Statistical Inference of Discretely Observed Compound Poisson Processes and Related Jump Processes

Suraj Shah

March 19, 2019

Abstract

Contents

1 Introduction **3**

2 Spectral Approach **4**
2.1 Van Es 4
2.1.1 Construction of Density Estimator via suitable inversion of characteristic functions 4
2.1.2 Kernel density estimators 7
2.1.3 Simulation Results 9

1 Introduction

Definition 1.1 (Counting Process)**.** A *counting process* is a stochastic process $\{N(t) : t \geq 0\}$ with values that are non-negative, integer and non-decreasing i.e. $\forall s, t \geq 0 : s \leq t$:

1. $N(t) \geq 0$,
2. $N(t) \in \mathbb{N}$,
3. $N(s) \leq N(t)$.

Definition 1.2 (Poisson Process)**.** A *Poisson process with intensity* λ is a counting process $\{N(t) : t \geq 0\}$ with the following properties:

1. $N(0) = 0$,
2. It has independent increments i.e. $\forall n \in \mathbb{N} : 0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, $N(t_n) - N(t_{n-1}), N(t_{n-1}) - N(t_{n-2}), \ldots, N(t_1)$ are independent,
3. The number of occurrences in any interval of length t is a Poisson random variable with parameter λt i.e. $\forall s, t : s \leq t$, $N(t) - N(s) \sim \text{Poisson}(\lambda(t - s))$.

Lemma 1.1. *A Poisson process with intensity λ has exponentially distributed inter-arrival times with rate λ.*

Definition 1.3 (Compound Poisson Process)**.** Let $N(t) : t \geq 0$ be a d-dimensional Poisson process with intensity λ.

Let $Y_1, Y_2, \ldots$ be a sequence of i.i.d random variables taking values in $\mathbb{R}^d$ with common distribution F.

Also assume that the Y_i's are independent of the Poisson process $\{N(t) : t \geq 0\}$.

Then, a *Compound Poisson process (CPP)* is a stochastic process $\{X(t) : t \geq 0\}$ such that

$$X(t) = \sum_{i=1}^{N(t)} Y_i$$

where, by convention, we take $X(t) = 0$ if $N(t) = 0$.

Suppose we take discrete observations of a CPP i.e. we consider $X(\Delta), X(2\Delta), \ldots$ where $X(t) : t \geq 0$ is a CPP. We want to estimate F. Note that the jump size $X(n\Delta) - X((n-1)\Delta)$ is equivalent in distribution to a Poisson random sum of intensity Δ:

$$\begin{aligned} X(n\Delta) - X((n-1)\Delta) = \sum_{i=1}^{N(n\Delta)} Y_i - \sum_{i=1}^{N((n-1)\Delta)} Y_i \\ = \sum_{i=1}^{N(n\Delta)-N((n-1)\Delta)} Y_i \\ =^d \sum_{i=1}^{N} Y_i \end{aligned}$$

where $N \sim \text{Poisson}(\Delta)$

2 Spectral Approach

Now we have formulated the problem, we visit some methods for estimating the unknown density f. Since adding a Poisson number of Y's is referred to as compounding, much of the literature refers to the problem of recovering density f of Y's from observations of X as decompounding.

The approach of decompounding was famously proposed by Buchmann and Grübel to estimate the density f for discrete and continuous cases of the distribution F of the Y's.

Van Es built on this idea for fixed sampling rate $\Delta = 1$ using the Lévy - Khintchine formula. We explain the idea behind this method and show its strength through various examples.

2.1 Van Es

2.1.1 Construction of Density Estimator via suitable inversion of characteristic functions

We first note the following property:

Proposition 2.1. *For Poisson random sum X, the characteristic function of X, denoted by ϕ_X, is given by $\phi_X(t) = \mathbb{E}e^{itX} = e^{-\lambda+\lambda\phi_f(t)}$*

Proof.

$$\begin{aligned}
\phi_X(t) &= \mathbb{E}e^{itX} \\
&= \mathbb{E}\left[\exp\left(it\sum_{i=1}^{N(\lambda)} Y_i\right)\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{N(\lambda)} \exp(itY_i)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{N(\lambda)} \exp(itY_i)\middle| N(\lambda)\right]\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{N(\lambda)} \mathbb{E}\left[\exp(itY_1) \mid N(\lambda)\right]\right] && \text{(by i.i.d assumption of the } Y_i\text{'s)} \\
&= \mathbb{E}\left[\prod_{i=1}^{N(\lambda)} \phi_f(t)\right] && (Y_1 \text{ and } N(\lambda) \text{ are independent)} \\
&= \mathbb{E}\left[\exp(N(\lambda)\ln\phi_f(t))\right] \\
&= \exp(\lambda(e^{\ln\phi_f(t)} - 1)) && \text{(MGF of a Poisson random variable)} \\
&= e^{-\lambda+\lambda\phi_f(t)}
\end{aligned}$$

