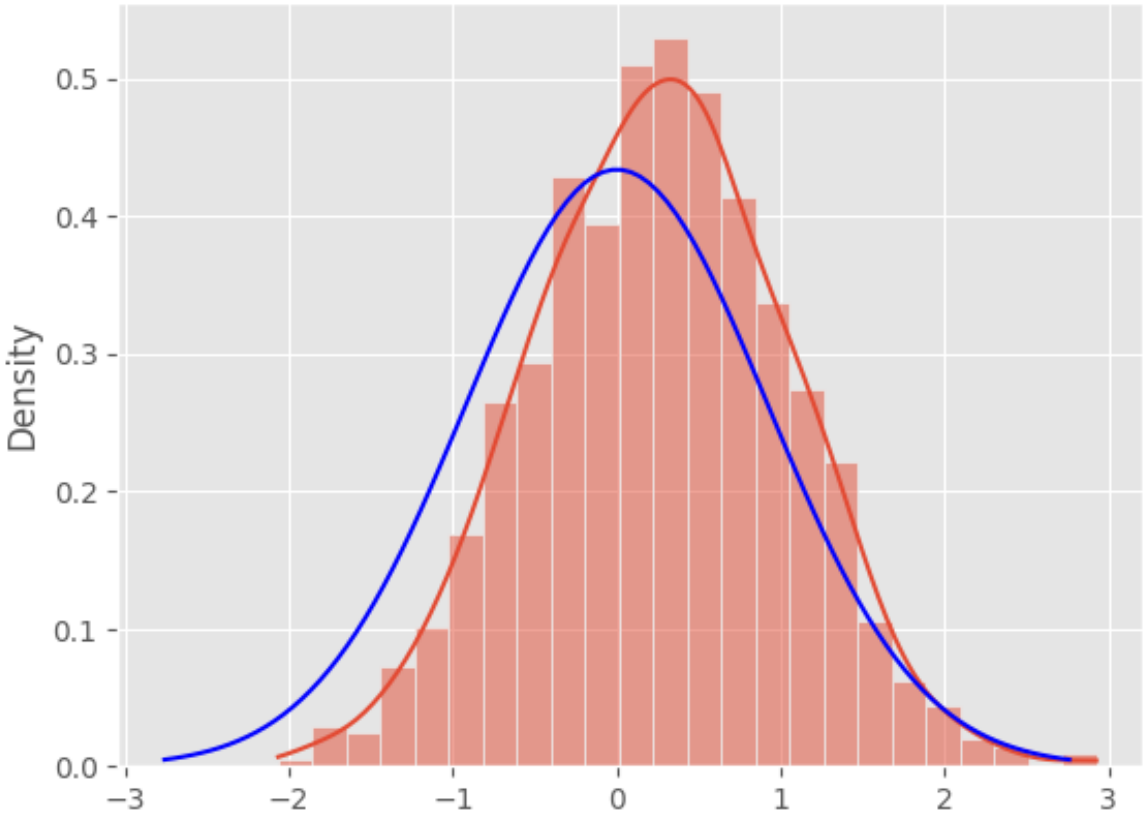
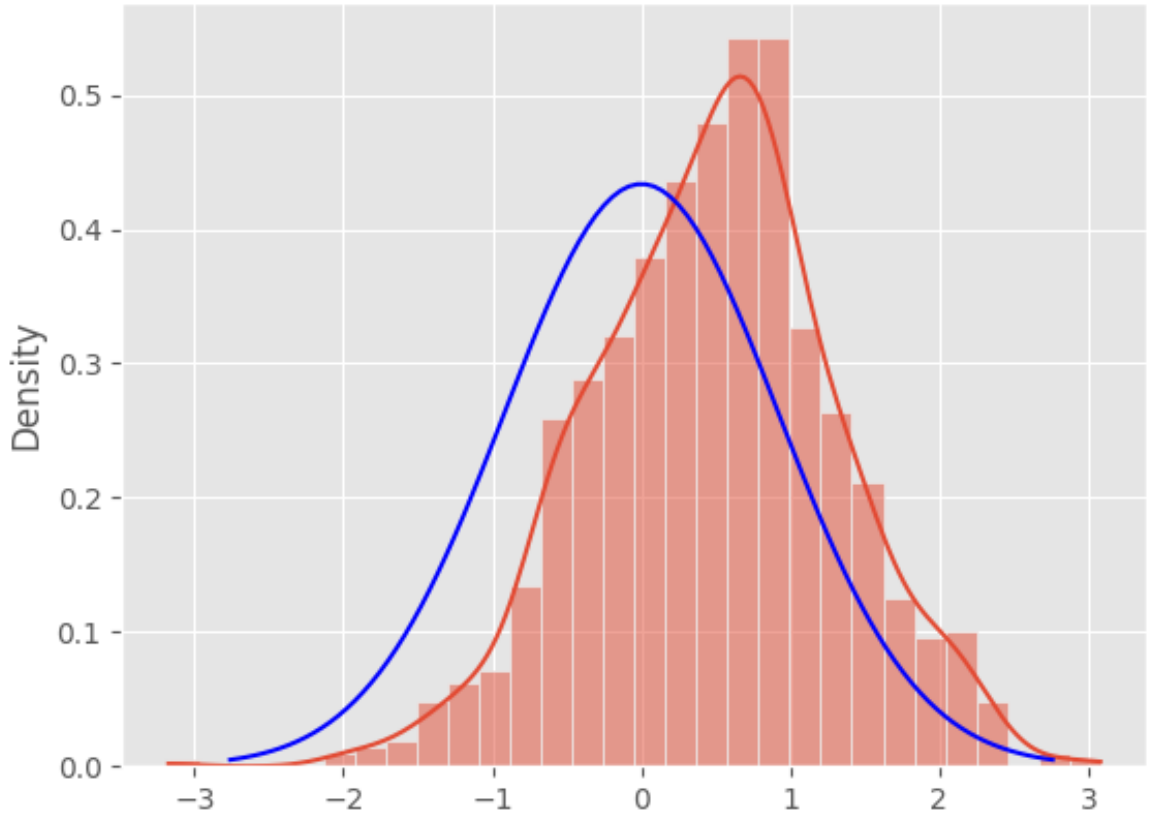
