# Exact Simulation of Tempered Stable Distributions with an Application to Pricing Barrier Options

MA5P1 Dissertation


**Matthew Ball**

Supervisor: Prof. Aleksander Mijatović

**Department of Mathematics**

University of Warwick


August 2023

A dissertation submitted for the partial fulfilment of the requirement for the degree of Master of Science in Mathematics

# Acknowledgements

I would like to thank Aleks, my dissertation supervisor, for his patience and invaluable guidance through this project.

# Abstract

This dissertation provides a rigorous introduction to Stable Distributions, before motivating and defining the rich and highly applicable class of Tempered Stable Distributions. An algorithm is presented which samples exactly from any given Tempered Stable Distribution, and this algorithm is subsequently analysed in detail. The algorithm is then applied to a method to sample simultaneously from both the final value and supremum of a Lévy Process to efficiently price Lévy - driven Barrier Options. This main contribution is to rigorously present an exact simulation algorithm for Tempered Stable Distributions when $\alpha \in (1, 2)$.

# Contents

# 1 Introduction

In many financial models, a key assumption is made: the returns of the underlying asset are distributed according to some variant of a normal distribution, for example the Black-Scholes-Merton options pricing formula assumes a log-normal distribution. However there is significant evidence, both rigorous and anecdotal, to suggest that this model does not accurately represent the reality of the markets. It is now widely accepted that Gaussian distributions have unrealistically light tails; they assign too low probabilities to the extreme events and market shocks which we have become accustomed to. Gaussian distributions are also symmetric about the origin, another assumption which has been observed as unrepresentative of the reality, where returns are often skewed.

So the search began for a new distribution, one which could both allow for skew and account for the probability of extreme events more accurately. In 1963, Mandelbrot [16] suggested the class of *Stable Distributions*, first defined by Lévy and Khintchine in the 1920's and 30's, as a heavy tailed alternative to the Gaussian in a financial context. Whilst there are many possible heavy tailed alternatives, stable distributions have the unique defining property that a linear combination of stable distributions still observes a stable distribution, albeit possibly scaled and shifted. Gaussian distributions are also a special case of stable distributions, so they are a natural generalisation. This makes them particularly applicable for financial modelling, though often stable distributions can exhibit too heavy tails, leading to infinite expectations and thus infinite options prices, which are useless in practice. This motivates the introduction of *Tempered Stable Distributions*, which approximate stable distributions around the mean but have tails which decay exponentially, ensuring all moments are finite. First defined as a concept in physics by Kopenen, 1995 [11], these distributions have begun to gain popularity in financial models as a practical yet accurate model for financial asset returns.

This dissertation begins in Sections 2 and 3 by introducing, in turn, infinitely divisible distributions, stable distributions and tempered stable distributions, defining each and exploring some basic and useful properties for understanding and applying these distributions. We then cover the problem of sampling from tempered stable distributions, and present and analyse an exact sampling algorithm. The dissertation concludes by applying this algorithm for the purpose of pricing barrier options, where the underlying asset returns are assumed to follow a tempered stable distribution.

# 2 Lévy Processes and Infinitely Divisible Distributions

## 2.1 Lévy Processes

When modelling a quantity evolving over time as a stochastic process, it is often useful for the model to satisfy a few basic requirements. Firstly, that the change over a certain period of time has the same distribution regardless of the time we started measuring, and secondly that the distribution of each change is independent of the observed values of the quantity up to that point. If we additionally ask for some basic continuity requirements so the process makes sense as a model, we arrive at the class of stochastic processes known as Lévy Processes, which are formally defined as follows:

**Definition 2.1.1.** A $d$-dimensional Lévy Process is a stochastic process $\{ X_t : t \geq 0 \}$ on $\mathbb{R}^d$ with $\mathbb{P}(X_0 = 0) = 1$ which satisfies the following conditions:

1. Independent Increments:
   For any integer $n > 0$ and real numbers $0 < t_1 < \cdots < t_n$, the 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. Stationary Increments:
   For any $s, t \geq 0$ we have $X_{t+s} - X_s \stackrel{d}{=} X_t$.

3. Stochastic Continuity:
   For any $\varepsilon > 0$ and $t \geq 0$ we have $\lim_{s \to t} \mathbb{P}(|X_t - X_s| > \varepsilon) = 0$.

4. Right Continuity and Left Limits:
   There exists $\Omega \in \mathcal{F}$ with $\mathbb{P}(\Omega) = 1$ and for every $\omega \in \Omega$, $X_t(\omega)$ is right continuous for $t \geq 0$ and has left limits for $t > 0$.

We give three basic examples of Lévy Processes, which will be useful later on. We will show later in this section that in fact every Lévy process can be characterised by superposing variants of these processes.

**Definition 2.1.2** (Poisson Process)**.** Let $(W_n)_{n \in \mathbb{N}}$ be a random walk on $\mathbb{R}$, starting at 0, such that each increment $T_{n+1} := W_{n+1} - W_n$ follows an exponential distribution with expectation $c > 0$. We define a *Poisson process* $X_t$ by $X_t(\omega) = n$ if and only if $t \in [W_n(\omega), W_{n+1}(\omega))$.

This process is well defined as $W_n \to \infty$ almost surely, as $\mathbb{P}(W_n \leq K) \leq \mathbb{P}(T_1 \leq K, \ldots, T_n \leq K) \leq \mathbb{P}(T_1 \leq K)^n \to 0$ as $n \to \infty$. It is known [20] that $W_n$ follows a Gamma distribution with parameters $c, n$, ie

$$\mathbb{P}(W_n \leq t) = \frac{c^n}{\Gamma(n)} \int_0^t x^{n-1} e^{-cx} \mathrm{d}x$$

and from this it is not hard to show that $X_t$ follows a Poisson distribution with mean $ct$, ie

$$\mathbb{P}(X_t = n) = \frac{(ct)^n e^{-ct}}{n!}$$

Some texts define a Possion process by this distribution of $X_t$, but I have chosen to define it by its construction as I believe this is more enlightening to the underlying behaviour of the process and why it is useful in modelling. This process is most often used to model quantities which increase only in increments of 1, for example customers arriving at a queue, or immigration in a population model. We now show from the definition that a Poisson process is indeed a Lévy Process.

**Proposition 2.1.3.** *The Poisson process is a Lévy process*

*Proof.* Let $X_t$ be a Possion process. By definition $\mathbb{P}(X_0 = 0) = \mathbb{P}\Big(0 \in \big[W_0(\omega), W_1(\omega)\big)\Big) = 1$. Then stochastic continuity and the continuity and limit requirements come from the paths of $X_t$ being right continuous step functions with jumps of size 1 at times $T_1, T_2, ...$. It remains to show the increments of the process are independent and stationary.

Then if $s, t > 0$,

$$\begin{aligned}
\mathbb{P}(X_{t+s} = n, X_t = n) &= \mathbb{P}(W_n \leq t, T_{n+1} > t + s - W_n) \\
&= \int_0^t \int_{t+s-x}^{\infty} f_{T_{n+1}}(y) f_{W_n}(x) \mathrm{d}y \mathrm{d}x \\
&= \int_0^t \int_{t+s-x}^{\infty} ce^{-cy} \frac{c^n}{\Gamma(n)} x^{n-1} e^{-cx} \mathrm{d}y \mathrm{d}x \\
&= \frac{c^{n+1}}{(n-1)!} \int_0^t x^{n-1} e^{-cx} \mathrm{d}x \int_{t+s-x}^{\infty} e^{-y} \mathrm{d}y \;=\; \frac{c^n t^n}{n!} e^{c(t+s)}
\end{aligned}$$

where the second equality follows since $W_n$ and $T_{n+1}$ are independent, and the third equality uses the relation $\Gamma(n) = (n-1)!$ for any integer $n$. So using the formula for conditional expectation, and the known distribution of $X_t$, we have

$$\mathbb{P}(X_{t+s} = n \,|\, X_t = n) = \mathbb{P}(X_{t+s} = n) / \, \mathbb{P}(X_t = n) = e^{-cs}.$$

Now let $n \geq 0$ and $m \geq 1$. Then we consider the distribution of $(W_{n+1} - t, T_{n+2}, ..., T_{n+m})$

conditioned on $X_t = n$. Let $s_1, \dots, s_m > 0$. Then

$$
\begin{aligned}
&\mathbb{P}(W_{n+1} - t > s_1, T_{n+2} > s_2, ..., T_{n+m} > s_m \,|\, X_t = n) \\
&= \mathbb{P}(X_t = n, W_{n+1} - t > s_1, T_{n+2} > s_2, \dots, T_{n+m} > s_m) / \,\mathbb{P}(X_t = n) \\
&= \mathbb{P}(X_t = n, W_{n+1} - t > s_1)\, \mathbb{P}(T_{n+2} > s_2, \dots, T_{n+m} > s_m) / \,\mathbb{P}(X_t = n) \\
&= \mathbb{P}(W_{n+1} - t > s_1 \,|\, X_t = n)\, \mathbb{P}(T_{n+2} > s_2, \dots, T_{n+m} > s_m) \\
&= e^{-s}\, \mathbb{P}(T_{n+2} > s_2, \dots, T_{n+m} > s_m) \\
&= \mathbb{P}(T_1 > s_1)\, \mathbb{P}(T_2 > s_2, \dots, T_m > s_m) \\
&= \mathbb{P}(T_1 > s_1, T_2 > s_2, \dots, T_m > s_m)
\end{aligned}
$$

where we have used that all of the $T_i$s are i.i.d. with $T_{n+i} \stackrel{d}{=} T_i$. So the distribution of $(W_{n+1} - t, T_{n+2}, ..., T_{n+m})$ conditioned on $X_t = n$ is equal to the distribution of $T_1, T_2, \dots T_m$. Notice that the event $\{X_{t+s} = n + m\} = \{W_{n+m} \le t + s < W_{n+m+1}\}$ can be rewritten as

$$\{(W_{n+1} - t) + T_{n+2} + \cdots + T_{n+m} \le s < (W_{n+1} - t) + T_{n+2} + \cdots + T_{n+m+1}\}$$

and hence $\mathbb{P}(X_{t+s} = n + m \,|\, X_t = n) = \mathbb{P}(W_{n+m} \le t + s < W_{n+m+1} \,|\, X_t = n) = \mathbb{P}(W_m < s < W_{m+1}) = \mathbb{P}(X_s = m)$. Therefore

$$\mathbb{P}(X_{t+s} = n + m)\, \mathbb{P}(X_t = n) = \mathbb{P}(X_s = m)\, \mathbb{P}(X_t = n)$$

Summing over all values of n gives $\mathbb{P}(X_{t+s} - X_t = m) = \mathbb{P}(X_s = m)$. So we have the stationary increments property.

By exactly the same argument, for $0 \le t_0 < \cdots < t_k$ and $n_0, n_1, \dots, n_k > 0$,

$$
\begin{aligned}
&\mathbb{P}(X_{t_0} = n_0, X_{t_1} - X_{t_0} = n_1, \dots, X_{t_k} - X_{t_{k-1}} = n_k) \\
&\quad = \mathbb{P}(X_{t_0} = n_0, X_{t_1} = n_0 + n_1, \dots, X_{t_k} = n_0 + n_1 + \cdots + n_k) \\
&\quad = \mathbb{P}(X_{t_0} = n_0)\, \mathbb{P}(X_{t_1} = n_0 + n_1, \dots, X_{t_k} = n_0 + n_1 + \cdots + n_k \,|\, X_{t_0} = n_0) \\
&\quad = \mathbb{P}(X_{t_0} = n_0)\, \mathbb{P}(X_{t_1 - t_0} = n_1, \dots, X_{t_k - t_0} = n_1 + \cdots + n_k)
\end{aligned}
$$

Repeating this inductively gives the independent increments property. So we have satisfied the definition and $X_t$ is a Lévy process. □

We now generalise Poisson processes by introducing the *Compound Poisson process*. Again I define the process by its construction.

**Definition 2.1.4** (Compound Poisson Process)**.** Let $N_t$ be a Poisson process as in Definition 2.1.2 with parameter $c > 0$, and let $(S_n)_{n\in\mathbb{N}}$ be a collection of i.i.d random variables, say $S_n \sim S$, defined on the same probability space, which are independent of $N_t$ and satisfy $\mathbb{P}(S = 0) = 0$. Then we define a (one-dimensional) Compound Possion process $X_t$ as

$$X_t = \sum_{n=1}^{N_t} S_n$$

The Compound Poisson process is a Lévy Process, and the proof is fairly similar to that of a Poisson process. For a full proof see [20, Ch 1.4]. This process is also a jump process, with the jumps arriving according to the Posssion process $N_t$, but the size of the jumps are also random, and distributed according to the law of $S$. So the Compound Poisson process has 2 parameters, the parameter $c > 0$ which governs $N_t$, and the law of $S$, which we call $\mathcal{F}$ and is a probability measure with no mass at 0. The Poisson process is a special case, where $F = \delta_1$.

The final basic process is probably the most famous and heavily used - Brownian motion. First observed by Brown in 1827 when he noticed the seemingly random movements of pollen particles in water, and formalised mathematically by Norbert Wiener in the 1920s, Brownian motion is a continuous random process where each increment is independent and normally distributed, with variance equal to the time spent observing. We define a generalised version where there is some linear drift beneath the random noise.

**Definition 2.1.5** (Brownian motion with drift)**.** A stochastic process $W_t$ is a Brownian Motion if

1. For any integer $n > 0$ and real numbers $0 < t_1 < \cdots < t_n$, the 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 and normally distributed with $X_{t_i} - Xt_{i-1} \sim \mathcal{N}(0, t_i - t_{i-1})$.

2. $W_0 = 0$ a.s.

3. $t \mapsto W_t$ is continuous almost surely

We then define Brownian Motion with a linear drift to be

$$Z_t = \sigma W_t + \gamma t, t \geq 0$$

where $\gamma \in \mathbb{R}, \sigma > 0$ are constants.

## 2.2 Infinitely Divisible Distributions

Lévy Processes are a very large subset of stochastic processes, and include many other processes commonly used in modelling. However, using this definition directly, it is hard to begin to generalise or parametrise Lévy Processes, and this prompted the introduction of a special type of probability distribution known as an *infintely divisible* distribution and show an intimate connection between these distributions and Lévy Processes which we will explore now. For a comprehensive resource on infinitely divisible distributions and Lévy Processes, see [20].

**Definition 2.2.1.** A distribution $\mu$ on $\mathbb{R}^d$ is infinitely divisible if for every positive integer $n$, there exists a distribution $\mu_n$ also on $\mathbb{R}^d$ such that if $X \sim \mu$ and $Y_n^{(1)}, Y_n^{(2)}, \ldots, Y_n^{(n)}$ are i.i.d with distribution $\mu_n$ then

$$X \stackrel{d}{=} \sum_{i=1}^{n} Y_n^{(i)}.$$

Equivalently $\mu = \mu_n^n$, recalling that $\mu_n^n$ is taken to mean the convolution of $\mu_n$ with itself $n$ times.

It is fairly straightforward to show that the distribution of a Lévy Process at any point in time is infinitely divisible.

**Proposition 2.2.2.** *Let $X_t$ be a Lévy Process. Then for any $t > 0$, the distribution of $X_t$ is infinitely divisible.*

*Proof.* [20]
Set $n > 0$ to be an integer. For $i = 0, 1, \ldots, n$ define

$$t_i := \frac{it}{n} \text{ and } Y_n^{(i)} = X_{t_k} - X_{t_{k-1}}.$$

Then the $Y_i$ are i.i,d by independence of increments for Lévy Processes , and by telescoping sum using $X_0 = 0$ almost surely,

$$X_t \stackrel{d}{=} \sum_{i=1}^{n} Y_n^{(i)}$$

□

It can be shown ([20, Thm 7.10]) that in fact the converse holds too; for every infintely divisible distribution $\mu$, there exists a Lévy process $X_t$ such that $X_1 \sim \mu$. Hence we have the following theorem:

**Theorem 2.2.3.** There is a one to one correspondence between infinitely divisible distributions and Lévy processes

**Remark 2.2.4.** Note that the distribution of $X_1$ entirely determines the distribution of $X_t$ for any $t > 0$. This is easiest to get a sense of by considering the characteristic function for each $X_t$, defined by $\varphi_{X_t}(y) = \mathbb{E}[\exp(iyX)]$. Knowledge of $\varphi_{X_t}(y)$ is sufficient to characterise the entire distribution of $X_t$, so we aim to express $\varphi_{X_t}$ in terms of $\varphi_{X_1}$ for all $t > 0$. We already know the distribution of $X_0$ is $\delta_0$ for all Lévy processes.