□

We can rewrite $\phi_X(t)$ as:

$$\begin{aligned}
\phi_X(t) &= e^{-\lambda}(e^{\lambda\phi_f(t)} - 1 + 1) \\
&= e^{-\lambda} + e^{-\lambda}(e^{\lambda\phi_f(t)} - 1) \\
&= e^{-\lambda} + e^{-\lambda}\frac{e^{\lambda} - 1}{e^{\lambda} - 1}(e^{\lambda\phi_f(t)} - 1) \\
&= e^{-\lambda} + \frac{1 - e^{-\lambda}}{e^{\lambda} - 1}(e^{\lambda\phi_f(t)} - 1) \qquad (1)
\end{aligned}$$

Since a zero jump size provides no additional information on the density f, we want to gain information about X conditional on the event that there is at least one jump. Seeing that $X \mid N(\lambda) > 0$ has a density is somewhat intuitive, but we provide a proof of this.

Lemma 2.1. *The random variable $X \mid N(\lambda) > 0$ has a density.*

Proof. By the Radon-Nikodym Theorem, a random variable X has a density if and only if $\mathbb{P}(X \in A) = 0$ for every Borel set A with Lebesgue measure zero.

Suppose that $\mathrm{Leb}(A) = 0$. Then

$$\begin{aligned}\mathbb{P}(X \in A | N(\lambda) > 0) &= \frac{1}{\mathbb{P}(N(\lambda) > 0)} \sum_{n=1}^{\infty} \mathbb{P}\left(Y_1 + \cdots + Y_n \in A, N(\lambda) = n\right) \\ &= \frac{1}{\mathbb{P}(N(\lambda) > 0)} \sum_{n=1}^{\infty} \mathbb{P}\left(Y_1 + \cdots + Y_n \in A\right) \mathbb{P}(N(\lambda) = n)\end{aligned}$$

Note that for each $n, Y_1 + \cdots + Y_n$ has a density so $\mathbb{P}\left(Y_1 + \cdots + Y_n \in A\right) = 0$. Thus the result follows.

□

Let g be the density of $X \mid N(\lambda) > 0$.
Let $\phi_g(t) = \mathbb{E}\left[e^{itX} \mid N(\lambda) > 0\right] = \frac{\mathbb{E}\left[e^{itX} \mathbb{1}(N(\lambda)>0)\right]}{\mathbb{P}(N(\lambda)>0)}$.
Then

$$\begin{aligned}\phi_X(t) &= \mathbb{E}\left[e^{itX} \mathbb{1}(N(\lambda) = 0)\right] + \mathbb{E}\left[e^{itX} \mathbb{1}(N(\lambda) > 0)\right] \\ &= \mathbb{P}(N(\lambda) = 0) + \mathbb{P}(N(\lambda) > 0) \phi_g(t) \\ &= e^{-\lambda} + (1 - e^{-\lambda}) \phi_g(t)\end{aligned}$$

Therefore, using (1), we get that

$$\phi_g(t) = \frac{1}{e^{\lambda} - 1} (e^{\lambda \phi_f(t)} - 1) \tag{2}$$

Thus, we can see from this that if we were to obtain an estimator for $\phi_g(t)$, then by suitable inversion of the formula in (2), we would obtain an estimator for $\phi_f(t)$.

In order to rewrite (2) in terms of $\phi_f(t)$, we must be able to invert the complex exponential function since $\phi_f(t)$ takes complex values. However, such function is not invertible since it is not bijective: in particular it is not injective as $e^{w+2\pi i} = e^w \ \forall w \in \mathbb{C}$.

Therefore, we use the following lemmas concerning the distinguished logarithm:

Lemma 2.2. *If $h_1 : \mathbb{R} \to \mathbb{C}$ and $h_2 : \mathbb{R} \to \mathbb{C}$ are continuous functions such that $h_1(0) = h_2(0) = 0$ and $e^{h_1} = e^{h_2}$, then $h_1 = h_2$.*

Proof. See Appendix. □

Lemma 2.3. *If $\phi : \mathbb{R} \to \mathbb{C}$ is a continuous function such that $\phi(0) = 1$ and $\phi_g(t) \neq 0 \ \forall t \in \mathbb{R}$ then there exists a unique continuous function $h : \mathbb{R} \to \mathbb{C}$ with $h(0) = 0$ and $\phi(t) = e^{h(t)}$ for $t \in \mathbb{R}$.*

Proof. See Appendix. □

Therefore, for such a function ϕ as described in the Lemma, we say that the unique function h is the distinguished logarithm and we denote $h(t) = \mathrm{Log}(\phi(t))$. Note also that for ϕ and ψ satisfying the assumptions of the Lemma, we have $\mathrm{Log}(\phi(t)\psi(t)) = \mathrm{Log}(\phi(t)) + \mathrm{Log}(\psi(t))$ as expected.

