Parameter Estimation for Infinite Lévy Processes via Sequential Monte Carlo

Liew Kuang Chen Joel

February 25, 2016

Acknowledgements

Abstract

Contents

1 Introduction

2 Preliminaries

2.1 The Lévy Process

In this paper, we will be using the Variance Gamma model from the family of Infinite Activity Lévy Processes. For the uninitiated, we will first go through the basics.

Definition 2.1. Lévy Process

A Lévy Process, X_t, is a cadlag stochastic process on $(\Omega, \mathcal{F}_t, P)$,with $X_0 = 0$ and has the following properties:

1. Independent increments: $X_{t+j+1} - X_{t+j} \perp X_{t+k+1} - X_{t+k}, \forall k, j \geq 0, k \neq j$

2. Stationary increments: The Law of $X_{t+h} - X_t$ does not change for any t.

3. Stochastic continuity: For any $\epsilon > 0, \lim_{h \to 0} P(|X_{t+h} - X_t| > \epsilon) = 0$

The Lévy Process also satisfies the Infinite Divisibility property.

Definition 2.2. Infinite Divisibility

For $n \geq 2$, a probability distribution F is infinitely divisibile if there exists n i.i.d. random variables $Y_1, Y_2, ...$ such that $\sum_{i=1}^{n} Y_i$ has distribution F.

One common distribution satisfying the infinite divisibility condition is the Normal Distribution. If $X \sim N(\mu, \sigma^2)$, then $X = \sum_{i=1}^{n} Y_i$,where $Y_i \sim N(\frac{\mu}{n}, \frac{\sigma^2}{n})$.

Due to fact that Lévy Processes are semimartingales and that their increments are independent, the process is Markovian and therefore in line with the Efficient Market Hypothesis. Hence, such processes are suitable to model asset prices (Li 2006).

2.1.1 Infinite Activity Processes - Variance Gamma

The probability densities of most Lévy processese are known in closed forms. However, their characteristics functions $\phi_{X_t}(u)$ is as follows:

$$\phi_{X_t}(u) = E[\exp^{iuX_t}] = \exp^{-t\psi_x(u)}, t \geq 0$$

where $\psi_x(u)$ is the characteristic exponent of X (Tankov) and satisfies the follow Lévy-Khintchine formula

$$\psi_x \equiv -i\mu u + \frac{\sigma^2 u^2}{2} + \int_{R_0}(1 - \exp^{iux} + iux1_{|x|<1})\pi(dx)$$
$$\text{where } R_0 = R \setminus \{0\} \text{ and } \pi \text{ is a } R_0 \text{ Radon measure and}$$
$$\int_{R_0} |x|^2\pi(dx) < \infty \qquad \int_{R_0} \pi(dx) < \infty$$

Unlike the Finite Activity models (such as the Merton Jump Model) under the family of Lévy Processes, the infinite-activity models allow infinite jumps within a finite time interval. Hence,

$$\int_{R_0} \pi(dx) \not< \infty$$

Within the category of Infinite Activity Lévy Processes, the jump process can have either finite or infinite variation. In the case of the finite/infinite variation, the sum of the absolute distance is finite/infinite over any finite time interval. For this paper, we will only be focussing on the Variance

Gamma Process - a member from the Infinite Activity with Finite Variation. This process,Z_t is created using a Brownian Motion, with drift γ and variance σ_j^2, subordinated by an independent gamma process, G_t, with parameter α.

$$G_{t+t_0} - G_t \sim Ga(\tfrac{t}{\alpha}, \tfrac{1}{\alpha})$$

2.2 The Model

For the purposes of this paper, we will be focusing on the Stochastic Volatility with Variance Gamma Jumps in Returns (SVVG). The model is as follows:

$$\begin{aligned} dY_t &= \mu dt + \sqrt{v_t}[\rho dW_{1t} + \sqrt{1-\rho^2} dW_{2t}] + dZ_t \\ dv_t &= \kappa(v - v_t)dt + \sigma_v \sqrt{v_t} dW_{1t} \\ dZ_t &= \gamma dG_t + \sigma_j \sqrt{dG_t} dB_t \end{aligned} \tag{1}$$

