Lévy Processes—the Normal Approximation in Simulation

Simon Riise Niemann

Master's thesis for the degree cand. act.
at University of Copenhagen

Supervisor: Thomas Mikosch

2nd December 2002

Preface

The use of jump processes for modelling purposes has expanded over the last couple of years. Lévy jump processes constitutes a subclass. By now Lévy jump processes are considered as more realistic models for applications in finance, mainly because the jump characteristics of the sample paths are closer to real-life financial data than the paths of Brownian motion.
We consider the approximation of small jumps of Lévy processes by a normal random variable. Especially, we investigate simulations by use of series expansions thoroughly, and we study how the approximation works in practice.
As an application of our results we consider a stochastic differential equation of the (shifted) Ornstein–Uhlenbeck type, driven by a non-Gaussian Lévy process—in our case an α-stable Lévy motion.
Our investigation mainly aims at the use of non-Gaussian Lévy processes in insurance and finance, but Lévy jump processes have by now been used extensively in many other areas, such as meteorology, seismology, turbulence and signal processing of sonar and radar data. This indicates, that this text could be relevant in other areas.
The basic theory of this thesis is based on the recent monographs of Bertoin [6] and Sato [29].

The source code developed for use in this thesis is quite comprehensive and therefore not printed in the appendix—the interested reader is encouraged to download source code and programs on my web page www.actuary.dk/thesis.

Contents

1 Lévy processes

Stochastic processes are mathematical models for random phenomena in real life. Whenever in mathematics we want to make a statement about future behavior of random phenomena, say the development of a stock index or some financial derivative, we need to come up with a model. The model reflects the way we think of the problem and how it is approached. It is indeed crucial for our results and we therefore need to be careful in choosing it. Obviously, the empirical features of real-life phenomena should match the ones given by a particular process, and therefore choosing the right model is a science of its own—no perfect solution can be expected.
Lévy processes constitutes one of the most important classes of stochastic processes. The stationary independent increments are often among the properties wanted when modelling and, needless to say, a lot of theory has been developed. We will start up introducing stochastic processes and Lévy processes—then move on to present the Lévy–Khintchine representation and the Lévy–Itô decomposition. These are the backbones for looking deeper into the main topic of this paper—approximation of small jumps of Lévy processes. The presentation below closely follows Sato [29].

1.1 Stochastic processes and Lévy processes

Definition 1.1 *A family $\{X_t : t \geq 0\}$ of random variables on $\mathbb{R}^d$ with parameter $t \in [0, \infty]$ defined on a common probability space is called a stochastic process.*

The parameter t is usually taken for time. The class of stochastic processes is very large and contains many interesting subclasses. One of them is the class of Lévy processes—stochastic processes with stationary independent increments as stated precisely in

Definition 1.2 *A stochastic process $\{X_t : t \geq 0\}$ on $\mathbb{R}^d$ is called a Lévy process if the following conditions are satisfied*

1. *$X_0 = 0$ a.s.*
2. *For all $n \geq 1$ and all $0 = t_0 < t_1 < \cdots < t_n$ the random variables $X_0, X_{t_1} - X_0, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.*
3. *The distribution of $X_{s+t} - X_s$ does not depend on s.*
4. *It is stochastically continuous.*
5. *There is $\Omega_0 \in \mathcal{F}$ with $\mathrm{P}[\Omega_0] = 1$ such that, for every $\omega \in \Omega_0$, $X_t(\omega)$ is right-continuous in $t \geq 0$ and has left limits in $t > 0$.*

Condition 2 is called the independent increments property and condition 3 is the stationary increments property. We define an additive process as a

stochastic process satisfying the conditions 1, 2, 4 and 5. Notice that under conditions 2 and 3, condition 4 can be replaced by

$$\lim_{t\to 0} \mathrm{P}[|X_t| > \epsilon] = 0, \quad \epsilon > 0. \tag{1.1}$$

Moreover we call a process satisfying 1–4 a Lévy process in law. Any Lévy process in law has a càdlàg version and thus satisfies all the above conditions 1–5—a proof of this non-trivial result can be found in Sato [29] chapter 2, together with a similar result on additive processes in law.
Because of the stationary independent increments we can think of Lévy processes as random walks in continuous time. The class of Lévy processes contains well known processes such as the Brownian motion and the Poisson process. The Brownian motion is used for modelling continuous random motions and the Poisson process for jumping random motions. These two processes are essential to the class of Lévy processes, as will become evident later. Other examples of Lévy processes are stable processes (which include the Cauchy process) and the Gamma process.