So if $n, m \in \mathbb{N}$, by independent stationary increments we have that

$$X_n \overset{d}{=} \sum_{i=1}^{n} X_1 \text{ and } X_1 \overset{d}{=} \sum_{i=1}^{m} X_{1/m}$$

and hence $\varphi_{X_n} = \varphi_{X_1}^n$ and $\varphi_{X_{1/m}} = \varphi_{X_1}^{1/m}$. Next we consider a general rational $t$, say $t = m/n$ . By the same methods, we have:

$$X_{m/n} \overset{d}{=} \sum_{i=0}^{m} X_{1/n}$$

and from this $\varphi_{X_{m/n}} = (\varphi_{X_{1/n}})^m = \varphi_{X_1}^{(m/n)}$ Finally we consider the case where $t$ is irrational. From a result in analysis, we can find a series $\sum_{i=0}^{\infty} r_i$ of rational numbers for which the sum converges to $t$. Then

$$X_t \overset{d}{=} \sum_{i=0}^{\infty} X_{r_i}$$

so $\varphi_{X_t} = \prod_{i=1}^{\infty} \varphi_{X_{r_i}} = \prod_{i=1}^{\infty} \varphi_{X_1}^{r_i} = \varphi_{X_1}^{\sum_{i=1}^{\infty} r_i} = \varphi_{X_1}^t$. For a more rigorous proof see [20, Lm 7.9]

Throughout this dissertation it will often be useful to consider the characteristic function of a distribution, and clearly $X$ follows an infinitely divisible distribution if and only if for every positive integer $n$, there is some other characteristic function for another random variable, say $\varphi_{Y_n}$ with $(\varphi_{Y_n})^n = \varphi_X$.

We calculate the characteristic functions for the infinitely divisible distributions associated to the three examples of Lévy Processes given earlier. A Poisson process is supported only on the nonnegative integers, so if $N_t$ is a Poisson process with parameter $c > 0$,

$$\begin{aligned}
\mathbb{E}[\exp iyN_1] &= \sum_{k=0}^{\infty} e^{iyk}\, \mathbb{P}(N_1 = k) \\
&= \sum_{k=0}^{\infty} e^{iyk} \frac{(ct)^k e^{(-c)}}{k!} \\
&= e^{-c} \sum_{k=0}^{\infty} \frac{(e^{iy}c)^k}{k!} \\
&= e^{-c} e^{cte^{iy}} = \exp\big\{c(e^{iy} - 1)\big\}
\end{aligned}$$

So using the remark, $\varphi_{N_t}(y) = \exp\big\{tc(e^{iy} - 1)\big\}$.

For a compound Poisson process $X_t$, and using the notation of definition 2.1.4, we have

$$\varphi_S(y) = \mathbb{E}[e^{iyS}] = \int_{\mathcal{A}} e^{iyz} F(\mathrm{d}z)$$

where $\mathcal{A} \subseteq \mathbb{R}$ is the support of the measure $F$. With this in hand, we can compute the characteristic function of $X_1$.

$$\begin{aligned}
\mathbb{E}[e^{iyX_1}] &= \sum_{k=0}^{\infty} \mathbb{E}[e^{iyX_1}|N_1 = k]\,\mathbb{P}(N_t = k) \\
&= \sum_{k=0}^{\infty} \big[\varphi_S(y)\big]^k \frac{(c)^k e^{(-c)}}{k!} \\
&= e^{-c} \sum_{k=0}^{\infty} \frac{\big[\varphi_S(y)c\big]^k}{k!} \\
&= e^{-c} e^{\varphi_S(y)c} = e^{c(\varphi_S(y)-1)}
\end{aligned}$$

Using the formula for $\varphi_S$ and the fact $F$ is a probability distribution, we get

$$\varphi_{X_1}(y) = \exp\left\{ c \int_{\mathcal{A}} (e^{iyz} - 1) F(\mathrm{d}z) \right\}$$

**Remark 2.2.5.** It is possible to use this representation to prove directly that a compound Poisson distribution is indeed infinitely divisble (and hene a compound poisson process is indeed a Lévy process) by writing its characteristic as the product of the characteristic function of $n$ Poisson random variables with parameter $c/n$, see [6, Ch 11].

For the final example, it is well known that the characteristic function of a normal random variable with mean 0 and variance $\sigma$ is given by $\exp\{-\frac{1}{2}\sigma^2 y^2\}$. Hence the characteristic function of Brownian Motion with drift $Z_t$ with parameters $\gamma$ and $\sigma$ is given by

$$\varphi_{Z_1}(y) = \exp\left\{ iy\gamma - \frac{1}{2} y^2 \sigma^2 \right\}$$

Characteristic functions are particularly important for Lévy Processes as it is possible, via a result of Lévy and Khintchine, to explicitly obtain an analytic formula for the characteristic function of any infinitely divisible distribution, and hence any Lévy process.

**Theorem 2.2.6** (Lévy - Khintchine representation)**.** A (one-dimensional) random variable $X$ follows an infinitely divisible distribution if and only if the characteristic function $\varphi_X$ is of the form

$$\varphi_X(y) = \exp\left\{ i\gamma y - \frac{(\sigma y)^2}{2} + \int_{\mathbb{R}} \left[ e^{iyz} - 1 - iyh(z) \right] F(\mathrm{d}z) \right\} \tag{1}$$

where $\gamma \in \mathbb{R}$, $\sigma > 0$ and $F$ is a $\sigma$-finite measure satisfying $F(\{0\}) = 0$ and

$$\int_{\mathbb{R}\setminus\{0\}} \min(1, z^2) F(\mathrm{d}z) < \infty$$

and $h : \mathbb{R} \to \mathbb{R}$ is a fixed, bounded measurable function, known as the *truncation function*, and satisfies $h(z) = z + O(z^2)$ as $z \to 0$ and $h(z) = O(1)$ as $|z| \to \infty$.

The proof is long and thus omitted but for a full proof see [20] (among many others).

We call $\gamma$ the *shift*, $\sigma$ the *Gaussian part* and $F$ the *Lévy measure*, and we will call the triplet $(\gamma, \sigma, F)$ the Lévy triplet, and this triplet completely characterises an infinitely divisible distribution. Using the correspondence between Lévy processes and infinitely divisible distributions, we define the Lévy triplet of a Lévy process $X_t$ to be the Lévy triplet of the law of $X_1$. We can then charatcerise the distribution of $X_t$ for general $t > 0$ by raising (1) to the power of $t$ (see remark 2.2.4).

The conditions on $h$ along with the condition on $F$ guarantee finiteness of the integral. Most commonly, I will use $h(z) = z\mathbb{1}_{|z|<1}$, which is a standard choice in the literature. For a useful collection of equivalent statements of this representation, see [9, Thm. 2.2]. It is also useful to note that different choices of $h$ yield the same measure $F$ and Gaussian part $\sigma$, but with a different constant $\gamma$; changing $h$ to $\tilde{h}$ corresponds to changing $\gamma$ to $\tilde{\gamma} = \gamma + \int_{\mathbb{R}} [h(z) - \tilde{h}(z)] F(\mathrm{d}z)$.

For our three example distributions, using the characteristic functions derived earlier , we can see that (using truncation function $h(z) = \mathbb{1}_{|z|<1}$) the Lévy triplet for the Poisson process is given by $(0, 0, c\delta_1)$, the triplet for Brownian Motion with drift is $(\gamma, \sigma, 0)$, and the Compound Poisson has Lévy triplet $(c\int_{|x|<1} xF(\mathrm{d}z), 0, cF)$. For the Compound Poisson process, we can use truncation function $h(z) = 0$, to get an alternative triplet $(0, 0, cF)$. This is a good example of how in general this triplet is not unique.

An interpretation of the representation using $h(z) = \mathbb{1}_{|z|<1}$ is that any Lévy process is made up of 2 components; a Brownian Motion scaled by $\sigma$ with drift $\gamma$ and a jump process governed by

$F$. This interpretation is formalised in [12] by rewriting (1) as

$$\varphi_{X_t}(y) = \exp\Bigg\{ t\left(iy\gamma - \frac{(y\sigma)^2}{2}\right) \tag{2}$$

$$+ tF(\mathbb{R}\setminus(-1,1)) \int_{|x|\geq 1} (e^{iyz}-1)\frac{1}{F\big(\mathbb{R}\setminus(-1,1)\big)} F(\mathrm{d}z) \tag{3}$$

$$+ t\int_{0<|x|<1} (e^{iyz}-1-iyz)\, F(\mathrm{d}z) \Bigg\} \tag{4}$$

Each summand (2), (3) and (4) relates to the characteristic function of a different Lévy process. (2) is the characteristic function of a scaled Brownian Motion with drift, say $B_t$, and 3 relates to a compound Poisson Process, say

$$C_t = \sum_{i=1}^{N_t} S_i$$

where $N_t$ is a Poisson process with rate $F\big(\mathbb{R}\setminus(-1,1)\big)$, which is finite by definition, and each $Z_i$ is i.i.d with law $F(dx)/F\big(\mathbb{R}\setminus(-1,1)\big)$ supported only on $\{\, x : |x| \geq 1 \,\}$. In the case $F\big(\mathbb{R}\setminus(-1,1)\big) = 0$, we simply say that the second term does not exist. Since a Poisson process can only have finitely many jumps in a finite interval, it follows that $X_t$ only has finitely many jumps larger than 1 over a finite interval.

The case of interest now is the final term (4). We want to show that it relates to the characteristic function of another Lévy process. Still using the methods of [12], we split the set $0 < |x| < 1$ into the disjoint partition $\mathcal{P} = \{P_1, P_2, \dots\}$ where

$$P_n = \{2^{-(n+1)} \leq |x| < 2^{-n}\}$$

Define $\lambda_n := F(P_n)$ and $F|_{P_n}$ to be the restriction of $F$ to the set $P_n$. Then we can rewrite (4) as

$$\begin{aligned}\int_{0<|x|<1} (1-e^{iyz}-iyz)F(\mathrm{d}z) = \sum_{n\geq 0}\Bigg[\lambda_n \int_{P_n} (1-e^{iyz})\lambda_n^{-1}F|_{P_n}(\mathrm{d}z) \\ - iy\lambda_n \int_{P_n} x\lambda_n^{-1}F|_{P_n}(\mathrm{d}z)\Bigg]\end{aligned}$$

(Again if $\lambda_n = 0$ we assume that the integral is 0). The first term is the characteristic function of a compound poisson process and the second term represents linear drift, so we conclude that the third Lévy process is the sum of at most countably many compound Poisson processes with drift. This result is known as the Lévy - Ito decomposition, and attributed (unsurprisingly) to Lévy [15] and Ito [7].

**Theorem 2.2.7** (Lévy - Ito decomposition)**.** Let $X_t$ be a Lévy process. Then we can express

$$X_t = Z_t + C_t + D_t$$

where $Z_t$ is a Brownian Motion with drift, $C_t$ is a compound Poisson process and $D_t$ is a square integrable martingale with characteristic function given by (4).

For rigorous details on this process $D_t$, see [12, Ch. 4] For our purposes, it will be sufficient to understand that the processes $C_t$ and $D_t$ characterise the jumps of the process of size $> 1$ and $\leq 1$ respectively. We call $C_t$ the large jump process and $D_t$ the small jump process.

The choice of 1 as the boundary between small and large jumps was arbitrary and can be replaced by any $\varepsilon > 0$, so we have that for any $\varepsilon > 0$, $X_t$ must have finitely many jumps greater than $\varepsilon$ in any finite time interval. The small jumps are characterised using the following definition.

**Definition 2.2.8.** Let $X_t$ be a Lévy process with triplet $(\gamma, \sigma, F)$ and $\varepsilon > 0$. Then if

$$\int_{|z|<\varepsilon} |z| F(\mathrm{d}z) < \infty$$

we say $X_t$ has *finite variation.* Infinite variation is defined analogously.

If we have finite variation, we no longer need the truncation function $h(z)$ to satisfy $h(z) = z + O(z^2)$ as $z \to \infty$ to ensure integrability, so in many cases we can simply choose $h = 0$ to simplify the formulas and remove the indicator function. In the infinite variation case, we have that $\mathbb{E}[X_t] = \infty$ for all $t$, as summing the jumps results in an undefined integral. In fact we can generalise which moments of $X_t$ are finite using the following lemma, [20, Cor 25.8].

**Lemma 2.2.9.** If $X_t$ is a Lévy process with triplet $(\gamma, \sigma, F)$, then $\mathbb{E}[X_t^p] < \infty$ is and only if $\int_{|z|\geq 1} |z|^p F(\mathrm{d}z) < \infty$

## 2.3 Changing the Measure

It is important to remember that whenever we are taking expectations or considering laws of random variables, we are doing so with respect to the underlying probability measure $\mathbb{P}$. It is useful now for us to consider under what conditions would be required on a new measure $\mathbb{Q}$ for $X_t$ to still be a Lévy process, and what is the triplet of $X_t$ under $\mathbb{Q}$? All of these are answered in a version of a theorem from Sato [20, Thm 33.1] presented in the thesis of Wannenwetsch [22]. First we define some concepts from measure theory.

**Definition 2.3.1.** If $\mathbb{P}$ and $\mathbb{Q}$ are measures, then we say $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$, written $\mathbb{Q} \ll \mathbb{P}$ if for any measurable set $A$, $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$. If $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$, then we say $\mathbb{P}$ and $\mathbb{Q}$ are equivalent and write $\mathbb{P} \sim \mathbb{Q}$.

**Theorem 2.3.2** (Radon-Nikodym derivative)**.** Let $\mathbb{P}$, $\mathbb{Q}$ be probability measures on the same measure space $(\Omega, \mathcal{F})$. Then if $\mathbb{Q} \ll \mathbb{P}$, there exists a measurable non negative function $f : \Omega \to \mathbb{R}_+$ such that for any measurable set $A \subseteq \Omega$,

$$\mathbb{Q}(A) = \int_A f(z)\,\mathbb{P}(\mathrm{d}z)$$

We say $f$ is the *Radon-Nikodym* derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ and write $f = \dfrac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}$.

**Theorem 2.3.3.** [22, Thm 1.20] Let $X_t$ be a Lévy process with respect to the measure $\mathbb{P}$, with Lévy triplet $(\gamma, \sigma, \mathcal{F})$. Then there exists a probability measure $\mathbb{Q}$ with $\mathbb{Q} \sim \mathbb{P}$ and $X_t$ is a Lévy process under $\mathbb{Q}$ with triplet $(\tilde{\gamma}, \tilde{\sigma}, \tilde{F})$ if and only if there exists $a \in \mathbb{R}$ and a function $y : \mathbb{R} \to \mathbb{R}_+$ satisfying:

$$\begin{aligned}
&\int_{\mathbb{R}\setminus\{0\}} |h(z)(1 - y(z))| F(\mathrm{d}z) < \infty \\
&\int_{\mathbb{R}\setminus\{0\}} \left(1 - \sqrt{y(z)}\right)^2 F(\mathrm{d}z) < \infty \\
&\tilde{\gamma} = \gamma - a\sigma^2 + \int_{\mathbb{R}\setminus\{0\}} h(z)(1 - y(z)) F(\mathrm{d}z) \\
&\tilde{\sigma} = \sigma \\
&\frac{\mathrm{d}\tilde{F}}{\mathrm{d}F} = y
\end{aligned}$$

For proof see [20, Thm 33.1] or Jacod and Shiryaev [8].