Using Euler Discretization, we have the following:

$$\begin{aligned} Y_{t+1} &= Y_t + \mu\Delta + \sqrt{v_t \Delta}\epsilon_{t+1}^y + Z_{t+1} \\ v_{t+1} &= v_t + \kappa(v - v_t)\Delta + \sigma_v \sqrt{v_t \Delta}\epsilon_{t+1}^v \\ Z_{t+1} &= \gamma G_{t+1} + \sigma_j \sqrt{G_{t+1}}\epsilon_{t+1}^Z \end{aligned} \tag{2}$$

where both ϵ_{t+1}^y and ϵ_{t+1}^v follow $N(0,1)$ with $corr(\epsilon_{t+1}^y, \epsilon_{t+1}^v) = \rho$ while $\epsilon_{t+1}^y, \epsilon_{t+1}^v \perp G_{t+1}$. The Jump Process Z_{t+1} follows a variance gamma process where ϵ_{t+1}^Z follows $N(0,1)$ and $\epsilon_{t+1}^Z \perp \epsilon_{t+1}^y, \epsilon_{t+1}^v, G_{t+1}$. G_{t+1} follows a Gamma Distribution, $\Gamma(\alpha, \beta)$.

Hence, we have observations $(Y_t)_{t=0}^T$; latent variables $(v_t)_{t=0}^T, (Z_t)_{t=1}^T, (G_t)_{t=1}^T$; and parameters $\Theta = \{\mu, \rho, \kappa, \upsilon, \sigma_v, \gamma, \sigma_j\}$. For convenience, we let $\theta = \{\Theta, (v_t)_{t=0}^T, (Z_t)_{t=1}^T, (G_t)_{t=1}^T\}$.

2.3 Derivation of Posterior Density

Since $corr(\epsilon_{t+1}^y, \epsilon_{t+1}^v) = \rho$ and $\epsilon_{t+1}^y, \epsilon_{t+1}^v \sim N(0,1)$, conditioning on the values of Z_{t+1},v_t and Θ, we get

$$\begin{pmatrix} Y_{t+1} - Y_t \\ v_{t+1} - v_t \end{pmatrix} | v_t, Z_{t+1}, \Theta = N\left(\begin{pmatrix} \mu\Delta + Z_{t+1} \\ \kappa(\theta - \upsilon)\Delta \end{pmatrix}, v_t\Delta \begin{pmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{pmatrix} \right) \quad (3)$$

As for the variance gamma process, Li et. al (2006) shows that conditioning on G_{t+1} and Θ, we get

$$J_{t+1}|G_{t+1}, \Theta \sim N(\gamma G_{t+1}, \sigma_j^2 G_{t+1})$$
$$G_{t+1}|\Theta \sim \Gamma(\tfrac{\Delta}{v}, v)$$

However, in Jasra (2011), it is shown that G_{t+1} can be integrated out:

$$\begin{aligned} p(Z_{t+1}|\Theta) &= \int p(Z_{t+1}|G_{t+1}, \Theta) p(G_{t+1}|\Theta)\, dG_{t+1} \\ &= \frac{2exp(\frac{\gamma Z_{t+1}^2}{\sigma})}{\alpha^{\frac{t-u}{\alpha}}\sqrt{2\pi}\Gamma(\frac{t-u}{\alpha}, \frac{1}{\alpha})} \left(\frac{Z_{t+1}^2}{\gamma^2 + 2\frac{\sigma^2}{\alpha}} \right)^{\frac{t-u}{2\alpha} - \frac{1}{4}} K_{\frac{t-u}{\alpha} - \frac{1}{2}} \left(\frac{\sqrt{Z_{t+1}^2(\gamma^2 + \frac{2\sigma^2}{\alpha})}}{\sigma^2} \right) \end{aligned} \quad (4)$$

where $K_a(.)$ is the modified Bessel Function of the second kind, and $\sigma = \sigma_j\sqrt{t-u}$. In this paper, we will let $\alpha = t - u$. Thus, the simulation proce-

dure is reduced by one dimension, and this reduces the simulation complexity, since there is one lesser latent variable to udpate.