1.2 Characteristic functions

For analyzing the distribution of a Lévy process, the characteristic functions are lending us a hand. Many calculations and representations can be done much easier using this transform.
First we need to define the inner product on $\mathbb{R}^d$. Let the column vectors $x = (x_j)_{j=1,\dots,d}$ and $y = (y_j)_{j=1,\dots,d}$ be elements in $\mathbb{R}^d$. The inner product is then defined as $\langle x, y\rangle = \sum_{j=1}^{d} x_j y_j$. The definition of characteristic functions is now as follows

Definition 1.3 *The characteristic function $\widehat{\mu}(z)$ of a probability measure μ on $\mathbb{R}^d$ is*

$$\widehat{\mu}(z) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\langle z,x\rangle} \mu(\mathrm{d}x), \quad z \in \mathbb{R}^d. \tag{1.2}$$

The characteristic function of the distribution P_X of a random variable X on $\mathbb{R}^d$ is denoted by $\widehat{P}_X(z)$. That is

$$\widehat{P}_X(z) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\langle z,x\rangle} P_X(\mathrm{d}x) = \mathrm{E}[\mathrm{e}^{\mathrm{i}\langle z,X\rangle}]. \tag{1.3}$$

Characteristic functions have many nice properties. Though we will be using many of the properties, we will only list the most important from the perspective of this paper. Write $\mu_1 * \mu_2$ for the convolution of the finite measures μ_1 and μ_2 and denote the n-fold convolution of μ by $\mu^n = \mu^{n*}$.

Proposition 1.4 *Let μ, μ_1, μ_2 be distributions on $\mathbb{R}^d$. If $\mu = \mu_1 * \mu_2$, then $\widehat{\mu}(z) = \widehat{\mu}_1(z)\widehat{\mu}_2(z)$. If X_1 and X_2 are independent random variables on $\mathbb{R}^d$, then $\widehat{P}_{X_1+X_2}(z) = \widehat{P}_{X_1}(z)\widehat{P}_{X_2}(z)$.*

A proof of this property (and others) of the characteristic functions can be found at many places in the literature—see e.g. Sato [29] for relevant references.
Suppose $\{X_t : t \geq 0\}$ is a Lévy process and let for $t \geq 0$, μ_t denote the distribution of $X_{s+t} - X_s$ for any $s \geq 0$. Note that $\mu_0 = \delta_0$, the degenerate probability distribution at 0. If we write for $s, t \geq 0$

$$X_{s+t} - X_0 = (X_s - X_0) + (X_{s+t} - X_s), \tag{1.4}$$

it is obvious that

$$\mu_{s+t} = \mu_s * \mu_t. \tag{1.5}$$

This indeed suggests a close relationship between Lévy processes and infinitely divisible distributions. First we properly introduce infinitely divisible distributions.