Therefore, noting that $\phi(t) = e^{\lambda(\phi_f(t)-1)}$ is a continuous function satisfying $\phi(0) = 1$ and $\phi(t) \neq 0 \ \forall t \in R$, we get that

$$\begin{aligned}
\lambda(\phi_f(t) - 1) &= \mathrm{Log}\left(e^{\lambda(\phi_f(t)-1)}\right) && \text{(Lemma 2.2)} \\
&= \mathrm{Log}\left(e^{-\lambda}\left[(e^{\lambda} - 1)\phi_g(t) + 1\right]\right) \\
&= -\lambda + \mathrm{Log}\left((e^{\lambda} - 1)\phi_g(t) + 1\right)
\end{aligned}$$

Therefore,

$$\phi_f(t) = \frac{1}{\lambda}\mathrm{Log}\left((e^{\lambda} - 1)\phi_g(t) + 1\right) \tag{3}$$

By Fourier inversion, for integrable ϕ_f we have

$$f(x) = \frac{1}{2\pi\lambda}\int_{-\infty}^{\infty} e^{-itx}\mathrm{Log}\left((e^{\lambda} - 1)\phi_g(t) + 1\right)dt \tag{4}$$

This suggests that if we can estimate $\phi_g(t)$, then we have an estimate of f.

2.1.2 Kernel density estimators

We provide the intuition behind choosing our estimator for g on observations of non-zero jump size as a kernel density estimator.

Let X be a random variable with probability density p with respect to the Lebesgue measure on $\mathbb{R}$. The corresponding distribution function is $F(x) = \int_{-\infty}^{x} p(t)dt$.

Consider n i.i.d observations $X_1, \ldots, X_n$ with same distribution as X. The empirical distribution function is given by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$$

By the Strong Law of Large Numbers, since for fixed x, $I(X_i \leq x)$ are i.i.d , we have that

$$F_n(x) \to \mathbb{E}\left[I(X_1 \leq x)\right] = \mathbb{P}(X \leq x) = F(x)$$

almost surely as $n \to \infty$.

Therefore, $F_n(x)$ is a consistent estimator of $F(x)$ for every $x \in \mathbb{R}$. Also note that $p(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x)$, so for sufficiently small $h > 0$ we can write an approximation

$$p(x) \approx \frac{F(x+h) - F(x-h)}{2h}$$

Thus, intuitively we can replace F by our empirical distribution function F_n to give us an estimator $\hat{p}_n(x)$ of $p(x)$

$$\begin{aligned}\hat{p}_n(x) &= \frac{F_n(x+h) - F_n(x-h)}{2h} \\ &= \frac{1}{2nh}\sum_{i=1}^{n} I(x-h < X_i \leq x+h) \\ &= \frac{1}{nh}\sum_{i=1}^{n} K_0\left(\frac{x-X_i}{h}\right)\end{aligned}$$

where $K_0(u) = \frac{1}{2}I(-1 < u \leq 1)$.

A simple generalisation is to replace K_0 by some arbitrary (but well-chosen) integrable function $K : \mathbb{R} \to \mathbb{R}$ such that $\int K(u)du = 1$ and $K(u) = K(-u)$ for every $u \in \mathbb{R}$. Such a function K is called a *kernel* and the parameter h is called a *bandwidth* of the estimator

$$\hat{p}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \tag{5}$$

We call this estimator a *kernel density estimator*.

Thus, for some kernel w with characteristic function ϕ_w and observations $Z_1, \ldots, Z_n$, we estimate density g by the kernel density estimator

$$g_{nh}(x) = \frac{1}{nh}\sum_{i=1}^{n} w\left(\frac{x-Z_i}{h}\right)$$

Letting $\phi_{\text{emp}}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itZ_j}$ be the empirical characteristic function, we get that

$$\begin{aligned}\phi_{g_{nh}}(t) &= \int_{-\infty}^{\infty} e^{itx} g_{nh}(x)dx \\ &= \int_{-\infty}^{\infty} e^{itx}\frac{1}{nh}\sum_{j=1}^{n} w\left(\frac{x-Z_j}{h}\right)dx \\ &= \frac{1}{n}\sum_{j=1}^{n} e^{itZ_j}\int_{-\infty}^{\infty} e^{ithy}w(y)dy \qquad \left(\text{by the substitution } y = \frac{x-Z_j}{h}\right) \\ &= \phi_{\text{emp}}(t)\phi_w(ht)\end{aligned}$$