We can also study the form of the new measure $\mathbb{Q}$, and relate how specific transformations of measure impact the Lévy triplet. We will focus on specific case, known as the *Esscher transform.*

**Definition 2.3.4.** Let $\mathbb{P}$ be a probability measure on a space $(\Omega, \mathcal{F})$. Let $\theta \in \mathbb{R}$ such that $\int_{\mathbb{R}\setminus\{0\}} e^{\theta z} F(\mathrm{d}z) < \infty$. Define the function

$$\mathcal{E}_\theta(z) = \frac{e^\theta z}{\mathbb{E}^{\mathbb{P}}[e^{\theta X_1}]}$$

Then we define the *Esscher transform* $\mathrm{Ess}_\theta(\mathbb{P})$ as the probability measure satisfying

$$\frac{\mathrm{d}(\mathrm{Ess}_\theta(\mathbb{P}))}{\mathrm{d}\,\mathbb{P}} = \mathcal{E}_\theta(z)$$

**Theorem 2.3.5.** Let $X_t$ be a Lévy process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a distribution characterised by $(\gamma, \sigma, F)$. Then if $\theta \in \mathbb{R}$, under the probability measure $\mathrm{Ess}_\theta(\mathbb{P})$, $X_t$ is still a Lévy process with triplet

$$\begin{aligned}
&\tilde{\gamma} = \gamma - \theta\sigma^2 + \int_{\mathbb{R}\setminus\{0\}} h(z)(1 - e^{\theta z})F(\mathrm{d}z) \\
&\tilde{\sigma} = \sigma \\
&\frac{\mathrm{d}\tilde{F}}{\mathrm{d}F} = e^{\theta z}
\end{aligned}$$

This is very similar to [13, Thm. 3.9], but here I do not specify a truncation function $h$. A much more general version is given in [20, Thm 33.2].

This theorem is particularly applicable in the case of the distribution having very heavy tails. For many applications, it is not desirable for the Lévy process driven model to have infinite moments, and since by Lemma 2.2.9 the moments of the process is determined entirely by the Lévy measure, using the relations in Theorem 2.3.3, it is possible to modify the Lévy measure to ensure all moments are finite by transforming the underlying probability measure. This process is known as *tempering*, and will play a big role in future chapters on modifying the tails of stable distributions.

# 3 Stable Distributions

As discussed in the introduction, many market models, to obtain workable solutions require that the returns of stocks follow a normal distribution. We now introduce the subclass of infinitely divisible distributions known as stable distributions, which are seen to generalise normal distributions, maintaining the useful property of invariance under linear combinations, whilst allowing for heavy tails and skew. We will explore in detail the characteristics of stable distributions which can be derived from this relation, and show how heavy the tails of these distributions are. In subsection 3.2, we use the discussion of stable distributions to motivate and define tempered stable distributions.

In this section, if $X$ is a random variable, then $X_1, X_2, \ldots$ will always denote independent copies of $X$.

## 3.1 Stable Distributions

**Definition 3.1.1.** Let $X$ be a random variable and recall $X_1, X_2$ are independent copies of $X$. Then $X$ has a stable distribution there exists $c > 0$ and $d \in \mathbb{R}$ such that

$$aX_1 + bX_2 \overset{d}{=} cX + d \tag{5}$$

The distribution is strictly stable if $d = 0$ for all $a$ and $b$.

**Lemma 3.1.2.** If $X$ is a stable random variable, then for any $n \geq 2$ there exists $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$\sum_{i=1}^{n} X_i \overset{d}{=} c_n X + d_n$$

*Proof.* [17, Ch 3] Firstly let $X$ be stable as in definition 3.1.1. We proceed by induction on $n$. If $n = 2$, by definition we can find $c > 0, d \in \mathbb{R}$ with $X_1 + X_2 \overset{d}{=} cX + d$, and thus the result follows with $c_2 = c, d_2 = d$. So assume the result holds for some $n \geq 2$. Then

$$\sum_{i=1}^{n+1} X_i = X_{n+1} + \sum_{i=1}^{n} X_i \overset{d}{=} X_{n+1} + c_n X + d_n$$

where $c_n > 0$ and $d_n \in \mathbb{R}$ exist by the inductive hypothesis. As $X$ and $X_{n+1}$ are i.i.d and follow a stable distribution, by definition we can find $\tilde{c} > 0$ and $\tilde{d} \in \mathbb{R}$ such that $X_{n+1} + c_n X \overset{d}{=} \tilde{c}X + \tilde{d}$. So

$$\sum_{i=1}^{n+1} X_i \overset{d}{=} X_{n+1} + c_n X + d_n \overset{d}{=} \tilde{c}X + \tilde{d} + d_n$$

and the result follows by setting $c_{n+1} = \tilde{c}, d_{n+1} = \tilde{d} + d_n$.

□

**Corollary 3.1.3.** If $X$ is strictly stable, lemma 3.1.2 holds with $d_n = 0$ for each $n$

*Proof.* If $X$ is strictly stable, we can exactly replicate the previous proof, noting that $d, d_n$ and $\tilde{d}$ are all 0 and hence $d_{n+1} = 0$. □

**Lemma 3.1.4.** Let $n, m$ be positive integers. Then $c_{nm} = c_n c_m$

*Proof.* Consider the sum of $nm$ independent copies of $X$. Then by Lemma 3.1.2, we can find

$c_{nm} > 0$ with

$$\begin{aligned} c_{nm}X &\stackrel{d}{=} X_1 + X_2 + \cdots + X_{nm} \\ &\stackrel{d}{=} (X_1 + \cdots + X_n) + (X_{n+1} + \cdots + X_{2n}) + \cdots + (X_{nm-n+1} + \cdots + x_{nm}) \\ &\stackrel{d}{=} c_n X_1 + \cdots + c_n X_m \\ &\stackrel{d}{=} c_n(X_1 + \cdots + X_m) \stackrel{d}{=} c_n(c_m X) \end{aligned}$$

Thus $c_{nm} = c_n c_m$ as required. □

**Corollary 3.1.5.** If $m, k$ are positive integers then $c_{m^k} = (c_m)^k$.

Generally, it is not possible to write down a closed form expression for the density function for a stable distribution. However we can gain information about the characteristic function, and the constant $c_n$. Specifically, we can show $c_n = n^{\frac{1}{\alpha}}$, where $0 < \alpha \leq 2$, and $\alpha$ will be known as the *index of stability* for the distribution, and is an important parameter when defining general stable distributions.

We can simplify a lot of the arguments by initially only considering symmetric random variables. Recall that a random variable $X$ is symmetric if $X \stackrel{d}{=} -X$.

**Lemma 3.1.6.** If $X$ is a stable and symmetric random variable, then $X$ is strictly stable.

*Proof.* If $X$ is stable, then we have $aX_1 + bX_2 \stackrel{d}{=} cX + d$ so

$$cX \stackrel{d}{=} aX_1 + bX_2 - d$$

Then $-cX \stackrel{d}{=} -aX_1 - bX_2 + d \stackrel{d}{=} aX_1 + bX_2 + d$ as $X_1$ and $X_2$ are also symmetric.
Thus

$$cX \stackrel{d}{=} -cX \stackrel{d}{=} aX_1 + bX_2 + d$$

and hence $cX - d \stackrel{d}{=} cX + d$ and we have $d = 0$. □

Now we come onto an important result, and will be key when deriving a closed form for the characteristic function of a general stable distribution. We start by proving for symmetric stable distributions, and hence strictly stable by the previous lemma. Constant random variables are trivially stable, so we assume $X$ is non constant.

**Theorem 3.1.7.** Let $X$ be a non constant symmetric random variable satisfying

$$\sum_{i=1}^{n} X_i \stackrel{d}{=} c_n X$$

Then there exists $\alpha \in (0, 2]$ and $k > 0$ such that $c_n = n^{\frac{1}{\alpha}}$ and the characteristic function of $X$ is given by

$$\varphi(y) = e^{-k|y|^\alpha}$$

We will eventually derive a much more detailed form for the characteristic function, but this form is useful for proving the theorem. The main result here is the form for $c_n$.

*Proof.* [17, Ch 3.1] In terms of its characteristic function,

$$\varphi(y)^n = \varphi(c_n y) \tag{6}$$

and replacing $y$ with $y/c_n$ gives

$$\varphi(y/c_n)^n = \varphi(y) \tag{7}$$

If $c_n = n^{\frac{1}{\alpha}}$, and $\varphi(y) = e^{-k|y|^\alpha}$, then

$$\varphi(y)^n = \left(e^{-k|y|^\alpha}\right)^n = e^{-kn|y|^\alpha} = e^{-k|n^{\frac{1}{\alpha}}y|^\alpha} = \varphi(n^{\frac{1}{\alpha}}y) = \varphi(c_n y)$$

So it remains to show that this is the only solution for which $\phi$ is a characteristic function.

Define $r_n := \log_n(c_n)$ which is allowed since by definition $c_n > 0$. Suppose for a contradiction that there exist integers $m, n > 1$ such that $r_n \neq r_m$. Without loss of generality let $r_m < r_n$. Then $\frac{r_m}{r_n} < 1$ and by density of rational numbers we can find integers $a, b$ such that

$$\frac{r_m}{r_n} < \frac{a}{b}\log_m(n) < 1$$

Raising $m$ to the power of each term gives

$$m^{\frac{r_m}{r_n}} < n^{\frac{a}{b}} < m$$

So

$$m^{br_m} < n^{ar_n} < m^{br_n}$$

The last inequality gives $n^a < m^b$, and recalling Corollary 3.1.5 we have

$$c_{n^a} = (c_n)^a = n^{ar_n} > m^{br_m} = (c_m)^b = c_{m^b}$$

Set $p_k := n^{ak}$ and $q_k := m^{bk}$. Then we have integer sequences with $p_k < q_k$ for every $k$ and

$$\frac{c_{q_k}}{c_{p_k}} = \left(\frac{c_{m^b}}{c_{n^a}}\right)^k \to 0$$

Now as $X$ is symmetric, its characteristic function $\varphi$ is a real valued function, with $\varphi(0) = 1$ and $|\varphi(y)| \leq 1$ for all $y \in \mathbb{R}$. Since $X$ is non constant and continuous, we can find a point $y^* > 0$ with $0 < \varphi(y^*) < 1$.

**Claim:** $\varphi(y) > 0$ for all $y$

*Proof.* Assume there exists $y_0$ with $\varphi(y_0) = 0$. Then

$$\varphi(c_2 y_0) = \varphi(y_0)^2 = 0$$

$$\varphi(y_0/c_2)^2 = \varphi(y_0) = 0$$

so $\varphi(y_0/c_2) = 0$. Notice $c_2$ cannot equal 1 as otherwise $\varphi(y)^2 = \varphi(y)$ for all $y$, and since $\varphi$ is continuous with $\varphi(0) = 1$ this would mean $\varphi = 1$, but this contradicts $X$ being non constant. So set

$$y_1 = \begin{cases} y_0 c_2 & c_2 < 1 \\ y_0/c_2 & c_2 > 1 \end{cases}$$

Then we can continue this to obtain a sequence $y_n$ with $y_n \to 0$ and $\varphi(y_n) = 0$ for every $n$. But this implies $\varphi(0) = 0$ as $\varphi$ is continuous - a contradiction. Hence $\varphi(y) \neq 0$ for all $y$ and since $\varphi$ is continuous with $\varphi(0) = 1$, $\varphi(y) > 0$ for all $y$ □

Let $\psi(y) := \log \varphi(y)$ which is well defined and continuous by the claim. Then by taking logarithms of (6) and (7), for every $n \geq 2$, we have $n\psi(y) = \psi(c_n u)$ and $n\psi(u/c_n) = \psi(u)$.
In particular, if $p, q$ are integers greater than or equal to 2, and $y \in \mathbb{R}$ then

$$\psi\left(\frac{c_q}{c_p} y\right) = q\psi\left(\frac{1}{c_p} y\right) = \frac{q}{p}\psi(u) \tag{8}$$

So, putting everything together, we have

$$\psi\left(\frac{c_{q_n}}{c_{p_n}} y^*\right) = \frac{q_n}{p_n}\psi(y^*)$$

But this is impossible as letting $n \to \infty$, by continuity of $\psi$ the LHS tends to $\psi(0) = 0$, but since

$\varphi(y^*) < 1$, $\psi(y^*) < 0$ and thus the RHS goes to $-\infty$. Hence $r_n = r_m$ for all integers $m, n$, say $r_n = r_m = r$. Then by definition $c_n = n^r$ for every $n$. So we can simplify equation (8) to

$$\psi\left(\left(\frac{q}{p}\right)^r y^*\right) = \frac{q}{p}\psi(y^*)$$

Define $\psi^*(y) := \psi(y^*)\left(\frac{y}{y^*}\right)^{\frac{1}{r}}$ for $y > 0$. Then $\psi^*$ is continuous and for all $p, q > 1$ and $\psi^*\left(\left(\frac{p}{q}\right)^r y^*\right) = \frac{p}{q}\psi(y^*)$. Therefore $\psi$ and $\psi^*$ are continuous functions which agree on a dense set, so they are equal. So if $\alpha = \frac{1}{r}$ and $k = -\frac{\psi(y^*)}{(y^*)^\alpha} > 0$, we have $\psi(y) = \psi^*(y) = -ky^\alpha$ and by symmetry of $X$ $\psi(-y) = \psi(y) = \psi(|y|)$ so $\psi(y) = -k|y|^\alpha$. Therefore $\varphi(y) = \exp(\psi(y)) = \exp(-k|y|^\alpha)$.

The last thing to prove is $0 < \alpha \leq 2$. If $\alpha < 0$ then $\exp(-k|u|^\alpha) \to 0$ as $u \to 0$, and if $\alpha = 0$ then the limit is $\exp(-k) \neq 1$ since $k > 0$. So if $\alpha \leq 0$ then $\varphi(y)$ is not a characteristic function. If $\alpha > 2$ then $\varphi$ is twice differentiable at 0, and $\phi''(0) = 0$. But then this implies $\mathrm{Var}(X) = 0$ contradicting the fact $X$ is non constant. $\square$

It is not difficult to extend this result to non-symmetric random variables using the trick that if $X_1$ and $X_2$ follow the same distribution, $X_1 - X_2$ must be symmetric.

**Corollary 3.1.8.** If $X$ is a stable random variable, such that $X_1 + X_2 + \cdots + X_n \stackrel{d}{=} c_n X + d_n$, then we have $c_n = n^{\frac{1}{\alpha}}$ for some $0 < \alpha \leq 2$.

*Proof.* Define $Y := X_1 - X_2$. Then $Y$ is symmetric, and if $Y_1, Y_2, \ldots$ are i.i.d copies of $Y$,

$$\begin{aligned} Y_1 + Y_2 + \cdots + Y_n &\stackrel{d}{=} (X_1 + X_2 + \cdots + X_n) - (X_{n+1} + X_{n+2} + \cdots + X_{2n} \\ &\stackrel{d}{=} (c_n X_1 + d_n) - (c_n X_2 + d_n) \\ &\stackrel{d}{=} c_n(X_1 - x_2) \stackrel{d}{=} c_n Y \end{aligned}$$

Hence $Y$ is symmetric and strictly stable, so applying the previous lemma gives $c_n = n^{\frac{1}{\alpha}}$ as required. $\square$

We can now define explicitly the Lévy triplet for any stable distribution.

**Theorem 3.1.9.** If $X$ is a nondegenerate stable random variable with index of stability $\alpha$, then if $\alpha = 2$ $X$ has Lévy triplet $(\gamma, \sigma, 0)$ and if $0 < \alpha < 2$, the triplet is given by $(\gamma, 0, F)$ where

$$F(\mathrm{d}z) = \left(\frac{\beta^+}{z^{\alpha+1}}\mathbb{1}_{(0,\infty)} - \frac{\beta^-}{|z|^{\alpha+1}}\mathbb{1}_{(-\infty,0)}\right)\mathrm{d}z$$

for some positive constants $\gamma, \beta^+, \beta^- > 0$.