Definition 1.5 *A probability measure μ on $\mathbb{R}^d$ is infinitely divisible*[1] *if, for any positive integer n, there is a probability measure μ_n on $\mathbb{R}^d$ such that $\mu = \mu_n^n$.*

Developing the above idea further, one can easily show that for any Lévy process X on $\mathbb{R}^d$, the distribution of X_t is infinitely divisible for any t—e.g. let $t_k = \frac{kt}{n}$, $\mu = P_{X_t}$ and $\mu_n = P_{X_{t_k} - X_{t_{k-1}}}$. Since

$$X_t = (X_{t_1} - X_0) + \cdots + (X_{t_n} - X_{t_{n-1}}), \tag{1.6}$$

we have $\mu = \mu_n^n$ because of the stationary independent increments. Conversely, it follows from Kolmogorov's consistency theorem (see Kolmogoroff [20] or with this regard Sato [29]) that if μ is infinitely divisible, then there exists a Lévy process X such that $X_{s+t} - X_s$ has distribution μ_t for any $s, t \geq 0$.
Notice, it is only in that case when X is a Lévy process in law, we need the assumption of stochastic continuity in order to have the link from Lévy processes in law to infinitely divisible distributions. In that case, the distribution of X_t is infinitely divisible but not always equal to μ_1^t (μ_1 being the distribution of X_1). E.g. let $f(t)$ be a function such that $f(t) + f(s) = f(t+s)$, $t, s \geq 0$ but such that $f(t)$ is not a constant multiple of t, and let $X_t = f(t)$. If $t \geq 0$ is irrational, choose rational numbers r_n such that $r_n \to t$. X_t being stochastically continuous ensures the desired convergence $P_{X_{r_n}} \to P_{X_t}$.

1.3 The Lévy–Khintchine representation

Now the strong link between Lévy processes and infinitely divisible distributions is about to pay off. The Lévy–Khintchine representation (or Lévy–Khintchine formula) gives us a representation of characteristic functions of all infinitely divisible distributions. Let $D = \{x : |x| \leq 1\}$, the closed unit ball.

[1] Mathematicians who study algebra would say that the distributions $(\mu_t)_{t \geq 0}$ on $\mathbb{R}^d$ form a convolution semigroup.

Theorem 1.6 (Sato [29] theorem 8.1) (i) *If μ is an infinitely divisible distribution on $\mathbb{R}^d$, then*

$$\begin{aligned}\widehat{\mu}(z) &= \exp\bigg[-\frac{1}{2}\langle z, Az\rangle + \mathrm{i}\langle\gamma, z\rangle \\ &\qquad + \int_{\mathbb{R}^d}(\mathrm{e}^{\mathrm{i}\langle z,x\rangle} - 1 - \mathrm{i}\langle z, x\rangle 1_D(x))\nu(\mathrm{d}x)\bigg], \quad z \in \mathbb{R}^d, \quad (1.7)\end{aligned}$$

where A is a symmetric nonnegative-definite $d \times d$ matrix, ν is a measure on $\mathbb{R}^d$ satisfying

$$\nu(\{0\}) = 0 \quad \textit{and} \quad \int_{\mathbb{R}^d}(|x|^2 \wedge 1)\nu(\mathrm{d}x) < \infty, \tag{1.8}$$

and $\gamma \in \mathbb{R}^d$.
(ii) *The representation of $\widehat{\mu}(z)$ in (i) by A, ν and γ is unique.*
(iii) *Conversely, if A is a symmetric nonnegative-definite $d\times d$ matrix, ν is a measure satisfying (1.8), and $\gamma \in \mathbb{R}^d$, then there exists an infinitely divisible distribution μ whose characteristic function is given by (1.7).*

We call (A,ν,γ) in theorem 1.6 the generating triplet of μ. A and ν are called, respectively, the Gaussian covariance matrix and the Lévy measure of μ. To improve our intuition we have a look at the Lévy–Khintchine representation in one dimension—let X_t be a Lévy process, $z \in \mathbb{R}$ and theorem 1.6 gives us the representation

$$\begin{aligned}\widehat{P}_{X_t}(z) &= \mathrm{E}[\exp(\mathrm{i}zX_t)] \\ &= \exp\left[t\left(\mathrm{i}az - \frac{b^2z^2}{2} + \int_{-\infty}^{\infty}(\mathrm{e}^{\mathrm{i}zx} - 1 - \mathrm{i}zx1_{(|x|\le 1)})\nu(\mathrm{d}x)\right)\right]. \quad (1.9)\end{aligned}$$