In view of (4) It is tempting to introduce an estimator $\hat{f}_{nh}$ of f

$$\hat{f}_{nh}(x) = \frac{1}{2\pi\lambda}\int_{-\infty}^{\infty} e^{-itx}\text{Log}\Big((e^{\lambda}-1)\phi_{\text{emp}}(t)\phi_w(ht) + 1\Big)dt \tag{6}$$

but this brings two main issues:

1. In light of Lemma 2.3, we may have some Borel set A with non-zero Lebesgue measure such that $(e^{\lambda}-1)\phi_{\text{emp}}(t)\phi_w(ht)+1$ is zero for $t \in A$. The distinguished logarithm is undefined under such sets and thus our estimator of f is undefined in this case.

2. There is no guarantee that the integral is finite. For example,

$$\phi_{g_{nh}}(t) = \frac{\exp(e^{it}) - 1}{e^{\lambda} - 1}$$

would give $\hat{f}_{nh}(1)$ to be infinity.

In order to prove asymptotic properties, we must adjust our estimators by bounding $\hat{f}_{nh}$ for each n using a suitable sequence $(M_n)_{n\geq 1}$. However, for our discussion, we note such limitations and provide simulations for examples where these two cases do not occur.

2.1.3 Simulation Results

We note that for $\lambda < \log 2$, the distinguished logarithm in (6) reduces to the principal branch of the logarithm. This is the logarithm whose imaginary part lies in the interval $(-\pi, \pi]$. We also note, as written above, that bounding $\hat{f}_{nh}$ by a suitable sequence is not needed in practice. Therefore, we can use (6) to compute our estimator with the principal branch of the logarithm, provided $\lambda < \log 2$.

We use the following kernel w given by

$$w(t) = \frac{48t(t^2-1)\cos t - 144(2t^2-5)\sin t}{\pi t^7}$$

This kernel has a fairly complicated form but its characteristic function $\phi_w(t)$ has a much simpler expression given by

$$\phi_w(t) = (1-t^2)^3 \mathbb{1}\{|t| < 1\}$$

We can rewrite (6) as $\hat{f}_{nh}(x) = \hat{f}^{(1)}_{nh}(x) + \hat{f}^{(2)}_{nh}(x)$ where

$$\hat{f}^{(1)}_{nh}(x) = \frac{1}{2\pi\lambda}\int_0^{\infty} e^{-itx}\text{Log}\Big((e^{\lambda}-1)\phi_{\text{emp}}(t)\phi_w(ht)+1\Big)dt \tag{7}$$

$$\begin{aligned}\hat{f}^{(2)}_{nh}(x) &= \frac{1}{2\pi\lambda}\int_{-\infty}^{0} e^{-itx}\text{Log}\Big((e^{\lambda}-1)\phi_{\text{emp}}(t)\phi_w(ht)+1\Big)dt \\ &= \frac{1}{2\pi\lambda}\int_0^{\infty} e^{itx}\text{Log}\Big((e^{\lambda}-1)\phi_{\text{emp}}(-t)\phi_w(ht)+1\Big)dt \end{aligned} \tag{8}$$

since ϕ_w is symmetric. We use a bandwidth of 0.14. Such a bandwidth is arbitrary and we may use better methods to compute a bandwidth estimator that yields better results. This can be done via cross-validation.

We approximate (7) and (8) by the Trapezoid Rule:

Trapezoid Rule. *Let* $\{t_j\}_{j=0}^{N-1}$ *be a set of* N *equally spaced values partitioning* $[a,b]$*, with spacing* $\Delta t_k = \Delta t = \frac{b-a}{N}$*. Then, for integrable function* f *we get the following approximation*

$$\int_a^b f(x)dx \approx \Delta t \left(\frac{f(t_0) + f(t_{N-1})}{2} + \sum_{j=1}^{N-2} f(t_j) \right) \tag{9}$$

We can approximate (7) (and similarly (8)) by computing the integrand from 0 to some sufficiently large M.