*Proof.* Using what we have just proven and the infinite divisibility properties, if $X$ is stable then $nX$ is equal in distribution to $n^{1/\alpha}X + d_n$. So the respective characteristic functions must be equal and we have

$$n \log \varphi_X(y) = \log \varphi_X(n^{1/\alpha}y) + iyd_n$$

Using the Lévy -Khintchine representation, if $X$ has Lévy triplet $(\gamma, \sigma, F)$, we can write

$$\begin{aligned} iny\gamma - \frac{n}{2}y^2\sigma^2 + \int_{\mathbb{R}} [e^{iyz} - 1 - iyz\mathbb{1}_{|z|<1}]nF(dz) \\ = iy(\gamma + d_n) - \frac{1}{2}y^2\sigma^2 + \int_{\mathbb{R}} [e^{iyz} - 1 - iyz\mathbb{1}_{|z|<1}]F(\frac{dz}{n^{1/\alpha}}) \end{aligned} \tag{9}$$

Using (9), and letting $\chi_x : \mathbb{R} \to \mathbb{R}$ by multiplication by $x$, we derive

$$\frac{1}{2}ny^2\sigma^2 = \frac{1}{2}y^2n^{\frac{2}{\alpha}}\sigma^2$$

$$nF = F \circ \chi^{-1}_{n^{1/\alpha}}$$

Comparing the $\sigma$ terms, either $\alpha = 2$ or $\sigma = 0$. Then considering $F$, for any $z > 0$ we have

$$nF[zn^{1/\alpha}, \infty) = F\,[z, \infty)$$

and taking $z = 1$ we get

$$F[n^{1/\alpha}, \infty) = n^{-1}F\,[1, \infty)$$

Substituting $m = n^{1/\alpha}$ gives

$$F[m, \infty) = m^{-\alpha}F[1, \infty)$$

and hence

$$F(-\infty, m) = \int_{\mathbb{R}\setminus\{0\}} F(\mathrm{d}z) - m^{-\alpha}F[1, \infty)$$

where the integral is finite by definition of the Lévy measure. We now differentiate with respect to $m$ to retrieve the density function

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}m}\Big[F(-\infty, m)\Big] &= \frac{\mathrm{d}}{\mathrm{d}m}\Big[\int_{\mathbb{R}} F(\mathrm{d}z) - m^{-\alpha}F[1, \infty)\Big] \\ &= \frac{\alpha}{m^{\alpha+1}}F[1, \infty) \end{aligned}$$

So on $\mathbb{R}_+$, we have $F(\mathrm{d}z) = \frac{\beta_+}{x^{\alpha+1}}\mathrm{d}z$, where $\beta_+ = \alpha F[1, \infty) \geq 0$. Identical calculations give the form for the density over $\mathbb{R}_-$.

Finally notice that if $F \neq 0$ then the integral

$$\int_{\mathbb{R}\setminus\{0\}} \min(1, z^2)F(\mathrm{d}z) = \int_{\mathbb{R}\setminus\{0\}} \min(1, z^2)\left(\frac{\beta^+}{z^{\alpha+1}}\mathbb{1}_{(0,\infty)} - \frac{\beta^-}{1 + z^{\alpha+1}}\mathbb{1}_{(-\infty,0)}\right)\mathrm{d}z$$

is convergent only if $0 < \alpha < 2$. So if $\alpha = 2$, we must have $F = 0$ as required.

□

**Corollary 3.1.10.** If $X$ is a stable distribution with $\alpha = 2$, then $X$ has a Gaussian distribution with mean $\gamma$ and variance $\sigma^2$.

*Proof.* Using the previous theorem, the characteristic function of $X$ is

$$\varphi_X(y) = \exp\left\{iy\gamma + \frac{1}{2}\sigma^2 y^2\right\}$$

which is that of the required Gaussian distribution. □

As properties of the Gaussian distribution are well known, from now on we take the phrase "$X$ follows a stable distribution" to implicitly mean that $\alpha \neq 2$. So, bringing everything together we have another equivalent distribution for a stable distribution.

**Definition 3.1.11.** If $X$ is a non-degenerate random variable, we say $X$ follows a stable distribution with parameters $\alpha \in (0, 2)$, $\beta^+, \beta^- > 0$ and $\gamma \in \mathbb{R}$ if the characteristic function is given by

$$\varphi_X(y) = iy\gamma + \int_{\mathbb{R}\setminus\{0\}} [e^{iyz} - 1 - iyz\mathbb{1}_{|z|<1}]F(\mathrm{d}z) \tag{10}$$

where

$$F(\mathrm{d}z) = \left(\frac{\beta^+}{z^{\alpha+1}}\mathbb{1}_{(0,\infty)} - \frac{\beta^-}{|z|^{\alpha+1}}\mathbb{1}_{(-\infty,0)}\right)\mathrm{d}z$$

Note the (arbitrary) choice of $h(z) = \mathbb{1}_{|z|<1}$. We can also define a one-sided stable distribution.

**Definition 3.1.12.** If $X$ is a non-degenerate random variable, we say $X$ follows a *totally positively skewed* stable distribution with parameters $\alpha \in (0, 2), \beta > 0$ and $\gamma \in \mathbb{R}$ if the characteristic function is given by

$$\varphi_X(y) = iy\gamma + \int_{\mathbb{R}\setminus\{0\}} [e^{iyz} - 1 - iyz\mathbb{1}_{|z|<1}]F(\mathrm{d}z) \tag{11}$$

where

$$F(\mathrm{d}z) = \frac{\beta^+}{z^{\alpha+1}}\mathbb{1}_{(0,\infty)}\mathrm{d}z$$

Totally negatively skewed stable variates are defined analogously. It is clear to see that a bilateral stable random variate is equal in distribution to the sum of a totally positively and a totally negatively skewed stable random variable. For this reason it is often useful for simulation purposes to only consider one sided distributions, and then simulate twice and take the difference.

**Remark 3.1.13.** Some specific instances of stable distributions have particularly convenient calculations which often greatly simplify the form for the characteristic function. The case where $\alpha = 2$ is the most obvious, as this is simply a Gaussian distribution. Other examples include the Cauchy distribution when $\alpha = 1, \beta^+ = \beta^-$, and the Lévy distribution when $\alpha = 1/2, \beta^+ = \pm\beta^-$.

I will write $X \sim S(\alpha, \beta^+, \beta^-, \gamma)$ to mean that $X$ follows the distribution governed by (10), and $X \sim S_+(\alpha, \beta, \gamma)$ to represent $X$ has distribution governed by (11).

**Remark 3.1.14.** The integral in (10) can be explicitly evaluated (see [13, Section 1.2.6 and Ex 1.4]) to obtain

$$-\log \varphi(y) = c|y|^\alpha \left[1 - i\tilde{\beta} \tan \frac{\pi\alpha}{2} \mathrm{sgn}(y)\right] - iya$$

where $\tilde{\beta} = \dfrac{\beta^+ - \beta^-}{\beta^+ + \beta^-} \in [-1, 1], \quad c = -\Gamma(-\alpha)\cos(\pi\alpha/2)(\beta^+ + \beta^-)$ and $a \in \mathbb{R}$ depends on $\alpha$ such that

$$a = \begin{cases} \gamma + \displaystyle\int_{\mathbb{R}} z \mathbb{1}_{|z|<1} F(\mathrm{d}z) & \delta \in (0, 1) \\ \gamma - \displaystyle\int_{\mathbb{R}} z \mathbb{1}_{|z|\geq 1} F(\mathrm{d}z) & \delta \in (1, 2) \end{cases}$$

The constant $\delta$ in the case $\alpha = 1$ which results in this distribution exactly aligning with our definition and choice of truncation function is complicated to calculate and hence excluded for brevity. The evaluation is still a stable distribution, but represents a shifted version of our definition.

This representation is often used as the definition for stable processes and then the other properties are derived from this formula. However it is often difficult to gain intuition about the behaviour of the process from this expression, so we will primarily use the Lévy - Khintchine representation.

It is now of interest to study the tails of a given stable distribution. As seen in the previous section, if $X \sim S_+(\alpha, \beta, \gamma)$ and $\varepsilon > 0$, $X$ has finite variation if and only if $\int_{|z|<\varepsilon} zF(\mathrm{d}z) < \infty$. Using our knowledge of the Lévy measure of this distribution,

$$\int_{|z|<\varepsilon} zF(\mathrm{d}z) = \int_0^\varepsilon z \frac{\beta}{z^{\alpha+1}} \mathrm{d}z$$

and this behaves like $\int_0^\varepsilon z^{-\alpha}$ near 0, so is finite if and only if $\alpha \leq 1$. Hence $X$ has finite variation if and only if $\alpha \leq 1$. In this case, we often use truncation function $h(z) = 0$ and say the characteristic function is given by

$$\varphi_X(y) = iy\gamma + \int_{\mathbb{R}\backslash\{0\}} [e^{iyz} - 1]F(\mathrm{d}z)$$

This also tells us that if $\alpha > 1$, the expectation of $X$ is undefined. Using Lemma 2.2.9, we calculate, for $p > 0$

$$\int_{|z|\geq 1} |z|^p F(\mathrm{d}z) = \int_1^\infty z^p \frac{\beta}{z^{\alpha+1}} \mathrm{d}z$$

which behaves like $z^{p-\alpha+1}$ and is hence finite if and only if $p < \alpha$. So we have that the only finite moments of $X$ are those up to $\alpha$. Since a bilateral stable distribution is the difference of 2 one sided distributions, we can extend this to $X \sim S(\alpha, \beta^+, \beta^-, \gamma)$ by the following lemma

**Lemma 3.1.15.** If $X \sim S(\alpha, \beta^+, \beta^-, \gamma)$, where $\alpha \in (0, 2)$, $\beta^+, \beta^- > 0$ and $\gamma \in \mathbb{R}$, then $\mathbb{E}[X^p] < \infty$ if and only if $p < \alpha$.

This result demonstrates just how heavy tailed stable distributions are, and provides the motivation for the rest of this dissertation. In many financial contexts this is undesirable, for example pricing options where the underlying asset price is modelled by a Stable Lévy process can often lead to infinite prices. So the problem now is finding a way to lighten the tails of these distributions whilst retaining the intuition of stable property.

## 3.2 Tempered Stable Distributions

Having seen how heavy tailed stable distributions can be, and the issues this can cause, it becomes of interest to define a class of distributions which encapsulate much of the underlying properties of stable distributions, but have finite moments. One such distribution class are *tempered stable* distributions, first introduced by Kopenen in 1995 [11]. To lighten the tails of a stable distribution, we aim modify the Lévy measure such that the integral $\int_{|z|\geq 1} |z|^p F(\mathrm{d}z)$ is finite for all $p$, and hence by Lemma 2.2.9 all moments are finite. Recalling the discussion at the end of Section 2, we can achieve this by transforming the underlying measure space. Firstly I define the target distribution, then we consider firstly if a change of measure is possible to achieve this, and if so what does the new measure looks like and what are the properties of the distribution in this new transformed space? Again I begin by simplifying things slightly and only considering a one sided stable distribution. We know the Lévy measure of a $S_+(\alpha, \beta, \gamma)$ distribution is given by

$$F(\mathrm{d}z) = \frac{\beta}{z^{\alpha+1}} \mathbb{1}_{(0,\infty)} \mathrm{d}z$$

Ideally, we would like to define a measure $\tilde{F}$ such that $\int_{|z|\geq 1} |z|^p \tilde{F}(\mathrm{d}z) < \infty$. Using Theorem 2.3.3, we also need $\tilde{F}$ to satisfy $\frac{\mathrm{d}\tilde{F}}{\mathrm{d}F} = y$ where $y$ is a non negative function $\mathbb{R} \to \mathbb{R}$ satisfying

$$\int_{\mathbb{R}_+} |h(z)(1 - y(z))| F(\mathrm{d}z) < \infty \tag{12}$$

$$\int_{\mathbb{R}_+} \left(1 - \sqrt{y(z)}\right)^2 F(\mathrm{d}z) < \infty \tag{13}$$

Since $\sigma = 0$ for stable distributions, the existence of the constant $a$ in Theorem 2.3.3 is trivially satisfied for any $a \in \mathbb{R}$. There are many possible functions $y$ which satisfy (12) and (13), and this function is known as the *tilting* or *tempering* function. For example various polynomials are possible but we will focus solely on the choice $y(z) = e^{-\lambda z}$ for some $\lambda > 0$.
These integrability conditions depend on $h$, but due to the rapid decay of $e^{- \ z}$, they will be satisfied for a large class of functions $h$. The conditions on $h$ are also loosened with this exponential factor; in particular we can drop the need for $h(z) = O(1)$ as $z \to \infty$. It will be most convenient for us to use $h(z) = z$, which does not truncate big jumps, and we will show later that this choice of $h$ ensures the mean of the variable is equal to the drift $\gamma$.

It remains to show that we have actually achieved our goal, and prove that under this new measure, all moments of $X_t$ are finite. Using Lemma 2.2.9, for any $p > 0$,

$$\begin{aligned}\int_{|z|\geq 1} |z|^p \tilde{F}(\mathrm{d}z) &= \int_{|z|\geq 1} |z|^p e^{-\lambda z} F(\mathrm{d}z) \\ &= \int_{|z|\geq 1} |z|^p e^{-\lambda z} \frac{\beta}{z^{\alpha+1}} \mathrm{d}z < \infty\end{aligned}$$

The $e^{-\lambda z}$ term will dominate as $z \to \infty$ so we have that the integral is finite for all $p$ and hence every moment is defined. We call this new class of distributions *tempered stable* distributions, and in this section we will explore their properties. They are formally defined as follows.

**Definition 3.2.1.** Let $\alpha \in (0, 2)$ $\beta, \lambda > 0$, and $\gamma \in \mathbb{R}$. Then if $\eta$ is an infinitely divisible distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with characteristic function

$$\varphi(y) = \exp\left\{iy\gamma + \int_{\mathbb{R}} (e^{iyz} - 1 - iyz)\tilde{F}(\mathrm{d}z)\right\}$$

where the Levy measure $\tilde{F}$ is given by

$$\tilde{F}(\mathrm{d}z) = \frac{\beta}{z^{1+\alpha}} e^{-\lambda z} \mathbb{1}_{(0,\infty)}(z)\mathrm{d}z$$

we say that $\eta$ is a one sided, or totally positively skewed, tempered stable distribution with parameters $\alpha, \beta, \lambda, \gamma$, which I will denote by $\eta = \text{TS}_+(\alpha, \beta, \lambda, \gamma)$.

Note the choice of $h(z) = z$, as discussed at the start of this section. I use $\tilde{F}$ for the Lévy measure as a reminder that this is a modified version of the Lévy measure for the corresponding stable process. The parameter $\lambda$ is known as the tempering parameter, and controls the speed at which the tail decays. Clearly $\lambda = 0$ corresponds to the associated stable distribution with no tempering. The closer the value of $\lambda$ to 0, the more the similarity to the corresponding stable distribution, and for large values of $\lambda$ the exponential term takes over much faster, so we resemble a Gaussian distribution. We can now define a general bilateral tempered stable distribution, which has jumps of both positive and negative size.

**Definition 3.2.2.** Let $\alpha- \in (0,2)$, $\beta^+, \beta^-, \lambda^-, \lambda^+ \in (0, \infty)$, and define $\eta^+, \eta^-$ as one sided tempered stable distributions with parameters $\eta_+ = \text{TS}_+(\alpha^+, \beta^+, \lambda^+)$, $\eta_- = \text{TS}_+(\alpha^-, \beta^-, \lambda^-)$.

Then $\eta = \eta_+ * \overline{\eta_-}$, where $\overline{\eta}$ denotes the dual of $\eta$, defined by $\overline{\eta}(A) = -\eta(A)$, is a bilateral tempered stable distribution with parameters $\alpha, \beta^+, \lambda^+, \beta^-, \lambda^-$, $LKshift$ denoted by $\eta = \text{TS}(\alpha, \beta^+, \lambda^+, \beta^-, \lambda^-, \gamma)$.