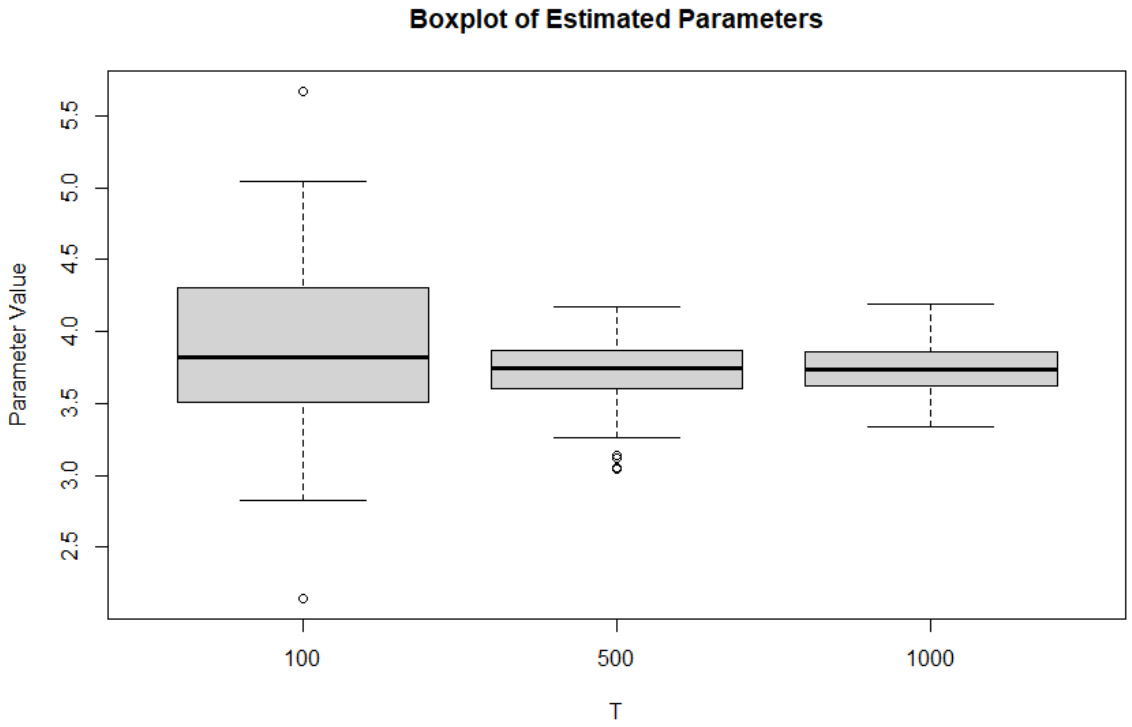