Notice that the right hand side of (1.9) can be written as a product of three terms. Thus recalling proposition 1.4 we can write X_t as the independent sum $X_t = at + bW_t + J_t$ of a drift term at, a Brownian component bW_t where W_t is a standard Brownian motion and a (compensated) pure jump part J_t with Lévy measure ν. One can think of $\{J_t\}$ as a pure jump process where jumps of size x occur with intensity $\nu(\mathrm{d}x)$. If ν has finite mass $\lambda = \int_{-\infty}^{\infty}\nu(\mathrm{d}x) < \infty$, then $\{J_t\}$ is a compound Poisson process with intensity λ and jump size distribution $\frac{\nu(\mathrm{d}x)}{\lambda}$. The compensation of $\{J_t\}$ by the last term in the integral is only done in order to make sure the integral converges. Thus it is not needed if $\int_{-\infty}^{\infty}(|x| \wedge 1)\nu(\mathrm{d}x) < \infty$, i.e. the paths of $\{J_t\}$ are of finite variation. Since the role of the interval $[-1, 1]$ in the centering is arbitrary, a compensated Lévy process is only given canonically up to a drift term—a clear presentation of the centering procedure can be found in Asmussen [3] appendix A3.
The following example shows the Lévy–Khintchine representation in action for three specific Lévy processes in one dimension.

Example 1.7 Using theorem 1.6 we have for X_t' being a Brownian motion in $\mathbb{R}$ with drift $a \in \mathbb{R}$ and variance $b^2 > 0$

$$\widehat{P}_{X_t'}(z) = \mathrm{E}[\exp(\mathrm{i}zX_t')] = \exp\left[t\left(\mathrm{i}az - \frac{b^2z^2}{2}\right)\right], \quad z \in \mathbb{R}, \tag{1.10}$$

and for X_t'' being a Gamma process[2] in $\mathbb{R}$ with parameters $a, b > 0$

$$\widehat{P}_{X_t''}(z) = \mathrm{E}[\exp(\mathrm{i}zX_t'')] = \exp\left[at\int_0^\infty (\mathrm{e}^{\mathrm{i}zx} - 1)\frac{\mathrm{e}^{-\frac{x}{b}}}{x}\mathrm{d}x)\right], \quad z \in \mathbb{R}, \tag{1.11}$$

and for X_t''' being a one-sided strictly $\frac{1}{2}$-stable process[3] in $\mathbb{R}_+$ with parameter $c > 0$

$$\widehat{P}_{X_t'''}(z) = \mathrm{E}[\exp(\mathrm{i}zX_t''')] = \exp\left[\frac{ct}{\sqrt{2\pi}}\int_0^\infty (\mathrm{e}^{\mathrm{i}zx} - 1)x^{-\frac{3}{2}}\mathrm{d}x)\right], \ z \in \mathbb{R}. \tag{1.12}$$

The relevant calculations can be found in Sato [29].

α-stable random variables are often represented by characteristic functions, as it is the best analytic way of characterizing all members of this class. Even though the α-stable laws are absolutely continuous, their densities (excluding a few exceptions) can only be expressed by complicated special functions—see Zolotarev [33].

1.4 The Lévy–Itô decomposition

The Lévy–Itô decomposition gives us a decomposition of sample functions of additive processes. It expresses a sample function as a sum of two independent parts—a continuous part and a (compensated) sum of independent jumps. Before writing up the theorem we need some preliminary definitions.