Using what we know about 1 sided tempered stable distributions, $\eta$ is an infinitely divisible distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with characteristic function

$$\varphi(y) = \exp\left\{iy\gamma + \int_{\mathbb{R}} (e^{iyz} - 1 - iyz)\tilde{F}(\mathrm{d}z)\right\}$$

and Levy measure

$$\tilde{F}(\mathrm{d}z) = \left[\frac{\beta^+}{z^{1+\alpha}} e^{-\lambda^+ z} \mathbb{1}_{(0,\infty)}(z) + \frac{\beta^-}{|z|^{1+\alpha}} e^{-\lambda^- z} \mathbb{1}_{(-\infty,0)}(z)\right] \mathrm{d}z$$

Where the distribution is symmetric, so $\beta^+ = \beta^-$ and $\lambda^+ = \lambda^-$, I will use the notation $\eta = \text{TS}(\alpha, \beta, \lambda, \gamma)$.

There are a few things to remark about our new class of distributions.

**Remark 3.2.3.** In Theorem 2.3.5, we have already seen the form for the new measure $\mathbb{Q}$; that $\mathbb{Q}$ must in fact be the Esscher transform of $\mathbb{P}$ with parameter $-\lambda$. So if $\mathbb{P}$ is the measure under which the random variable $X$ follows a stable distribution and $\mathbb{Q}$ is the tilted measure under which $X$ observes a tempered stable distribution with tempering parameter $\lambda > 0$, we have that

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\,\mathbb{P}} = \frac{e^{-\lambda x}}{\mathbb{E}^{\mathbb{P}}[e^{-\lambda X}]} \tag{14}$$

It is worth noting that due to our different choice of truncation function for tempered stable distributions, simply tilting the stable distribution as defined in Definition 3.1.1 will result in a shifted version of the tempered stable distribution defined in Definition 3.2.2. In particular, if the stable distribution has characteristic function

$$\varphi_S(y) = iy\gamma + \int_{\mathbb{R}\setminus\{0\}} [e^{iyz} - 1 - iyz\mathbb{1}_{|z|<1}]F(\mathrm{d}z)$$

then simply tilting, using Theorem 2.3.3 will result in a distribution with characteristic function

$$\varphi_{\mathrm{TS}'}(y) = iy(\gamma - p) + \int_{\mathbb{R}\setminus\{0\}} [e^{iyz} - 1 - iyz\mathbb{1}_{|z|<1}]\tilde{F}(\mathrm{d}z) \tag{15}$$

where

$$p = \int_{\mathbb{R}\setminus\{0\}} (1 - e^{\lambda z})zF(\mathrm{d}z)$$

So to exactly align with our definition of tempered stable distributions (using truncation function $h(z) = z$) an additional shift of

$$\xi = p - \int_1^\infty z\tilde{F}(\mathrm{d}z)$$

is required. For absolute clarity, I will use the notation $\mathrm{TS}'$ to denote a distribution with characteristic function as in (15), and TS to denote the distribution defined in Definition 3.2.2. We then have the final identity

$$f_{\mathrm{TS}}(x) = f_{\mathrm{TS}'}(x+\xi) = \frac{e^{-\lambda(x+\xi)}}{\mathbb{E}^{\mathbb{P}}[e^{-\lambda X}]} f_S(x+\xi) \tag{16}$$

where $f_\eta$ represents the density function for the distribution $\eta$.

**Remark 3.2.4.** This idea of tempering can be generalised for any monotone tempering function $q$, or into higher dimensions where the tempering function can be expressed as a measure around the unit ball. See [19] for results in this more general setting, but we will keep to this exponential, one dimensional definition.

**Remark 3.2.5.** It is also worth noting that since this tempering function does not alter the behaviour of the stable Lévy measure around 0, the distribution still is of finite variation if and only if $\alpha < 1$.

We will often only consider one sided tempered stable distributions, and when we come to simulation algorithms, it is straightforward to obtain a bilateral simulation algorithm from a one sided one by simply sampling twice from a one sided distribution and taking the difference. It will also be useful to have an easily differentiable form for the characteristic function, with the goal of using this representation to calculate explicitly the moments for a given tempered stable distribution, which we know now are all finite. For simplicity, throughout the rest of this section we exclude the case $\alpha = 1$.

**Lemma 3.2.6.** [14, Lemma 2.5 - Remark 2.8]
If $\alpha \in (0,1) \cup (1,2)$, then we can express

$$\int_0^\infty (e^{iyz} - 1 - iyz)\tilde{F}(\mathrm{d}z) = \beta\,\Gamma(-\alpha)\Big[(\lambda - iy)^\alpha - \lambda^\alpha + iy\alpha\lambda^{\alpha-1}\Big]$$

*Proof.* The proof makes use of power series. First assume $\alpha \neq 1$. We have

$$\begin{aligned}
\int_{\mathbb{R}} (e^{iyz} - 1 - iyz)\tilde{F}(\mathrm{d}z) &= \int_{\mathbb{R}} (e^{iyz} - 1 - iyz)\frac{\beta}{z^{1+\alpha}} e^{-\lambda z}\mathbb{1}_{(0,\infty)}(z)\mathrm{d}z \\
&= \beta \int_0^\infty \left(\left(\sum_{n=0}^\infty \frac{(iyz)^n}{n!}\right) - 1 - iyz\right)\frac{e^{-\lambda z}}{z^{1+\alpha}}\mathrm{d}z \\
&= \beta \int_0^\infty \left(\sum_{n=2}^\infty \frac{(iy)^n}{n!} z^{n-\alpha-1}e^{-\lambda z}\right)\mathrm{d}z \\
&= \beta \sum_{n=2}^\infty \left(\frac{(iy)^n}{n!}\int_0^\infty z^{n-\alpha-1}e^{-\lambda z}\mathrm{d}z\right) \\
&= \beta \sum_{n=2}^\infty \frac{(iy)^n}{n!}\lambda^{1+\alpha-n}\frac{1}{\lambda}\int_0^\infty w^{n-\alpha-1}e^{-w}\mathrm{d}w \\
&= \beta\lambda^\alpha \sum_{n=2}^\infty \frac{(iy/\lambda)^n}{n!}\,\Gamma(n-\alpha)
\end{aligned}$$

where the penultimate step came from a change of variables $w = \lambda z$. The change of order of summation and integration is permitted in the radius of convergence, where $|z| < \lambda$, but can be extended to the half-plane $\{\, z \in \mathbb{C} : \Im(z) > -\lambda \,\}$ using analytic continuation. In particular the formula holds for the whole real line. Now consider the well known series

$$(1-x)^p = 1 - px + p(p-1)\frac{x^2}{2!} - p(p-1)(p-2)\frac{x^3}{3!} + \dots$$

So

$$(1 - \frac{iy}{\lambda}))^\alpha = 1 - \alpha\frac{iy}{\lambda} + \alpha(\alpha-1)\frac{(iy/\lambda)^2}{2!} - \alpha(\alpha-1)(\alpha-2)\frac{(iy/\lambda)^3}{3!} + \dots$$

And hence

$$\begin{aligned}
\Gamma(-\alpha)\left[(1-\frac{iy}{\lambda}))^{\alpha}-1+\frac{iy\alpha}{\lambda}\right] &= \Gamma(-\alpha)(-\alpha)(1-\alpha)\frac{(iy/\lambda)^2}{2!}+\Gamma(-\alpha)(-\alpha)(1-\alpha)(2-\alpha)\frac{(iy/\lambda)^3}{3!}+\dots \\
&= \Gamma(2-\alpha)\frac{(iy/\lambda)^2}{2!}+\Gamma(3-\alpha)\frac{(iy/\lambda)^3}{3!}+\dots \\
&= \sum_{n=2}^{\infty}\frac{(iy/\lambda)^n}{n!}\Gamma(n-\alpha)
\end{aligned}$$

Plugging this in gives

$$\begin{aligned}
\int_{\mathbb{R}}(e^{iyz}-1-iyz)F(\mathrm{d}z) &= \beta\lambda^{\alpha}\,\Gamma(-\alpha)\left[(1-\frac{iy}{\lambda}))^{\alpha}-1+\frac{iy\alpha}{\lambda}\right] \\
&= \beta\,\Gamma(-\alpha)\left[(\lambda-iy)^{\alpha}-\lambda^{\alpha}+iy\alpha\lambda^{\alpha-1}\right]
\end{aligned}$$

For the $\alpha = 1$ case, the argument is almost identical, using the well known power series for $\log(1+x)$. □

So using this new representation, we can rewrite the characteristic function for a one sided tempered stable random variable in a way which is more friendly for differentiation.

**Theorem 3.2.7.** Let $X \sim \mathrm{TS}_{+}(\alpha, \beta, \lambda, \gamma)$. Then we can express

$$\log\varphi_X(y) = iy\gamma + \beta\,\Gamma(-\alpha)\left[(\lambda-iy)^{\alpha}-\lambda^{\alpha}+iy\alpha\lambda^{\alpha-1}\right]$$

*Proof.* This follows by simply substituting the result from the previous lemma into the definition of the tempered stable distribution (remembering we are excluding $\alpha = 1$.) □

Now an analytic form for the characteristic function is known, we can exploit the relation

$$\mathbb{E}[X^n] = \frac{1}{i^n}\frac{\mathrm{d}}{\mathrm{d}y}\varphi(y)\Big|_{y=0}$$

to determine the moments of the distribution. We start in the case $\alpha \neq 1$. Differentiating the formula for $\varphi$ we just derived, we have

$$\begin{aligned}
\varphi'(y) &= \varphi(y)\frac{\mathrm{d}}{\mathrm{d}y}\left(iy\gamma+\beta\,\Gamma(-\alpha)\left[(\lambda-iy)^{\alpha}-\lambda^{\alpha}+iy\alpha\lambda^{\alpha-1}\right]\right) \\
&= \varphi(y)\left(i\gamma+\beta\,\Gamma(-\alpha)\left[-i\alpha(\lambda-iy)^{\alpha-1}+i\alpha\lambda^{\alpha-1}\right]\right) \\
&= \varphi(y)\left(i\gamma+\beta i\,\Gamma(1-\alpha)\left[(\lambda-iy)^{\alpha-1}-\lambda^{\alpha-1}\right]\right)
\end{aligned}$$

Now $\varphi(0) = 1$ by definition of characteristic function, so $\varphi'(0) = i\gamma + \beta i\,\Gamma(1-\alpha)\left[(\lambda)^{\alpha-1} - \lambda^{\alpha-1}\right] = i\gamma$ and thus if $X$ follows a $\mathrm{TS}_+(\alpha, \beta, \lambda)$ distribution,

$$\mathbb{E}[X] = \gamma \tag{17}$$

We now differentiate again to find the second moment of $X$.

$$\begin{aligned}
\varphi''(y) &= \varphi'(y)\left(i\gamma + \beta i\,\Gamma(1-\alpha)\left[(\lambda - iy)^{\alpha-1} - \lambda^{\alpha-1}\right]\right) + \varphi(y)\beta i\,\Gamma(1-\alpha)\left[-i(\alpha-1)(\lambda - iy)^{\alpha-2}\right] \\
&= \varphi(y)\left[i\gamma + \beta i\,\Gamma(1-\alpha)\left[(\lambda - iy)^{\alpha-1} - \lambda^{\alpha-1}\right]\right]^2 - \varphi(y)\beta\,\Gamma(2-\alpha)(\lambda - iy)^{\alpha-2}
\end{aligned}$$

So $\varphi''(0) = (i\gamma)^2 - \beta\,\Gamma(2-\alpha)\lambda^{\alpha-2}$ and hence

$$E[X^2] = \frac{i^2\gamma^2 - \beta\,\Gamma(2-\alpha)\lambda^{\alpha-2}}{i^2} = \gamma^2 + \beta\,\Gamma(2-\alpha)\lambda^{\alpha-2} \tag{18}$$

We can see now why the choice $h(z) = z$ was particularly convenient, since then the mean of the process is exactly equal to the shift. These explicit formulas for the moments of tempered stable random variables will be used when evaluating the sampling algorithms presented in the next section. These formulae depend of course on the choice of $h$.

# 4 Sampling Algorithms for Tempered Stable Distributions

Efficient algorithms to sample from stable and tempered stable distributions are vital for these distributions to be of any practical use. However this is not always a straightforward task due to the inability to write down neither the density nor distribution function analytically. For stable random variables, this problem was effectively solved by Chambers, Mallow and Stuck who in 1976 proposed what is now often known as the CMS method for sampling from a stable distribution in [2]. However the problem remains for Tempered Stable distributions, and in this section I will present and analyse a proposed exact sampling algorithm. First, I introduce accept-reject sampling, which will be instrumental in our sampling algorithm.

## 4.1 Accept-Reject Sampling

I begin by introducing the accept-reject method. Accept-reject sampling is a well known technique where variables distributed according to the target distribution are sampled by exploiting a relation with another distribution for which an efficient sampling method is already known. For this section, assume we are trying to sample from a distribution $\nu$ with density function $f(x)$, and we already have an efficient algorithm for sampling from another distribution $\mu$ with density function $g(x)$. Assume also there exists a constant $d > 0$ such that

$$\sup_{x\in\mathbb{R}}\left[\frac{f(x)}{g(x)}\right] \leq d \tag{19}$$

Then we can sample from $\nu$ using the following algorithm [21]:

**Algorithm 1** Accept-Reject Sampling

1: Sample $Y$ from distribution $\mu$
2: Sample $U$ as Uniform[0,1]
3: **if** $U \leq \frac{f(Y)}{dg(Y)}$ **then** return Y ("accept")
4: **else** Go back to step 1 ("reject")
5: **end if**

*Proof of correctness.* We want to show that the distribution of $Y$ conditional on $U < f(Y)/dg(Y)$ is indeed $\nu$. So let $y \in \mathbb{R}$. Then using Bayes' formula,

$$\mathbb{P}\left(Y \leq y \,|\, U \leq \frac{f(Y)}{dg(Y)}\right) = \frac{\mathbb{P}(U \leq \frac{f(Y)}{dg(Y))} \,|\, Y \leq y)\,\mathbb{P}(Y \leq y)}{\mathbb{P}(U \leq \frac{f(Y)}{dg(Y)})} \tag{20}$$

Let $F$ and $G$ be the cumulative density functions for $\nu$ and $\mu$ respectively. Then noting that due to the bound $f(x) \leq dg(x)$ the ratio is always between 0 and 1, and remembering if $U$ is a uniform [0,1] random variable, for any $a \in [0,1]$, $\mathbb{P}(U < a) = a$,

$$\begin{aligned}
\mathbb{P}\left(U \leq \frac{f(Y)}{dg(Y))} \,|\, Y \leq y\right) &= \frac{\mathbb{P}\left(U \leq \frac{f(Y)}{dg(Y)}, Y \leq y\right)}{\mathbb{P}(Y \leq y)} \\
&= \frac{1}{G(y)}\int_{-\infty}^{y} \mathbb{P}\left(U \leq \frac{f(Y)}{dg(Y)} \,|\, Y = v\right) g(v)\mathrm{d}v \\
&= \frac{1}{G(y)}\int_{-\infty}^{y} \frac{f(v)}{dg(v)} g(v)\mathrm{d}v \\
&= \frac{1}{G(y)}\int_{-\infty}^{y} \frac{f(v)}{d}\mathrm{d}v = \frac{F(y)}{dG(y)}
\end{aligned}$$

Conditioning on $Y$, we have

$$\begin{aligned}
\mathbb{P}\left(U \leq \frac{f(Y)}{dg(Y)}\right) &= \int_{\mathbb{R}} \mathbb{P}\left(U \leq \frac{f(Y)}{dg(Y)} \,|\, Y = y\right) g(y)\mathrm{d}y \\
&= \frac{f(y)}{dg(y)} g(y)\mathrm{d}y = \frac{1}{d}\int_{\mathbb{R}} f(y)\mathrm{d}y = \frac{1}{d}
\end{aligned}$$

So plugging these into (20) we have

$$\mathbb{P}\left(Y \leq y \,|\, U \leq \frac{f(Y)}{dg(Y)}\right) = \frac{F(y)G(y)}{dG(y) \cdot 1/d} = F(y)$$

Thus the conditional distribution of $Y$ given $U < f(Y)/dg(Y)$ is indeed $\nu$, as required. □

Since this algorithm involves repeating a process until a suitable value is found, it is useful from a computational expense point of view to know the average number of samples considered before one is accepted. The probability a sample is accepted is precisely $\mathbb{P}\left(U \leq \frac{f(Y)}{dg(Y)}\right)$, which we computed in the proof to be $\frac{1}{d}$. Then the distribution of the number of times the algorithm runs follows a geometric distribution, with parameter $p = 1/d$. It is well known that this distribution has expected value of $1/p = d$. Thus the ideal bounding constant is 1, and the algorithm becomes more efficient the closer $d$ is to 1.