As for the priors, we will follow the ones subscribed by Li (2006).

$p(\Theta) = p(\mu)p(\gamma)p(\sigma_j)p(\kappa)p(\sigma_v, \rho)p(\upsilon)$

$\mu, \gamma, \sigma_j \sim N(0, 1)$

$\kappa, \upsilon \sim N(0, 1)$ truncated at 0

Reparameterize (ρ, σ_v) to (ϕ, w), where $\phi = \sigma_v\rho, w = \sigma_v^2(1 - \rho^2)$

$w \sim IG(1.0, 0.5), \phi|w\ N(0, 0.5w)$

Now, the posterior distribution is as follows:

$$p(\Theta, v_{1:n}, Z_{1:n}|Y_{0:n}) \propto p(\Theta) \prod_{i=0}^{n} p(Y_{i+1}, v_{i+1}|Y_i, v_i, Z_{i+1}, \Theta)p(Z_{i+1}|\Theta) \quad (5)$$

2.4 Sequential Monte Carlo

2.4.1 Perfect Monte Carlo

Suppose we are able to simulate N i.i.d. random samples/particles $\{\theta_{0:t}^i, i = 1, 2, ..., N\}$ from a particular distribution (which in this paper, we are referring to the posterior distribution) - $p(\theta_{0:t}|y_{0:t})$. Then, the empirical estimate of this distribution is

$$P_N(d\theta_{0:t}|y_{0:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\theta_{0:t}^i}(d\theta_{0:t}) \quad (6)$$

where $\delta_{\theta^i_{0:t}}(d\theta_{0:t})$ refers to the delta-dirac mass located at $\theta^i_{0:t}$. One can then obtain the estimate of $E[f(\theta_{0:t})]$ via

$$E[f(\theta_{0:t})]_{est} = \int f(\theta^i_{0:t})\, P_N(d\theta_{0:t}|y_{0:t}) = \frac{1}{N}\sum_{i=1}^{N} f(\theta^i_{0:t}) \tag{7}$$

If the posterior variance of $f(\theta_{0:t})$ satisfies $\sigma^2 := E[f(\theta_{0:t})^2] - E[f(\theta_{0:t})]^2 < +\infty$,

$$var(E[f(\theta_{0:t})]_{est}) = \frac{\sigma^2}{N} \tag{8}$$

and by the Strong Law of Large Numbers,

$$E[f(\theta_{0:t})]_{est} \underset{N\to\infty}{\overset{\text{a.s.}}{\to}} E[f(\theta_{0:t})] \tag{9}$$

Also, if $\sigma^2 < +\infty$, the Central Limit Theorem holds

$$\sqrt{N}(E[f(\theta_{0:t})]_{est} - E[f(\theta_{0:t})]) \underset{N\to\infty}{\Longrightarrow} N(0,\sigma^2)] \tag{10}$$

Unfortunately, it often is difficult to sample from the posterior distribution, especially in high dimensional cases. Markov Chain Monte Carlo (MCMC) methods is a possible alternative, but are unsuited for recursive problems. Hence, we will introduce a new method in this paper.

2.5 Importance Sampling

One classical method is the use of Importance Sampling method (Geweke 1989). Suppose we are interested in evaluating $E[f(\theta_{0:t})]$. Using a pro-

posal density $\pi(\theta_{0:t})|y_{0:t})$, from which we can easily sample from, we get the following:

$$\begin{aligned} E[f(\theta_{0:t})]_{est} &= \int f(\theta^i_{0:t})\, P_N(d\theta_{0:t}|y_{0:t}) \\ &= \int f(\theta^i_{0:t}) p(\theta_{0:t})\, \theta_{0:t} \\ &= \int f(\theta^i_{0:t}) \frac{p(\theta_{0:t})}{\pi(\theta_{0:t})|y_{0:t})} \pi(\theta_{0:t})|y_{0:t})\, d\theta_{0:t} \\ &= \int f(\theta^i_{0:t}) w(\theta^i_{0:t}) \pi(d\theta_{0:t}|y_{0:t})\, d\theta_{0:t} \end{aligned} \tag{11}$$