Definition 1.8 *Let $(\Theta, \mathcal{B}, \rho)$ be a σ-finite measure space. A family of $\overline{\mathbb{Z}}_+$-valued random variables $\{N(B) : B \in \mathcal{B}\}$ is called a Poisson random measure on Θ with intensity measure ρ, if the following hold:*

1. *For every B, $N(B)$ has Poisson distribution with mean $\rho(B)$.*
2. *If $B_1, \ldots, B_n$ are disjoint, then $N(B_1), \ldots, N(B_n)$ are independent.*
3. *For every ω, $N(\cdot, \omega)$ is a measure on Θ.*

Moreover write

$$\begin{aligned} D_{a,b} &= D(a,b] = \{x \in \mathbb{R}^d : a < |x| \le b\} \qquad \text{for } \ 0 \le a < b < \infty, \\ D_{a,\infty} &= D(a,\infty) = \{x \in \mathbb{R}^d : a < |x| < \infty\} \qquad \text{for } \ 0 \le a < \infty, \\ H &= (0,\infty) \times (\mathbb{R}^d \backslash \{0\}) = (0,\infty) \times D_{0,\infty}. \end{aligned}$$

We are now ready to formulate the Lévy–Itô decomposition.

[2]Define the Gamma distribution as having density $f(x) = \frac{1}{b^a\Gamma(a)}x^{a-1}\mathrm{e}^{-\frac{x}{b}}, \ 0 < x < \infty$.

[3]Define the one-sided strictly stable distribution of index $\frac{1}{2}$ (sometimes referred to as the Lévy distribution) as having density $f(x) = \frac{c}{\sqrt{2\pi}}\mathrm{e}^{-\frac{c^2}{2x}}x^{-\frac{3}{2}}, \ 0 < x < \infty$.

Theorem 1.9 (Sato [29] theorem 19.2) *Let $\{X_t : t \geq 0\}$ be an additive process on $\mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{F}, P)$ with system of generating triplets $\{(A_t, \nu_t, \gamma(t))\}$ and define the measure $\widetilde{\nu}$ on H by $\widetilde{\nu}((0,t] \times C) = \nu_t(C)$ for $C \in \mathcal{B}(\mathbb{R}^d)$. Using Ω_0 from definition 1.2 of an additive process, define, for $B \in \mathcal{B}(H)$,*

$$N(B,\omega) = \begin{cases} \#\{s : (s, X_s(\omega) - X_{s-}(\omega)) \in B\} & for\ \omega \in \Omega_0, \\ 0 & for\ \omega \notin \Omega_0. \end{cases} \tag{1.13}$$

Then the following hold

1. *$\{N(B) : B \in \mathcal{B}(H)\}$ is a Poisson random measure on H with intensity measure $\widetilde{\nu}$.*

2. *There is $\Omega_1 \in \mathcal{F}$ with $P[\Omega_1] = 1$ such that, for any $\omega \in \Omega_1$,*

$$\begin{aligned} X_t^1(\omega) &= \lim_{\epsilon \to 0} \int_{(0,t] \times D(\epsilon,1]} \{xN(\mathrm{d}(s,x),\omega) - x\widetilde{\nu}(\mathrm{d}(s,x))\} \\ &\quad + \int_{(0,t] \times D(1,\infty)} xN(\mathrm{d}(s,x),\omega) \end{aligned} \tag{1.14}$$

   is defined for all $t \in [0,\infty)$ and the convergence is uniform in t on any bounded interval. The process $\{X_t^1\}$ is an additive process on $\mathbb{R}^d$ with $\{(0, \nu_t, 0)\}$ as the system of generating triplets.

3. *Define*

$$X_t^2(\omega) = X_t(\omega) - X_t^1(\omega) \qquad for\ \omega \in \Omega_1. \tag{1.15}$$

   There is $\Omega_2 \in \mathcal{F}$ with $P[\Omega_2] = 1$ such that, for any $\omega \in \Omega_2$, $X_t^2(\omega)$ is continuous in t. The process $\{X_t^2\}$ is an additive process on $\mathbb{R}^d$ with $\{(A_t, 0, \gamma(t))\}$ as the system of generating triplets.