## 4.2 Accept-Reject scheme for Tempered Stable distributions

In this subsection, we show that it is possible to utilise the idea of accept-reject sampling to exactly simulate from tempered stable distributions for all $\alpha$. We consider only one sided distributions, as to simulate a bilateral distribution we can simply sample from two one-sided distributions and take the difference. Throughout the section, $f_S$ refers to the density of a $S_+(\alpha, \beta, 0)$ distribution as defined in Definition 3.1.12 (with $h(z) = \mathbb{1}_{|z|<1}$), $f_{\mathrm{TS}}$ refers to the density function for a $\mathrm{TS}_+(\alpha, \beta, \lambda, 0)$ random variable (with $h(z) = z$) and $f_{\mathrm{TS}'}$ refers to the density function for the distribution obtained by directly tilting $S_+(\alpha, \beta, 0)$, with characteristic function given by (15). Recall that $\mathrm{TS}'$ is the same distribution as TS, just shifted by the centring term $\xi$, defined in (16). I use $\varphi_\eta$ to denote the characteristic function of the distribution $\eta$.

It is possible to utilise a method of Devroye [3] to employ accept-reject techniques to simulate a random variable directly from its characteristic function. This would be particularly applicable to our cause of simulation of tempered stable variates, as their characteristic functions are well known. We begin by constructing an integrable function $g$ which dominates $f_{\mathrm{TS}}$. Remember from Subsection 3.2 that $\varphi_{\mathrm{TS}}$ is twice differentiable.

**Lemma 4.2.1.** Define

$$c = \frac{1}{2\pi} \int_{\mathbb{R}} |\varphi_{\mathrm{TS}}(z)| \mathrm{d}z \tag{21}$$

$$k = \frac{1}{2\pi} \int_{\mathbb{R}} |\varphi''_{\mathrm{TS}}(z)| \mathrm{d}z \tag{22}$$

Then

$$f_{\mathrm{TS}}(x) \leq \min\left\{c, \frac{k}{x^2}\right\}$$

*Proof.* [3] We know that

$$f_{\mathrm{TS}}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{ixz}\varphi_{\mathrm{TS}}(z)\mathrm{d}z$$

Using integration by parts,

$$\begin{aligned} f_{\mathrm{TS}}(x) &= \left[-\frac{e^{ixz}\varphi_{\mathrm{TS}}(z)}{2\pi ix}\right]_{-\infty}^{\infty} + \frac{1}{2\pi ix}\int_{-\infty}^{\infty} e^{-ixz}\varphi'_{\mathrm{TS}}(z)\mathrm{d}z \\ &= \frac{1}{2\pi ix}\int_{-\infty}^{\infty} e^{-ixz}\varphi'_{\mathrm{TS}}(z)\mathrm{d}z \end{aligned}$$

Doing the same thing again yields

$$f_{\mathrm{TS}}(x) = -\frac{1}{2\pi x^2}\int_{-\infty}^{\infty} e^{-ixz}\varphi''_{\mathrm{TS}}(z)\mathrm{d}z \tag{23}$$

Taking absolute values then gives that $f_{\mathrm{TS}}$ is bounded above by both functions, and hence must be bounded above by the minimum. □

So let $g(x) = \min\{c, k/x^2\}$. Noting that $c < k/x^2$ whenever $-\sqrt{k/c} < x < \sqrt{k/c}$, we can integrate $g$ piecewise to get $A := \int_{-\infty}^{\infty} g(z)\mathrm{d}z = 4\sqrt{kc}$. So we can normalise $g$ to get a probability distribution $A^{-1}g$. If we can find a way to sample from this distribution, then in theory we can construct an acceptance-rejection scheme to sample from $f_{\mathrm{TS}}$. The next lemma gives exactly that.

**Lemma 4.2.2.** Let $V_1, V_2$, be i.i.d uniform $(-1, 1)$ random variables. Then if $k, c$ are arbitrary positive numbers, the random variable

$$V := \frac{V_1}{V_2}\sqrt{\frac{k}{c}}$$

has density $A^{-1}g$, where $A$ and $g$ are as defined earlier.

*Proof.* We start by deriving the density function for $W = \frac{V_1}{V_2}$.

$$\begin{aligned} f_W(x) &= \int_{\mathbb{R}} |z| f_{V_1}(z) f_{V_2}(xz)\mathrm{d}z \\ &= \int_{\mathbb{R}} |z|\frac{1}{2}\mathbb{1}_{\{z\in[-1,1]\}} \cdot \frac{1}{2}\mathbb{1}_{\{xz\in[-1,1]\}}\mathrm{d}z \\ &= \frac{1}{4}\int_{\mathbb{R}} |z|\mathbb{1}_{\{z\in[-1,1]\}} \cdot \mathbb{1}_{\{xz\in[-1,1]\}}\mathrm{d}z \end{aligned}$$

Considering the 2 indicator functions, we have 2 cases. If $|x| \leq 1$, then the set $\{\, z : xz \in [-1, 1] \,\}$

contains the set $\{\, z : z \in [-1,1] \,\}$, so the second indicator disappears. So,

$$\begin{aligned} f_W(x) &= \frac{1}{4}\int_{\mathbb{R}} |z| \mathbb{1}_{\{z\in[-1,1]\}} \mathrm{d}z \\ &= \frac{1}{4}\int_{-1}^{1} |z| \mathrm{d}z = \frac{1}{4} \end{aligned}$$

The second case is where $|x| > 1$. In this case, the larger set is $\{\, z : |xz| \in [-1,1] \,\}$. In this case,

$$\begin{aligned} f_W(x) &= \frac{1}{4}\int_{\mathbb{R}} |z| \mathbb{1}_{\{xz\in[-1,1]\}} \mathrm{d}z \\ &= \frac{1}{4}\int_{-1/|x|}^{1/|x|} |z| \mathrm{d}z = \frac{1}{4|x|^2} \end{aligned}$$

Then remembering that if $k \in \mathbb{R}$, and $X$ is a random variable, then $f_{kX}(z) = \frac{1}{k} f_X\left(\frac{z}{k}\right)$, we have

$$\begin{aligned} f_V(x) &= \sqrt{\frac{c}{k}} \cdot f_W\left(\sqrt{\frac{c}{k}} z\right) \\ &= \begin{cases} \frac{1}{4\sqrt{k/c}} & |x| \leq \sqrt{\frac{k}{c}} \\ \frac{1}{4\sqrt{k/c}} \cdot \frac{c}{kx^2} & |x| > \sqrt{\frac{k}{c}} \end{cases} \\ &= \begin{cases} \frac{1}{4\sqrt{kc}} c & |x| \leq \sqrt{\frac{k}{c}} \\ \frac{1}{4\sqrt{kc}} \frac{1}{kx^2} & |x| > \sqrt{\frac{k}{c}} \end{cases} \\ &= \frac{1}{4\sqrt{kc}} \min\left(c, \frac{1}{kx^2}\right) = A^{-1} g(x) \end{aligned}$$

□

So we can derive an algorithm to sample from a tempered stable distribution [10, Alg. 1].

**Algorithm 2** Accept-Reject algorithm for TS random variables

1: Generate $V_1, V_2$ as independent Uniform $(-1,1)$ random variables, and $U$ as Uniform$[0,1]$
2: Set $V \leftarrow \sqrt{k/c} \cdot (V_1/V_2)$
3: **if** $|V| < \sqrt{k/c}$ **and** $cU < f_{\mathrm{TS}}(V)$ **then return** $V$ ("accept")
4: **else** Go back to step 1 ("reject")
5: **end if**
6: **if** $|V| \geq \sqrt{k/c}$ **and** $kU/V^2 < f_{\mathrm{TS}}(V)$ **then return** $V$ ("accept")
7: **else** Go back to step 1 ("reject")
8: **end if**

**Remark 4.2.3.** This algorithm will work theoretically for all values of $a \in (0,2)$, including $\alpha = 1$, however I will focus on the case where $\alpha \in (1,2)$. This is because many exact simulation algorithms already exist for $\alpha \in (0,1)$, and these have been widely studied (see [4] for a few examples). The case $\alpha = 1$ has been excluded for convenience. The algorithm in theory works perfectly well in this case, but as mentioned earlier conversion between the various parametrisations is cumbersome and requires calculation of complicated constants. There is also a separate form for the expectation (also given in [9]), and it is not possible to use the convenient representation derived in Lemma 3.2.6 to easily work with and differentiate $\varphi_{\text{TS}}$ to calculate $k$ and $c$. Hence, to keep the dissertation concise, this case is omitted.

Immediately we encounter an issue: how do we evaluate $f_{\text{TS}}$ when there is no known formula? There are a few workarounds, for example we can numerically perform the inverse Fourier transform to extract the density function directly from the characteristic function. However this is computationally expensive and relies on numerical integration which can lose accuracy near the tails of the distribution. So another possible solution is to use the change of measure discussions from previous sections. Recalling (16), and the exact form for the Laplace transform for a one sided stable distribution, we can efficiently evaluate $f_{\text{TS}}$ by evaluating

$$f_{\text{TS}}(x) = \frac{e^{-\lambda(x+\xi)}}{\mathbb{E}[e^{-\lambda S}]} f_S(x+\xi)$$

The expectation term, also known as the Laplace transform, has a well known exact form for a one sided stable distributions with $\alpha \neq 1$, namely

$$\mathbb{E}[e^{-\lambda S}] = \exp\left\{\beta\,\Gamma(-\alpha)\lambda^\alpha - \delta\lambda\right\}$$

where

$$\delta = \begin{cases} -\displaystyle\int_0^1 z\frac{\beta}{z^{\alpha+1}}\mathrm{d}z & \alpha \in (0,1) \\ \displaystyle\int_1^\infty z\frac{\beta}{z^{\alpha+1}}\mathrm{d}z & \alpha \in (1,2) \end{cases}$$

For a proof and discussion of this result see [9, Thm 3.13], remembering how to convert between parametrisations using Remark 3.1.14.

Substituting this in, we obtain the following formula to evaluate $f_{\mathrm{TS}}$.

$$f_{\mathrm{TS}}(x) = e^{-\lambda(x+\xi)-\beta\,\Gamma(-\alpha)\lambda^{\alpha}-\delta\lambda} f_S(x+\xi)$$

I have plotted a few example pdfs using this technique in Figure 1.

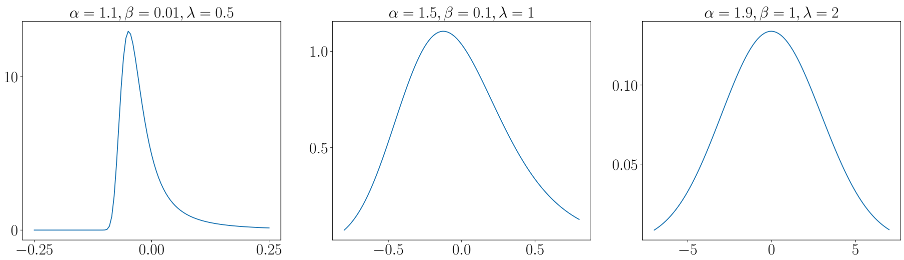


Figure 1: PDFs for Tempered Stable Distributions with varying parameters

Note this is an efficient method for evaluating the density function of a Tempered Stable distribution due the availability of highly efficient ways to evaluate $f_S$. In particular, this is orders of magnitude faster than simply calculating the inverse Fourier transform of the characteristic function via numerical integration. It is worth mentioning that this method does still rely on numerical integration to evaluate the constants $\delta$ and $\xi$, but these are shift constants chosen such that the various parametrisations agree with each other. If it is not necessary for the distribution to be centred, or for the underlying stable distribution to use the truncation function $\mathbb{1}_{|z|<1}$ then these constants can be omitted. I leave them in for continuity through the dissertation.

It is important for us now to analyse both the correctness and efficiency of this algorithm. I begin by analysing correctness, to verify the algorithm works across varying parameters. I will do this by sampling $N$ random values $X_1, X_2, \dots, X_N$ using Algorithm 4.2, then calculate an approximation of the $n$th moment using the formula

$$\mathbb{E}[X^n] \approx \frac{1}{N}\sum_{k=1}^{N} X_k^n$$

and then compare this to the analytical moments found earlier using differentiation of the characteristic function, and record the absolute error. In the following analysis I used $N = 100,000$. Figure 2 indicates that the algorithm performs well for values of $\alpha$ bounded away from 2 but with some degradation as $\alpha \to 2$, which is expected as the distribution approaches a tilted Gaussian distribution. The impacts of varying $\beta$ and $\lambda$ are shown similarly in Figure 3 and Figure 4. This analysis shows strong evidence that Algorithm 4.2 is indeed sampling from the correct distribution. We can see that performance improves notable for small values of $\beta$ and larger values of $\lambda$. In all of these figures, higher order moments than 2 were calculated using sympy symbolic differentiation of the characteristic function. There is another clear pattern: higher moments lead to larger error between the moment of the sample and the theoretical moment of the distribution. This is likely because the higher moments are governed by extreme values, especially as the tempering parameter $\lambda \to 0$ or the scale $\beta$ grows large. Therefore $N = 100,000$ may not be enough samples to accurately encapsulate these extreme values. As a test for this theory, Figure **??** re-plots the moments from Figure 3 over the range $\beta \in [1, 2]$ but this time with $N = 1,000,000$. The error is markedly reduced, and gives confidence of the error converging to 0 as $N \to \infty$.