Since we can easily simulate from N i.i.d. from $\pi(\theta_{0:t})|y_{0:t})$, we can get an estimate of $E[f(\theta_{0:t})]$

$$\begin{aligned} \hat{E}[f(\theta_{0:t})]_{est} &= \int f(\theta^i_{0:t})\, \hat{P}(d\theta_{0:t}) \\ &= \frac{\frac{1}{N}\sum_{i=1}^N f(\theta^i_{0:t}) w(\theta^i_{0:t})}{\frac{1}{N}\sum_{i=1}^N w(\theta^i_{0:t})} \\ &= \sum_{i=1}^N f(\theta^i_{0:t}) W(\theta^i_{0:t}) \end{aligned} \tag{12}$$

where $w(\theta^i_{0:t})$ and $W(\theta^i_{0:t})$ is known as the importance weights and normalised weights respectively. Note that we are now operating under the new measure

$$\hat{P}(\theta_{0:t}) = \frac{1}{N}\sum_{i=1}^N \delta_{\theta^i_{0:t}}(d\theta_{0:t}) W(\theta^i_{0:t}) \tag{13}$$

The Importance Sampling method does not work well in this form (especially in high dimensional cases). In addition, under $\mathcal{F}_{t+1}$, one has to recompute the weights/normalised weights despite having the weights/normalised

weights at $\mathcal{F}_t$ - the recursive problem has yet to be solved! However, this has spawned new innovations, one which we will discuss and use in this paper.

2.6 Sequential Importance Sampling

The Importance Sampling Method can be improved to include new information, $\mathcal{F}_{t+1}$, without recomputing $\{\theta^i_{0:t}, i = 1, 2, ..., N\}$. Using the following fact,

$$P(A_n, .., A_1|B) = P(A_n|A_{n-1}, .., B)P(A_{n-1}|A_{n-2}, .., B)..P(A_1|B) \quad (14)$$

we get,

$$\pi(\theta_{0:t+1}|y_{1:t+1}) = \pi(\theta_{t+1}|\theta_{0:t}, y_{1:t+1})\pi(\theta_{0:t}|y_{1:t}) \quad (15)$$

After iterating,

$$\pi(\theta_{0:t+1}|y_{1:t+1}) = \pi(\theta_0)\prod_{i=1}^{t+1}\pi(\theta_i|\theta_{0:i-1}, y_{1:i}) \quad (16)$$

Our weights can then be recursively updated using the following

$$W(\theta^i_{0:t+1}) \propto W(\theta^i_{0:t})\frac{p(y_{t+1}|\theta^i_{t+1})p(\theta^i_{t+1}|\theta^i_t)}{\pi(\theta^i_{t+1}|\theta^i_{0:t}, y_{1:t+1})} \quad (17)$$

One importance case (which we will use) is when we adopt the prior distribution as the importance distribution at $t = 0$.

2.7 Simulation using a Sequence of Densities

Using the following densities,

$$\pi_k(\theta_{1:n}|y_{0:n}) \propto p(\Theta)\prod_{i=1}^{n}[p(y_i, v_i|y_{i-1}, v_{i-1}, z_i, \Theta)^{\zeta_k}p(z_i|\Theta)] \tag{18}$$

where $0 \leqslant \zeta_1 < ... < \zeta_p = 1$. The idea is to simulate from an 'easy' density, before shifting the particles to a much more difficult density. At ζ_1, the tempered posterior density focusses more on the priors than when compared to ζ_k where $k > 1$. Overtime, as the algorithm iterates, the tempered posterior focusses less on the prior and more on the liklihood. Hence, when $\zeta_p = 1$, the particles now exist in the target posterior density.