4. *The two processes $\{X_t^1\}$ and $\{X_t^2\}$ are independent.*

We call $\{X_t^1\}$ and $\{X_t^2\}$ respectively the jump part and the continuous part of $\{X_t\}$. We notice that the first term on the right hand side of (1.14) is compensated (to having mean 0) in order to have the integral converging—its limit is called the compensated sum of jumps.
We can use the Lévy–Itô decomposition to write any Lévy jump process $\{J_t\}$ as the independent sum $J_t = J_t^{(1)} + J_t^{(2)}$, where the Lévy measures of $\{J_t^{(1)}\}$ and $\{J_t^{(2)}\}$ are the restrictions $\nu^{(1)}$, $\nu^{(2)}$ of ν to respectively $\mathbb{R}^d \backslash D(1,\infty)$ and $D(1,\infty)$. Notice that $\nu^{(2)}$ is finite so $\{J_t^{(2)}\}$ is a compound Poisson process. The choice of areas (intervals) will often be chosen differently in order to make as much use of this strong result as possible—e.g. to respectively $\mathbb{R}^d \backslash D(\epsilon,\infty)$ and $D(\epsilon,\infty)$.

2 Approximation of small jumps of Lévy processes

Let $\{X_t\}$ be a Lévy process on $\mathbb{R}$. By use of the Lévy–Khintchine representation we can write X_t as the independent sum $X_t = at + bW_t + J_t$ of a drift term at, a Brownian component bW_t and a compensated pure jump part J_t with Lévy measure ν. Then we use the Lévy–Itô decomposition and write J_t as the independent sum $J_t = J_t^{(1)} + J_t^{(2)}$, where the Lévy measures of $\{J_t^{(1)}\}$ and $\{J_t^{(2)}\}$ are the restrictions $\nu^{(1)}$, $\nu^{(2)}$ of ν to respectively $[-\epsilon, \epsilon]$ and $\{x : |x| > \epsilon\}$. All together we have

$$X_t = at + bW_t + J_t^{(1)} + J_t^{(2)}, \tag{2.1}$$

where all the terms on the right hand side are independent. Hence these building blocks can be used for simulating any Lévy process. The Brownian and compound Poisson terms are not that difficult to simulate—but in many cases ν is not finite and we are facing the problem of simulating $J_t^{(1)}$. Recall from theorem 1.6 that $\int_{\mathbb{R}}(|x|^2 \wedge 1)\nu(\mathrm{d}x) < \infty$, thus $J_t^{(1)}$ may have infinitely many jumps (the case where ν is not finite). This means that we cannot simulate $J_t^{(1)}$ by standard procedures. Thus we have to look at alternatives for approximating the infinite number of jumps in finite time.

2.1 The normal approximation

One could simply simulate the Lévy process by choosing ϵ "sufficiently" small in some sense and then neglecting $J_t^{(1)}$, i.e. simulating

$$X_1^\epsilon(t) = at + bW_t + J_t^{(2)}. \tag{2.2}$$

But because $J_t^{(1)}$ is in general not of finite variation this could be a bad idea—$J_t^{(1)}$ can behave in a manner which is crucial for the process. Also remember that a compensated Lévy process is only given canonically up to a drift term (a changes with the compensation).
An improvement could be achieved by assuming that the error

$$X_\epsilon(t) = X(t) - X_1^\epsilon(t) \tag{2.3}$$

made in (2.2) is approximately normal. This would mean incorporating the variation of the small jumps

$$X_2^\epsilon(t) = at + \sqrt{b^2 + \sigma^2(\epsilon)}W_t' + J_t^{(2)}, \tag{2.4}$$

where W_t' is a version of W_t, and independent of $J_t^{(2)}$,

$$\sigma^2(\epsilon) = \int_{|x|\le\epsilon} x^2\nu(\mathrm{d}x). \tag{2.5}$$

We notice a Brownian term appearing even when the original process X does not have one ($b = 0$). This approximation has been suggested on intuitive

grounds in some particular cases by Bondesson [8] and Rydberg [27]. Asmussen and Rosiński [4] have looked deeper into this subject and presented a theorem stating when the functional CLT underlying (2.4) is valid.
We notice that the approximation error $X_\epsilon(t)$ given by (2.3) is a Lévy process with $\mathrm{E}[X_\epsilon(t)] = 0$ (use appropriate compensation) and $\mathrm{Var}[X_\epsilon(1)] = \sigma^2(\epsilon)$ (given by (2.5)).