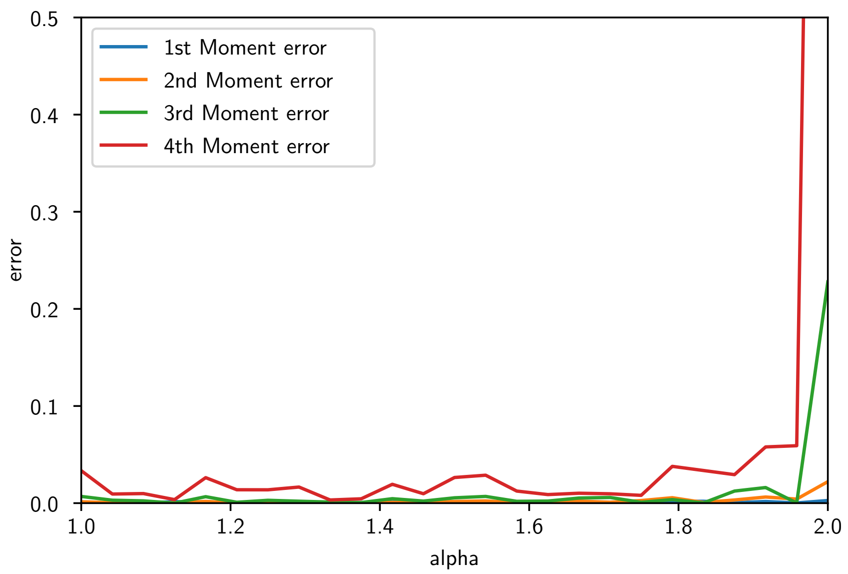


Figure 2: Absolute error between analytical moments vs observed moments calculated over 100,000 samples for $\alpha \in (1, 2)$, fixing $\beta = 0.1$ and $\lambda = 1$

We now investigate the efficiency of the algorithm. Using the discussion earlier, we know the expected number of times the algorithm runs before returning a value is given by the accept-reject bounding constant $d$ which here is $A = 4\sqrt{kc}$. Thus to analyse the efficiency of this algorithm, we need to consider how the values of $c$ and $k$ vary with the parameters $\alpha, \beta$ and $\lambda$. These are the integrals over the real line of $|\varphi_{\mathrm{TS}}|$ and $|\varphi''_{\mathrm{TS}}|$ respectively. Recalling the representations derived

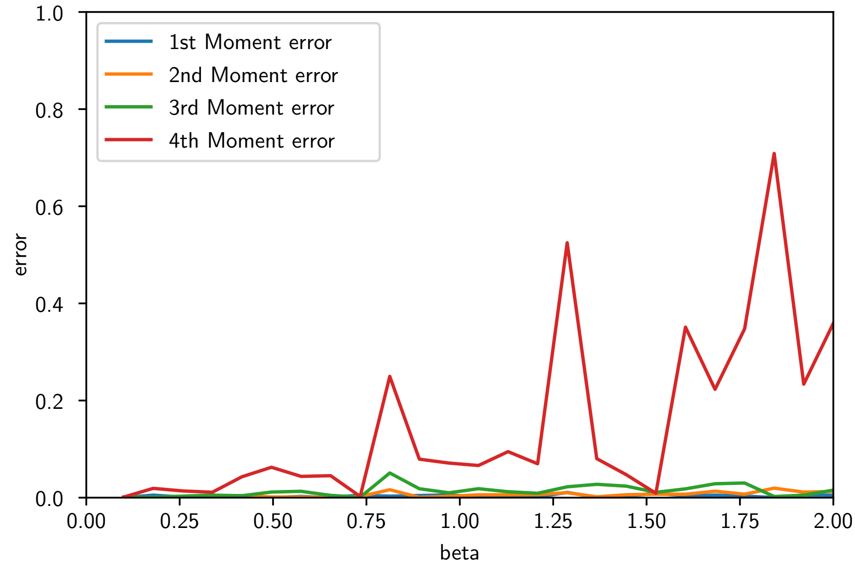


Figure 3: Absolute error between analytical moments vs observed moments over 100,000 samples for varying values of $\beta$, fixing $\alpha = 1.5$ and $\lambda = 1$

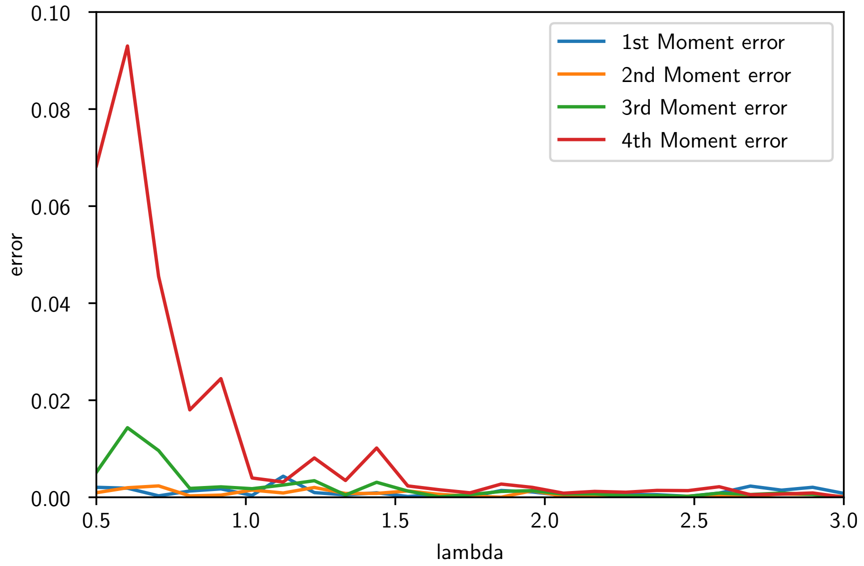


Figure 4: Absolute error between analytical moments vs observed moments over 100,000 samples for varying values of $\lambda$, fixing $\alpha = 1.5$ and $\beta = 0.1$

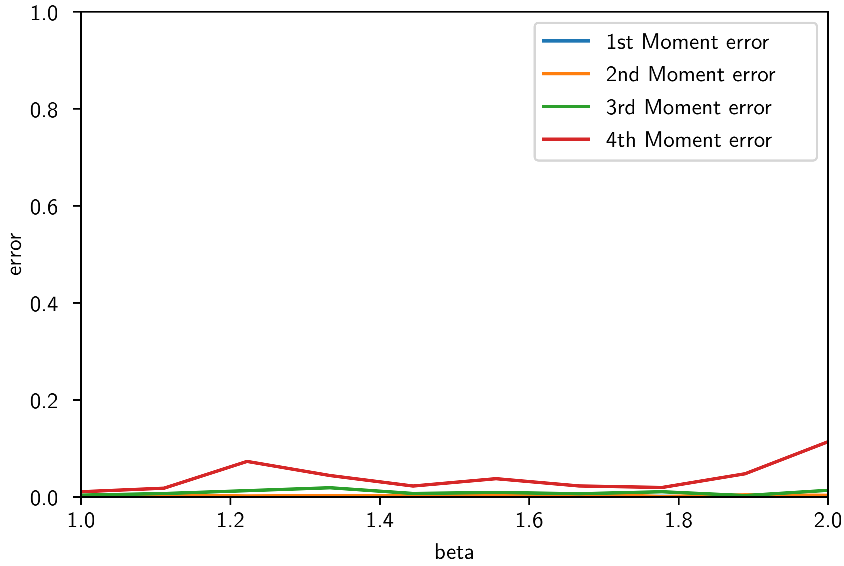


Figure 5: Absolute error between analytical moments vs observed moments over 1,000,000 samples for varying values of $\beta$, fixing $\alpha = 1.5$ and $\lambda = 1$

in section 3.2, we can express

$$|\varphi(y)| = \left|\exp\left\{\beta\,\Gamma(-\alpha)\left[(\lambda - iy)^{\alpha} - \lambda^{\alpha} + iy\alpha\lambda^{\alpha-1}\right]\right\}\right| \tag{24}$$

$$|\varphi''(y)| = \left|\beta\varphi(y)\left[\beta\,\Gamma(1-\alpha)^2\big((\lambda - iy)^{\alpha-1} - \lambda^{\alpha-1}\big)^2 - \Gamma(2-\alpha)(\lambda - iy)^{\alpha-2}\right]\right| \tag{25}$$

This is a very difficult problem to consider analytically, especially with the complex integration involved. The heatmaps in Figure 6 calculate $4\sqrt{kc}$ across a grid of parameters to give a sense of how the efficiency of the algorithm changes with the parameters. We can see the general pattern: the algorithm expects to take more time with smaller input parameters. However these expected values are still very promising, with the algorithm expecting to execute fewer than 3 times. So it becomes of interest to study these extremes, as $\alpha \to 1$, $\beta, \lambda \to 0$. One method to do this is to hold 2 of the variables constant and study the limit as the other one tends to these extreme values. But this is still a difficult problem analytically as to evaluate the limits of $c$ and $k$ would take careful use of integration convergence theorems, and it is not immediately clear which, if any, of $\varphi_{\text{TS}}$ or $\varphi''_{\text{TS}}$ are monotonic or dominated (when viewed as functions of $\alpha$, $\beta$ or $\lambda$.) So I will consider these limits numerically. Tables 1 - 3 display numeric calculations of how the expectation $d$ changes when we fix 2 parameters as constants and let the other approach its extreme value.

These were calculated using python and for smaller values, errors in the numerical integration and overflow warnings prevented getting any useful results. Nonetheless these results are strong,

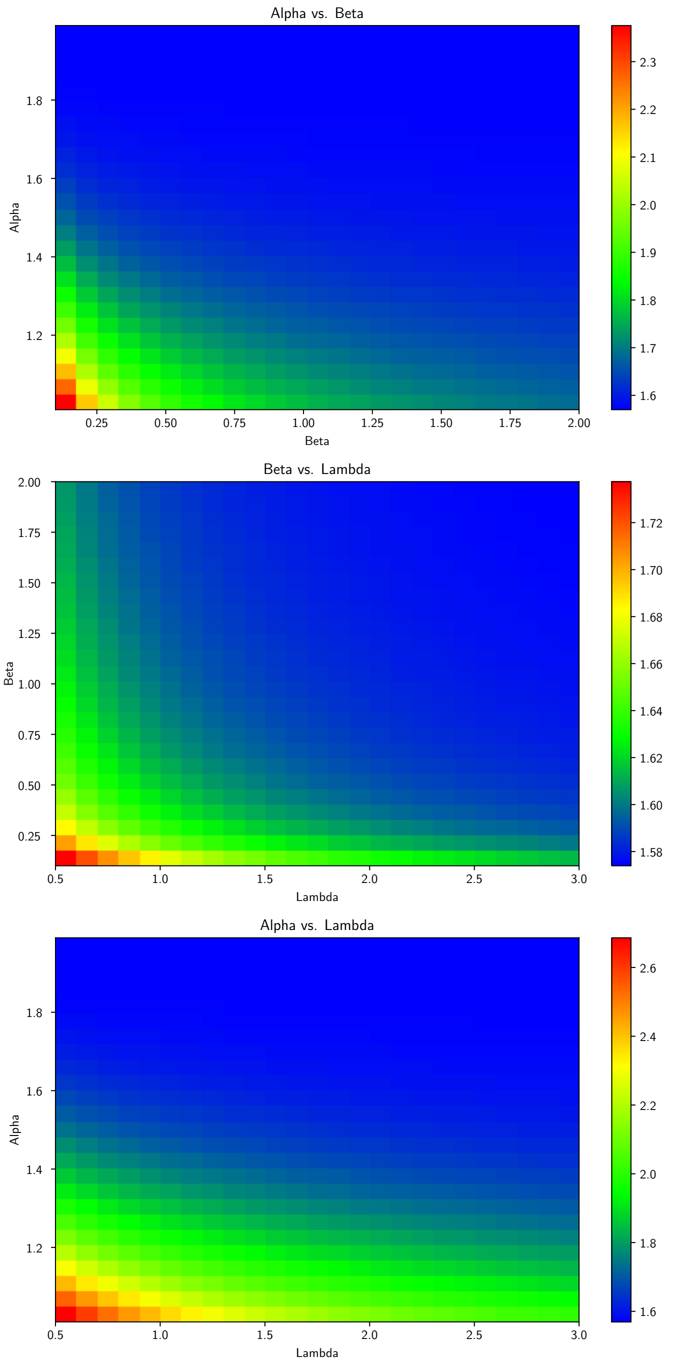


Figure 6: Heatmap showing how expected number of executions of Algorithm 4.2 varies as parameters vary. Where a parameter is excluded, it takes the value $\alpha = 1.5$, $\beta = 0.1$ and $\lambda = 1$.

Table 1: The expected number of executions of Algorithm 4.2 for extreme values of $\beta$, fixing $\alpha = 1.5$ and $\lambda = 1$.

| $\beta$ | 1 | 0.1 | 0.01 | 0.001 | 0.0001 | $10^{-5}$ | $10^{-6}$ | $10^{-7}$ |
|---|---|---|---|---|---|---|---|---|
| $d$ | 1.5983 | 1.6788 | 1.8177 | 1.9522 | 2.0430 | 2.0936 | 2.1192 | 2.1316 |

Table 2: The expected number of executions of Algorithm 4.2 for extreme values of $\alpha$, fixing $\beta = 0.1$ and $\lambda = 1$.

| $\alpha$ | 1.1 | 1.01 | 1.001 | 1.0001 | $1+10^{-5}$ | $1+10^{-6}$ | $1+10^{-7}$ | $1+10^{-8}$ |
|---|---|---|---|---|---|---|---|---|
| $d$ | 2.1574 | 2.3756 | 2.4014 | 2.4040 | 2.4043 | 2.4043 | 2.4043 | 2.4043 |

indicating that if only one parameter is approaching the extreme value, the algorithm is still very efficient. Table 2 especially implies that there does exist some limit for $d$ as the parameters approach extreme values, and this is a significant result as it implies the existence of a lower bound on the efficiency of the algorithm, which would be very useful to someone implementing it on a commercial scale. The problem remains of finding this bound by evaluating the integrals in (24), but I leave this for someone more experienced in complex integration.

As a final verification that all of this is correct, I compare the theoretical expected number of times the algorithm runs with the observed average when generating 100,000 samples. These results are displayed in Table 4 for a range of parameters. Given we expect a variance of 0.01 for a sample size of 100,000, these results are good, with a fall for $\alpha = 1.8$, which might be as expected, due to the worsening performance of the algorithm as $\alpha$ approaches 2.

## 4.3 Sampling from Increments of Tempered Stable driven Lévy Processes

We conclude this section with a short discussion on how to adapt these algorithms to sample from increments of $X_t$ where $X$ is a Lévy Process with $X_1 \sim \mathrm{TS}_+(\alpha, \beta, \lambda, 0)$ and $t > s > 0$. As seen when introducing Lévy Processes , we can express the characteristic function $\varphi_{X_t} = \varphi^t_{X_1} = \varphi^t_{\mathrm{TS}_+}$ which we can write as

$$\begin{aligned}
\varphi^t_{\mathrm{TS}_+} &= \exp\left\{ t\int_0^\infty [e^{iyz} - 1 - iyz]\frac{\beta e^{-\lambda z}}{z^{\alpha+1}}\mathrm{d}z \right\} \\
&= \exp\left\{ \int_0^\infty [e^{iyz} - 1 - iyz]\frac{(\beta t)e^{-\lambda z}}{z^{\alpha+1}}\mathrm{d}z \right\}
\end{aligned}$$

So we can simulate $X_t$ by sampling from $\mathrm{TS}_+(\alpha, t\beta, \lambda, 0)$, which is a direct application of Algorithm 4.2.

Table 3: The expected number of executions of Algorithm 4.2 for extreme values of $\lambda$, fixing $\alpha = 1.5$ and $\beta = 0.1$.

| $\lambda$ | 1 | 0.1 | 0.01 | 0.001 | 0.0001 | $10^{-5}$ | $10^{-6}$ | $10^{-7}$ |
|---|---|---|---|---|---|---|---|---|
| $d$ | 1.6788 | 1.8893 | 2.0430 | 2.1086 | 2.1316 | 2.1391 | 2.1415 | 2.2.1423 |

| Parameters | Number of executions before exiting | |
|---|---|---|
| | Observed | Theoretical |
| $\alpha = 1.3$ $\beta = 0.1$ $\lambda = 0.5$ | 1.9714 | 1.9718 |
| $\alpha = 1.5$ $\beta = 0.5$ $\lambda = 1$ | 1.6139 | 1.6150 |
| $\alpha = 1.8$ $\beta = 1$ $\lambda = 2$ | 1.5088 | 1.5706 |

Table 4: Theoretical vs Observed values of $d$ over 100,000 samples for varying parameters

# 5 Pricing Barrier Options

We conclude this dissertation by briefly describing a potential application of this algorithm. Traditionally, when pricing options stocks are modelled as some variant of Brownian Motion, for example the exponential stock model. These approaches all rely on the underlying assumption of asset returns following normal distributions, so to improve accuracy we would like to be able to model these processes as following tempered stable distributed increments. Now we have established an exact, efficient simulation algorithm for sampling from increments of tempered stable processes, we can use it for exactly this. We focus on pricing barrier options, where the payout depends not only on the final price, but on the supremum of process too. In [1], an efficient method to sample from both the final value and the supremum is presented, and this algorithm works for any Lévy Process where an algorithm to sample from $X_t$ is known. At the time of writing, the algorithm for exact sampling from Tempered Stable driven processes where $\alpha \in (1, 2)$ was unknown to the authors, so in this final section I introduce their algorithm, and discuss the suitability of applying Algorithm 4.2 in conjunction with their algorithm to price barrier options.

## 5.1 Introducing Barrier Options

I give a short introduction to barrier options. Much work has been done on the study of these derivatives, with many comprehensive resources available. There are two major categories of barrier options, **knock-out** options which become worthless if the underlying asset hits a predetermined price level, and **knock-in** options, which only come into existence if the barrier price level is hit. These in turn can be 'up', where the asset price must cross above the barrier level, or 'down', where the asset price must fall beneath the barrier level. Formally we define the payoff function as follows.