2.8 Simulation Procedure

To start off the simulation, we first sample from the prior distribution π_0 and Importance sampling is then conducted as below

$$\begin{aligned} w_1(\theta^i) &= \frac{\pi_1(\theta_1^i|y_{0:n})}{\pi_0(\theta_1^i|y_{0:n})} \\ W_1(\theta^i) &= \frac{w_1^i}{\sum_{i=1}^{M} w_1^i} \end{aligned} \tag{19}$$

At the second iteration, the particles are shifted from π_1 to π_2 via a kernel of invariant distribution $P_2(\theta_1^i, \theta_2^i)$

$$w_2(\theta^i) = \frac{\pi_2(\theta_2^i|y_{0:n})}{\int \pi_1(\theta_1^i|y_{0:n})P(\theta_1^i, \theta_2^i)\, d\theta_1} \tag{20}$$

There are many choices of kernels, but in this paper we will adopt the Random Walk Metropolis Kernel. Hence, for $t \geqslant 2$,

$$w_k(\theta^i) = W_{k-1}^i(\theta^i) \prod_{i=1}^{n} p(y_i|y_{i-1}, v_i, z_i, \Theta)^{\zeta_k - \zeta_{k-1}} \tag{21}$$

For the Random Walk Metropolis Kernel, one has to specify the proposal variance ψ_{k+1} in order for one to update the parameter values. Using adaptive MCMC techniques (Andrieu, Moulines 2006), we approximate the mean and therefore the variance of the parameters at iteration k, for the next iteration $k+1$. Also, if the acceptance rate of the parameter is above 0.85, we tune up the variability by a factor of 5. Likewise, if the acceptance rate of the parameter is below 0.15, we tune down the variability by a factor of 1/5.

2.9 Resampling

Like the problem encountered in the Sequential Importance Sampling, the variability of the weights increases as the iteration increases. This is termed as weight degeneracy (Doucet et. al 2001). Hence, to counter this problem, we resample the particles according to the normalised weights within the cloud of particles. After resampling, the weights of the particles are reset to 1. In this paper, we will be using the Multinomial Resampling method, even though more sophisticated methods exist. One point to note is that the resampling should occur too often. When resampling occurs, the number of unique particles fall, hence reducing the particles' approximation of the target density. One criterion to measure the variability of the weights is the

Effective Sample Size (ESS) (Doucet et al. 2001).

$$ESS_k = \frac{(\sum_{i=1}^{M} w_k^i)^2}{\sum_{i=1}^{M}(w_k^i)^2} = \frac{(\sum_{i=1}^{M} W_k^i \prod_{i=1}^{n} p(y_i|y_{i-1}, v_i, z_i, \Theta)^{\zeta_k - \zeta_{k-1}})^2}{\sum_{i=1}^{M}(W_k^i \prod_{i=1}^{n} p(y_i|y_{i-1}, v_i, z_i, \Theta)^{\zeta_k - \zeta_{k-1}})^2} \quad (22)$$

The idea of this criterion is to give insight on the approximate number of samples relative to an independent simulation approach (Jasra et. al 2006). When ESS_k drops below some threshold l, we resample.

In this paper, at iteration k of the algorithm, we let $ESS_k = 0.95 ESS_{k-1}$. Using the bisection method, we are able to calculate ζ_k using (18).

2.10 The SMC Algorithm

Therefore, the algorithm is as follows: At iteration k=1, sample $\theta_1^i \sim \pi_1$ for $i = 1, .., M$ and compute

$$w_1(\theta^i) = \frac{\pi_1(\theta_1^i|y_{0:n})}{\pi_0(\theta_1^i|y_{0:n})}$$
$$W_1(\theta^i) = \frac{w_1^i}{\sum_{i=1}^{M} w_1^i}$$

If $ESS_k \leqslant l$, resample and set $W_1(\theta^i) = 1/M$. Set ψ_{k+1} for each kernel.

At iteration $k = 2, .., p$, set ζ_k and compute

$$w_k(\theta^i) = \frac{\pi_1(\theta_1^i|y_{0:n})}{\pi_0(\theta_1^i|y_{0:n})}$$
$$W_k(\theta^i) = \frac{w_k^i}{\sum_{i=1}^{M} w_k^i}$$

If $ESS_k \leqslant l$, resample and set $W_k(\theta^i) = 1/M$. If $k < l$, set ψ_{k+1} for each kernel.