We may also allow for (9) to be written as a 'nice' sum in order to compute the Fast Fourier Transform. Thus, we write

$$\int_a^b f(t)dt \approx \Delta t \left(\sum_{j=0}^{N-1} f(t_j) \right) \tag{10}$$

Applying this to (7) and (8) we get for $t_j = j\eta$ for some spacing parameter η

$$\hat{f}_{nh}^{(1)}(x) \approx \frac{\eta}{2\pi\lambda} \sum_{k=0}^{N-1} e^{-it_j x} \text{Log}\Big((e^{\lambda} - 1)\phi_{\text{emp}}(t_j)\phi_w(ht_j) + 1 \Big), \tag{11}$$

$$\hat{f}_{nh}^{(2)}(x) \approx \frac{\eta}{2\pi\lambda} \sum_{k=0}^{N-1} e^{it_j x} \text{Log}\Big((e^{\lambda} - 1)\phi_{\text{emp}}(-t_j)\phi_w(ht_j) + 1 \Big), \tag{12}$$

We apply the Fast Fourier Transform to evaluate our functions $\hat{f}_{nh}^{(1)}$ and $\hat{f}_{nh}^{(2)}$ at points $\{x_k\}_{k=0}^{N-1}$.

Fast Fourier Transform. *Let* $\{x_k\}_{k=0}^{N-1}$ *be a sequence of complex numbers. The Fast Fourier Transform computes the sequence* $\{Y_j\}_{j=0}^{N-1}$ *where*

$$Y_j = \sum_{k=0}^{N-1} x_k e^{-ij\frac{2\pi k}{N}} \tag{13}$$

The inverse transform is given by

$$Y_j = \frac{1}{N} \sum_{k=0}^{N-1} x_k e^{ij\frac{2\pi k}{N}} \tag{14}$$

Thus, we employ a regular spacing with parameter δ so that our values $\{x_k\}_{k=0}^{N-1}$ evenly spaced and given by

$$x_k = \frac{-N\delta}{2} + \delta k$$

Thus we have

$$\hat{f}_{nh}^{(1)}(x_k) \approx \frac{1}{2\pi\lambda}\sum_{k=0}^{N-1} e^{-ijk\eta\delta} e^{it_j\frac{N\delta}{2}} \psi^{(1)}(t_j)\eta, \tag{15}$$

$$\hat{f}_{nh}^{(2)}(x_k) \approx \frac{1}{2\pi\lambda}\sum_{k=0}^{N-1} e^{ijk\eta\delta} e^{-it_j\frac{N\delta}{2}} \psi^{(2)}(t_j)\eta, \tag{16}$$

Therefore, we take $\eta\delta = \frac{2\pi}{N}$ and we apply FFT on the sequence $\left\{e^{it_j\frac{N\delta}{2}}\psi^{(1)}(t_j)\eta\right\}_{j=0}^{N-1}$ to get values for $\hat{f}_{nh}^{(1)}$ and we apply IFFT on the sequence $\left\{e^{-it_j\frac{N\delta}{2}}\psi^{(2)}(t_j)\eta\right\}_{j=0}^{N-1}$ to get values for $\hat{f}_{nh}^{(2)}$.

We take N to be a power of 2 for computational speed up in calculating the Discrete Fourier Transforms and we choose η relatively small so that δ can be relatively larger and so points are relatively separate from one another.

The results for $N = 16384$ and $\eta = 0.01$ based on 1000 observations can be shown in Figure 1.

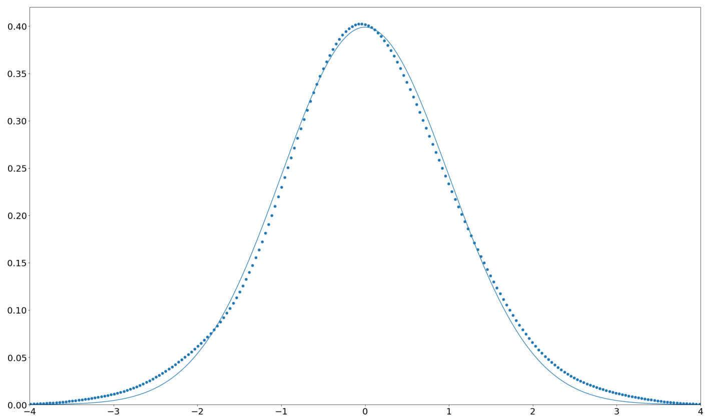

Figure 1: Density Estimator of Standard Normal

The second example we consider is the case of f being a mixture of two normal densities

$$f(\cdot) = \rho_1\psi(\cdot;\mu_1,\sigma_1^2) + \rho_2\psi(\cdot;\mu_2,\sigma_2^2)$$

where $\rho_1 = \frac{2}{3}, \rho_2 = \frac{1}{3}, \mu_1 = 0, \mu_2 = 3, \sigma_1^2 = 1, \sigma_2^2 = \frac{1}{9}$. We use a bandwidth of 0.1.