**Definition 5.1.1.** Let $(S_t)_{t\geq 0}$ be the price process of some risky asset. Then an up-and-out barrier option with strike $K > 0$, barrier $B > 0$ and maturity $T > 0$ has payoff at time $T$ given by

$$V_T = \begin{cases} (S_T - K)^+ & \sup_{t \leq T} S_t < B \\ 0 & \sup_{t \leq T} S_t \geq B \end{cases}$$

Down-and-out, up-and-in and down-and-in barrier options are defined analogously. The value of these options is path dependent, meaning it depends on the entire path $(S_t)_{t<T}$ not just the final value $S_T$. Therefore, to price 'up' options using a Monte-Carlo method, we need to be able to sample from both the final price of the asset and the supremum of the price process up to time $T$ (or infimum for down options). I will focus exclusively on pricing 'up' options, and the next section outlines an approach from [1] to sample both the final value and the supremum of a Lévy Process at the same time, provided you can sample from the price process $S_t$ for any $t$.

## 5.2 A Stick Breaking Approximation

We begin by defining a stick breaking process. For the rest of this subsection, $T > 0$ is some fixed time and $X_t$ is a tempered stable process, representing the payoff .

**Definition 5.2.1.** Let $U_1, U_2, \ldots$ be a sequence of i.i.d Uniform(0,1) random variables. Then define recursively the decreasing sequence $L_i$ by

$$L_0 = T$$
$$L_i = U_i * L_{i-1}$$

We also define $\ell_i := L_i - L_{i-1}$.

It is fairly easy to see $L_i \to 0$ as $i \to \infty$, and this can be proven formally by noticing that

$$L_i = T \prod_{j=1}^{i} U_j$$

and since each $U_j$ is less than 1 almost surely, the product is 0 with probability 1. We now want to use this to sample the triplet $\chi := (X_T, \overline{X}_T, \tau_T)$ where $X_T$ is the final value of the tempered stable process, $\overline{X}_T$ is the supremum of the process up to $T$ and $\tau_T$ is the time at which this supremum is attained. We do this using the *concave majorant*, defined as follows.

**Definition 5.2.2.** The concave majorant is the least concave pointwise function $C : [0, T] \to \mathbb{R}$ satisfying

$$X_t \leq C_t \ \forall t < T$$

This is a piecewise linear function, comprised of infinitely many line segments, which are known as *faces*. The length of these faces are a countable infinite collection of positive real numbers which sum to $T$. For any positive integer $n$, we define $d_n$ to be the left endpoint of face $n$ and $g_n$ be the right endpoint.

Now if $Y$ is an independent copy of $X$, which is also independent of the stick breaking process $L$, by a clever result of [18] (presented in [1]) that if we perform a size-biased sampling of the faces to obtain a numbering $\mathcal{N} = (d_n - g_n, C_{d_n} - C_{g_n})_{n\in\mathbb{N}}$ we have

$$\mathcal{N} = (d_n - g_n, C_{d_n} - C_{g_n})_{n\in\mathbb{N}} \stackrel{d}{=} (\ell_n, Y_{L_{n-1}} - Y_{L_n})_{n\in\mathbb{N}} \tag{26}$$

Therefore we obtain the following representation

$$\chi = (X_T, \overline{X}_T, \tau_T) \stackrel{d}{=} \sum_{k=1}^{\infty} (Y_{L_{k-1}} - Y_{L_k},\ \max(0, Y_{L_{k-1}} - Y_{L_k}),\ \ell_k \mathbb{1}_{Y_{L_{k-1}} - Y_{L_k} > 0}) \tag{27}$$

*Sketch proof for the representation (27)* . The first part is obvious via telescoping sum, and the second part comes from (26), as the supremum of the process will be equal to the sum of the faces of the concave majorant with positive slope, and then the time this is achieved in the sum of the length of the intervals over which $C_t$ has positive slope. □

As the sampling was size based and the total sum of the length of the faces is finite, in this numeration the faces get smaller as $n$ increases. We can therefore approximate $\chi$ by

$$\chi_n := \sum_{k=1}^{n} (Y_{L_{k-1}} - Y_{L_k},\ \max(0, Y_{L_{k-1}} - Y_{L_k}),\ \ell_k \mathbb{1}_{Y_{L_{k-1}} - Y_{L_k} > 0})$$

Hence we have the following algorithm, which samples a random vector equal in law to $\chi_n$ [1, SB-Alg]

**Algorithm 3** Stick-Breaking algorithm for sampling $\chi$

```
1: Set L_0 = T, 𝒳_0 = (0,0,0)
2: for k = 1, ..., n do
3:     Generate U_k ~ Uniform(0,1)
4:     Set ℓ_k = U_k L_{k-1}, L_k = L_{k-1} − ℓ_k
5:     Generate 𝒴_k with the distribution of X_{L_k}
6:     Set 𝒳_k = 𝒳_{k-1} + (𝒴_k, max(𝒴_k, 0), ℓ_k 𝟙_{𝒴_k>0})
7: end for
8: Generate 𝒴_{n+1} with the distribution of X_{L_n}
9: return 𝒳_n + (𝒴_{n+1}, max(𝒴_{n+1}, 0), L_n 𝟙_{𝒴_{n+1}>0})
```

This algorithm samples the increment for $X_t$ $n+1$ times, and if we sample $X_t$ using Algorithm 2, we expect each sample to take an average of $4\sqrt{kc}$ executions (where $k$ and $c$ are defined as in Section 4). This makes Algorithm 5.2 a very efficient sampling algorithm when compared to a novel Monte-Carlo method. By coupling $(X, \ell, Y)$ together, we can compare $\chi$ and $\chi_n$ on the same probability space to explicitly evaluate the convergence of $\chi_n$ to $\chi$ using the following theorem [1, Thm. 1].

**Theorem 5.2.3.** For $n \in \mathbb{N}$, define

$$\Delta_n = \overline{X}_T - \sum_{k=1}^{n} \max(Y_{L_{k-1}} - Y_{L_k}, 0)$$
$$\delta_n = \tau_T - \sum_{k=1}^{n} \ell_k \mathbb{1}_{Y_{L_{k-1}} - Y_{L_k}}$$

Then we can define the vector of errors by

$$\begin{aligned}\chi - \chi_n &= (0,\, \Delta_n - \max(0, Y_{L_n}),\, \delta_n - L_n \mathbb{1}_{Y_{l_n} > 0}) \\ &:= (0, \Delta_n^{SB}, \delta_n^{SB})\end{aligned}$$

Then (conditional on $L_n$),

$$(Y_{L_n}, \Delta_n, \delta_n) \stackrel{d}{=} (Y_{L_n}, \overline{Y}_{L_n}, \tau_{L_n}^{Y})$$

where $\tau^Y$ denotes the time the process $Y$ reaches its supremum. Hence

$$(\Delta_n^{SB}, \delta_n^{SB}) \stackrel{d}{=} (\overline{Y}_{L_n} - \max(0, Y_{L_n}), \tau_{L_n}^{Y} - L_n \mathbb{1}_{Y_{L_n} > 0})$$

Furthermore, $0 \le \Delta_{n+1}^{SB} \le \Delta_n^{SB} \le \Delta_n$ and $0 \le \delta_n \le L_n$ with $|\delta_n^{SB}| \le L_n$ with all inequalities holding almost surely.

*Proof.* See [1]. □

From this theorem we can deduce that the sequences $(\Delta_n^{SB})_{n \in \mathbb{N}}$ and $\delta_n^{SB}$ are non increasing and converge to 0 almost surely. We can also see from this theorem that the error $\Delta_n^{SB}$ is bounded above by the supremum of $X$ over the interval $[0, TL_n]$, where $\mathbb{E}[TL_n] = t \cdot 2^{-n}$, so the interval is geometrically small. This then implies that the error becomes geometrically small (for a full, rigorous exploration of this see [1, Section 2]). This is significant because the novel Monte-Carlo estimator will have error of order $O(1/\sqrt{N})$, so this algorithm is orders of magnitude faster.

**Remark 5.2.4.** Executing this algorithm involves sampling from $X_t$ for exponentially small $t$. So we would like to be able to say that the sampling algorithm is both accurate and efficient for small values of $t$. Subsection 4.3 tells us that this is equivalent to the behaviour for very small values

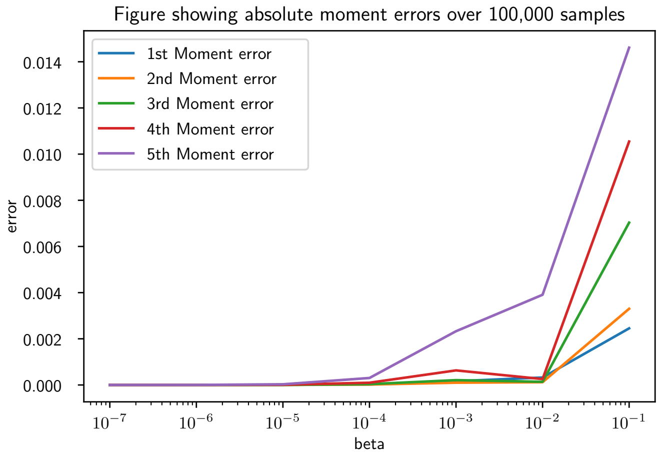


Figure 7: Absolute error between analytical moments vs observed moments over 100,000 samples for varying values of $\beta$, fixing $\alpha = 1.5$ and $\lambda = 1$

of $\beta$, and Table 1 implies that, as long as the values for $\alpha$ and $\lambda$ are not varying or too extreme, letting $\beta \to 0$ does not have a significant impact on the efficiency of the algorithm. To investigate correctness, Figure 7 extends Figure 3 for these small values of $\beta$. We can see that the algorithm performs exceptionally well for these small timesteps, making it highly appropriate to use with Algorithm 5.2.

# 6 Conclusion

In conclusion, this dissertation has provided a comprehensive resource on Tempered Stable distributions, exploring the motivation and technicalities for deriving them from the heavy-tailed Stable distributions. I then presented 2 algorithms from Kawai,Masuda [10] for exact sampling from these distributions, one for $\alpha \in (0, 1)$ and one for $\alpha \in (1, 2)$. These were rigorously derived, and both accuracy and efficiency were analysed in detail. Whilst many algorithms are known for sampling from these distributions when $\alpha \in (0, 1)$, the important result was the detailed presentation and analysis of the algorithm for exact sampling when $\alpha \in (1, 2)$. I made an effort to include every detail such that the algorithm is easily repeatable for any interested readers.

This class of Tempered Stable distributions is widely applicable in many areas of mathematical modelling, and can be used to generalise assumptions of normal distributions wherever they come up. I considered a particular application to finance and options pricing, where the assumption that asset returns are normally distributed is replaced by a Tempered Stable distribution. González Cázares, Mijatović et. al presented an algorithm for efficient simulation of extrema of Lévy Processes in [5], but as of writing an algorithm for sampling from increments of a Tempered Stable Lévy Process with $\alpha \in (1, 2)$ was unknown to the authors. In the last section, I briefly introduced this algorithm, and discussed the suitability of combining this with Algorithm 4.2 to efficiently price barrier options where the underlying asset returns are modelled by Tempered Stable Distributions.

Extensions to this dissertation might be an attempt to analytically evaluate the limits of the acceptance rate for Algorithm 4.2 as the parameters take extreme values, rigorously extending Algorithm 4.2 to include the case $\alpha = 1$ or applying Algorithm 4.2 to any situation where currently approximate simulation algorithms are used. Algorithm 4.2 can also be used to sample from CGMY distributions with $Y \in (1, 2)$, which may have many uses in finance. The code containing a class implementing the Tempered Stable distribution and Algorithm 4.2 is available at githib.com/mattball296/diss.

# References

[1] Jorge Ignacio Gonzá lez Cázares, Aleksandar Mijatović, and Gerónimo Uribe Bravo. Geometrically convergent simulation of the extrema of lévy processes. *Mathematics of Operations Research*, 47(2):1141–1168, may 2022. doi: 10.1287/moor.2021.1163. URL `https://doi.org/10.1287%2Fmoor.2021.1163`.

[2] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976. ISSN 01621459. URL `http://www.jstor.org/stable/2285309`.

[3] Luc Devroye. On the computer generation of random variables with a given characteristic function. *Computers and Mathematics with Applications*, 7(6):547–552, 1981. ISSN 0898-1221. doi: https://doi.org/10.1016/0898-1221(81)90038-9. URL `https://www.sciencedirect.com/science/article/pii/0898122181900389`.

[4] Luc Devroye. Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Trans. Model. Comput. Simul.*, 19, 10 2009. doi: 10.1145/1596519.1596523.

[5] Jorge Ignacio González Cázares and Aleksandar Mijatovic. Monte carlo algorithm for the extrema of tempered stable processes, 03 2021.

[6] Oliver C. Ibe. 11 - levy processes. In Oliver C. Ibe, editor, *Markov Processes for Stochastic Modeling (Second Edition)*, pages 329–347. Elsevier, Oxford, second edition edition, 2013. ISBN 978-0-12-407795-9. doi: https://doi.org/10.1016/B978-0-12-407795-9.00011-6. URL `https://www.sciencedirect.com/science/article/pii/B9780124077959000116`.

[7] Kiyosi ITOcirc. On stochastic processes (i). *Japanese journal of mathematics :transactions and abstracts*, 18:261–301, 1941. doi: 10.4099/jjm1924.18.0_261.

[8] J. Jacob and A. N. Shiryayev. *Limit Theorems for Stochastic Processes.* Springer, 2002.

[9] Svante Janson. Stable distributions, 2011 (revised 2022).

[10] Reiichiro Kawai and Hiroki Masuda. On simulation of tempered stable random variates. *Journal of Computational and Applied Mathematics*, 235(8):2873–2887, 2011. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2010.12.014. URL `https://www.sciencedirect.com/science/article/pii/S0377042710006643`.

[11] Ismo Koponen. Analytic approach to the problem of convergence of truncated lévy flights towards the gaussian stochastic process. *Phys. Rev. E*, 52:1197–1199, Jul 1995. doi: 10.1103/PhysRevE.52.1197. URL `https://link.aps.org/doi/10.1103/PhysRevE.52.1197`.

[12] Andreas Kyprianou. Levy processes and continuous-state branching processes: part i. 01 2014. doi: 10.1007/978-3-642-37632-0_12.

[13] Andreas Kyprianou. *Fluctuations of Lévy Processes with Applications*. 01 2014. ISBN 978-3-642-37632-0. doi: https://doi.org/10.1007/978-3-642-37632-0.

[14] Uwe Küchler and Stefan Tappe. Tempered stable distributions and processes. *Stochastic Processes and their Applications*, 123(12):4256–4293, 2013. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2013.06.012. URL `https://www.sciencedirect.com/science/article/pii/S0304414913001828`.

[15] P. Lévy. *Theorie de L'addition Des Variables Aleatoires*. Gauthier-Villars, 1954. URL `https://books.google.co.uk/books?id=yrlCnQAACAAJ`.

[16] Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36, 1963. URL `https://EconPapers.repec.org/RePEc:ucp:jnlbus:v:36:y:1963:p:394`.

[17] J.P. Nolan. *Univariate Stable Distributions: Models for Heavy Tailed Data*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, 2020. ISBN 9783030529154. URL `https://books.google.co.uk/books?id=c7r8DwAAQBAJ`.

[18] Jim Pitman and Geró nimo Uribe Bravo. The convex minorant of a lévy process. *The Annals of Probability*, 40(4), jul 2012. doi: 10.1214/11-aop658. URL `https://doi.org/10.1214%2F11-aop658`.

[19] Jan Rosiński. Tempering stable processes. *Stochastic Processes and their Applications*, 117(6):677–707, 2007. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2006.10.003. URL `https://www.sciencedirect.com/science/article/pii/S030441490600144X`.

[20] Ken-Iti Sato. *Lévy processes and Infinitely Divisible Distributions*. 1999.

[21] Karl Sigman. Lévy processes in finance : The change of measure and non-linear dependence, 2005. URL `https://hdl.handle.net/20.500.11811/2252`.

[22] Jens Wannenwetsch. Lévy processes in finance : The change of measure and non-linear dependence, 2005. URL `https://hdl.handle.net/20.500.11811/2252`.