Figure 2: Illustration of proposition 4.1 for no discrete jump process with $\lambda = 3$ and γ. $n = 1000$





As can be seen from graph, as the time grid increases, there is a convergence to the true value of the lambda parameter, but this convergence may be limited by the value of the bandwidth hyperparameter h_n. As for the forward step we can try to specify such function h_n which satisfies $h_n \to 0$ as $n \to \infty$ and minimizes error of the estimated parameters in terms of MSE.

Now let's evaluate how our estimate works for different intensity parameters. We will also use the same hyperparameters, but will only change the lambda parameter. Here we need to focus on how the put forward estimator will handle different intensity levels. To do this, we will use the methodology described earlier in []. We will include different levels of λ, specifically 0.5, 1, 2.5, 5 and 10

The resulting simulations generally show that as the simulated time interval increases, the estimate gets better on average, and is more likely to arrive at the true value. However, in low-intensity processes, there is a situation in which the variance of the estimate decreases, but on average the resulting estimate value begins to deviate as the number of observations increases, in contrast to medium-intensity and high-intensity processes. This may be due to the fact that low-intensity compound Poisson processes are more sensitive to the choice of the bandwidth parameter, which leads to more significant errors with increasing number of observations

Table 1 confirms the assumptions made above. On average, the quality metric for simulating trajectories of 1000 points has slightly worse quality in terms of MAPE compared to trajectories of 500 points in length, especially noticeable deterioration for stochastic processes with low intensity in the counting process. Also, the result of the experiment performed, shows that a smaller absolute error relative to the true value of the lambda parameter is achieved when estimating stochastic processes of high intensity. In the case of increasing the time horizon for these processes, the estimation has the strongest increase in quality, in other words, when estimating processes with

Figure 3: Absolute error for simulated lambda in case of different intensity

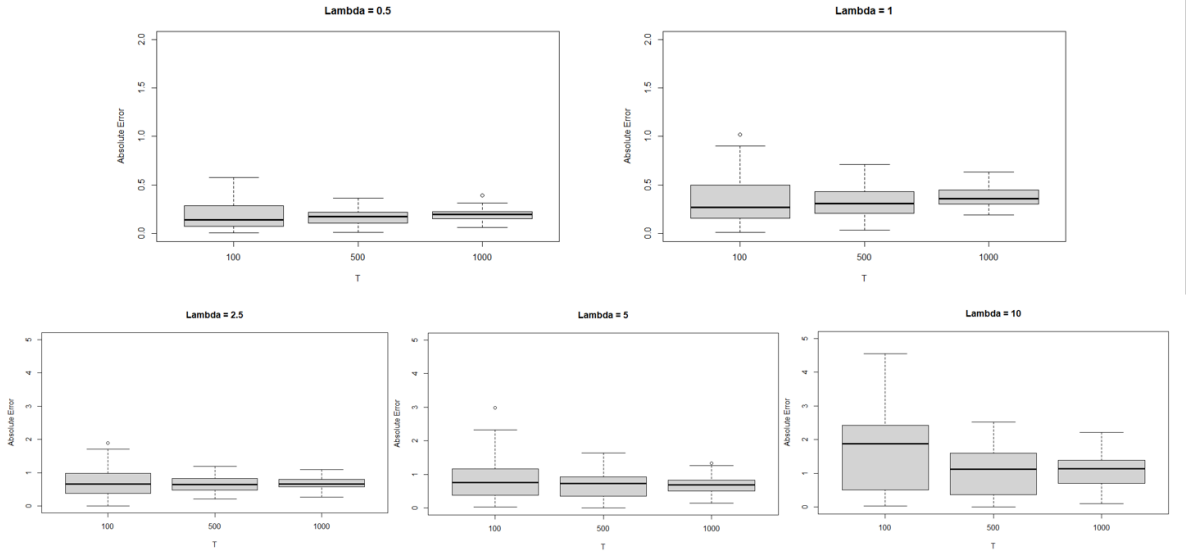


Table 1: MAPE for estimated lambda in case of different intensity

lambda	100	500	1000
0.5	37.039	35.766	39.382
1	33.758	32.137	37.120
2.5	29.106	27.306	27.895
5	17.528	13.530	13.723
10	16.362	10.330	10.618

high intensity, the number of observations is more significant relative to processes with lower intensity. However, low-intensity processes are more sensitive to the choice of the bandwidth parameter.