Theorem 2.1 (Asmussen and Rosiński [4] theorem 2.1)
$\frac{X_\epsilon}{\sigma(\epsilon)} \xrightarrow{\mathcal{D}} W$ *as* $\epsilon \to 0$ *if and only if for each* $\kappa > 0$

$$\sigma(\kappa\sigma(\epsilon) \wedge \epsilon) \ \sim \ \sigma(\epsilon), \qquad as\ \epsilon \to 0. \tag{2.6}$$

The weak convergence of $\sigma(\epsilon)^{-1}X_\epsilon$ to a standard Brownian motion is in the Skorokhod space $\mathbb{D}[0,1]$, the space of right continuous paths with left limits (càdlàg paths), i.e. see Billingsley [7]. We also have an intuitive condition for the validity of (2.6).

Proposition 2.2 (Asmussen and Rosiński [4] proposition 2.2)
Condition (2.6) is implied by

$$\lim_{\epsilon \to 0} \frac{\sigma(\epsilon)}{\epsilon} = \infty. \tag{2.7}$$

Moreover, if ν *does not have atoms in some neighborhood of the origin, then (2.6) and (2.7) are equivalent.*

That is, the normal approximation always holds when the standard deviation of the small jump part of a Lévy process converges to zero slower than the level of truncation. Asmussen and Rosiński [4] also give an example that shows (2.6) and (2.7) are not equivalent in general.
In order to decide which processes do or do not admit the normal approximation, we have a look at the processes from example 1.7. The Brownian motion does not have any jumps, so there is no use for the approximation in this case. Another case that does not need the approximation either is the compound Poisson process, because ν is finite, and hence, X has only finitely many jumps in any finite time interval.

Example 2.3 For the compound Poisson process ν has the form $\lambda\nu'$, where λ is the intensity of jumps and ν' is the jump size distribution. That is, around 0, $\nu(x)$ is of order x^k, $k > -1$. We have

$$\begin{aligned}
\frac{\sigma(\epsilon)}{\epsilon} &= \frac{1}{\epsilon}\left(\int_0^\epsilon x^2 \nu(\mathrm{d}x)\right)^{\frac{1}{2}} = \frac{1}{\epsilon}\left(\int_0^\epsilon x^2 x^k \mathrm{d}x\right)^{\frac{1}{2}} \\
&= \frac{1}{\epsilon}\left(\frac{1}{3+k}\Big[x^{3+k}\Big]_0^\epsilon\right)^{\frac{1}{2}} = \frac{1}{\epsilon}\left(\frac{1}{3+k}\epsilon^{3+k}\right)^{\frac{1}{2}} \\
&\overset{\epsilon}{\propto} \frac{\epsilon^{\frac{1}{2}(3+k)}}{\epsilon} = \epsilon^{\frac{1}{2}+\frac{k}{2}} \to 0, \quad as\ \epsilon \to 0,\ k > -1.
\end{aligned} \tag{2.8}$$

So $\sigma(\epsilon) = o(\epsilon)$ and therefore (2.6) fails!

Example 2.4 For the one-sided strictly $\frac{1}{2}$-stable process we have $\nu(\mathrm{d}x) = \frac{c}{\sqrt{2\pi}} x^{-\frac{3}{2}} \mathrm{d}x$. And therefore

$$\begin{aligned} \sigma^2(\epsilon) &= \mathrm{Var}(X_\epsilon(1)) = \int_{|x|<\epsilon} x^2 \nu(\mathrm{d}x) = \frac{c}{\sqrt{2\pi}} \int_0^\epsilon x^{\frac{1}{2}} \mathrm{d}x \\ &= \frac{2c}{3\sqrt{2\pi}} \left[x^{\frac{3}{2}} \right]_0^\epsilon = \frac{2c}{3\sqrt{2\pi}} \epsilon^{\frac{3}{2}} \end{aligned} \tag{2.9}$$

and $\sigma(\epsilon) = \sqrt{\frac{2c}{3\sqrt{2\pi}}} \epsilon^{\frac{3}{4}}$. Notice that ν does not have any atoms and therefore by proposition 2.2 the normal approximation holds. Generally the normal approximation holds for any stable process with $\alpha \in (0, 2)$ (see (3.11)).