Step 2. Here we will include all the parameters of the model

Step 3. Here we will try to estimate λ with different settings of compound Poisson process

5 Bibligraphy

Sato, K.-I.: *Lévy Processes and Infinitely Divisible Distribution.* Cambridge University Press, 1999.

Coca, A. J.: *Efficient nonparametric inference for discretely observed compound Poisson processes.* PhD thesis, University of Cambridge. https://albertococacabrero.wordpress.com/research/ (2016). Accessed 27 Jan 2017

M. Wand. *Finite sample performance of deconvolving density estimators*x. Statistics Probability Letters, 37(2):131–139, 1998. (Cited on page 117).

B. van Es, S. Gugushvili, and P. Spreij. *A kernel type nonparametric density estimator*

for decompounding. Bernoulli, 13(3):672–694, 2007. Nickl, R., Reiß, M.: A Donsker theorem for Lévy measures. J. Funct. Anal. 263, 3306–3332 (2012)
B. Buchmann. Weighted empirical processes in the nonparametric inference for L´evy processes. Mathematical Methods of Statistics, 18(4):281–309, 2009.
M. D. Donsker. Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. The Annals of Mathematical Statistics, 23(2):277–281, 1952.
J. E. Yukich. Weak convergence of the empirical characteristic function. Proceedings of the American Mathematical Society, 95(3):470–473, 1985.
S. Gugushvili, F. van der Meulen, and P. Spreij. Nonparametric Bayesian inference for multidimensional compound Poisson processes. Modern Stochastics: Theory and Applications
F. Comte and V. Genon-Catalot. Nonparametric adaptive estimation for pure jump L´evy processes. Annales de l'institut Henri Poincar´e (B), 46(3):595–617, 2010a

6 Additional Information

1. Decomposition of λ into real and complex integrals

$$\hat{\lambda}_n = \frac{1}{\Delta 2\pi}\int_{-T}^{T}\int (cos(ux) - i(sin(ux))Ln(\frac{1}{n}\sum_{k=1}^{n} e^{iuZ_k})dudx =$$
$$= \frac{1}{\Delta 2\pi}(\int_{-T}^{T}\int (cos(ux) - i(sin(ux))[\sqrt{(\sum cos(uZ_k))^2 + (\sum sin(uZ_k))^2}+$$
$$+ i * arctg\frac{\sum sin(uZ_k)}{\sum cos(uZ_k)} - log(n)]dudx =$$
$$= \frac{1}{\Delta 2\pi}\int_{-T}^{T}\int [cos(ux)\sqrt{(\sum cos(uZ_k))^2 + (\sum sin(uZ_k))^2} + sin(ux)arctg\frac{\sum sin(uZ_k)}{\sum cos(uZ_k)}$$
$$- cos(ux)log(n)]dudx + +i\frac{1}{\Delta 2\pi}\int_{-T}^{T}\int [cos(ux)arctg\frac{\sum sin(uZ_k)}{\sum cos(uZ_k)} -$$
$$- sin(ux)\sqrt{(\sum cos(uZ_k))^2 + (\sum sin(uZ_k))^2} + sin(ux)log(n)]dudx \tag{3}$$

2. More simple version of decomposition of λ into real and complex integrals if we change integrations limits

$$\hat{\lambda}_n = \frac{1}{\Delta 2\pi}\int \frac{2sin(Tu)}{u}[\sqrt{(\sum cos(uZ_k))^2 + (\sum sin(uZ_k))^2} + i * arctg\frac{\sum sin(uZ_k)}{\sum cos(uZ_k)} - log(n)]du$$
$$= \frac{1}{\Delta 2\pi}\int \frac{2sin(Tu)}{u}[\sqrt{(\sum cos(uZ_k))^2 + (\sum sin(uZ_k))^2} - log(n)]du+$$
$$+ i * \int \frac{2sin(Tu)}{u} * arctg\frac{\sum sin(uZ_k)}{\sum cos(uZ_k)}du \tag{4}$$