Example 2.5 For the Gamma process $\nu(\mathrm{d}x) = \lambda \frac{\mathrm{e}^{-\frac{x}{b}}}{x} \mathrm{d}x$. This gives

$$\begin{aligned} \sigma^2(\epsilon) &= \mathrm{Var}(X_\epsilon(1)) = \int_{|x|<\epsilon} x^2 \nu(\mathrm{d}x) = \lambda \int_0^\epsilon x \mathrm{e}^{-\frac{x}{b}} \mathrm{d}x \\ &\sim \lambda \int_0^\epsilon x \mathrm{d}x = \lambda \left[\frac{x^2}{2} \right]_0^\epsilon = \frac{\lambda}{2} \epsilon^2 \end{aligned} \tag{2.10}$$

and $\sigma(\epsilon) \sim \epsilon \sqrt{\frac{\lambda}{2}}$. Again ν does not have any atoms and therefore by proposition 2.2 the normal approximation does not hold.

But indeed in some sense the normal approximation is not 'far' away in the Gamma case. The standard deviation of the small jump part is proportional to ϵ and what we needed was ϵ to a power of $k < 1$. To know how 'far' away the normal approximation is—is it still good or should we not be using it—we investigate the actual distribution. Take a look at the logarithm of the moment generating function (the cumulant generating function), read $\sigma_\epsilon = \sigma(\epsilon)$

$$\begin{aligned} \log \mathrm{E} \left[\exp \left(s \frac{J^{(1)}(1) - \mu(\epsilon)}{\sigma(\epsilon)} \right) \right] &= \int_0^\epsilon \left(\mathrm{e}^{\frac{sx}{\sigma_\epsilon}} - 1 - \frac{sx}{\sigma_\epsilon} \right) \nu(\mathrm{d}x) \\ &= \int_0^\epsilon \frac{\lambda}{x} \left[\mathrm{e}^{\left(\frac{s}{\sigma_\epsilon} - \frac{1}{b} \right) x} - \mathrm{e}^{-\frac{x}{b}} - \frac{sx}{\sigma_\epsilon} \mathrm{e}^{-\frac{x}{b}} \right] \mathrm{d}x, \end{aligned} \tag{2.11}$$

and make a series expansion of the exponential function

$$\begin{aligned} (2.11) &= \int_0^\epsilon \frac{\lambda}{x} \Bigg[\left(1 - \frac{x}{b} + \frac{x^2}{2b^2} + O(x^3) \right) \left(1 + \frac{sx}{\sigma_\epsilon} + \frac{s^2 x^2}{2\sigma_\epsilon^2} + \sum_{n=3}^\infty \frac{s^n x^n}{n! \sigma_\epsilon^n} \right) \\ &\quad -1 + \frac{x}{b} - \frac{x^2}{2b^2} + O(x^3) - \frac{sx}{\sigma_\epsilon} + \frac{sx^2}{\sigma_\epsilon b} - \frac{sx^3}{2\sigma_\epsilon b^2} + O\left(\frac{sx^4}{\sigma_\epsilon} \right) \Bigg] \mathrm{d}x \end{aligned}$$

$$
\begin{aligned}
&= \int_0^\epsilon \frac{\lambda}{x}\left[\frac{s^2x^2}{2\sigma_\epsilon^2} - \frac{s^2x^3}{2\sigma_\epsilon^2 b} + \frac{s^2x^4}{4\sigma_\epsilon^2 b^2} + \left(1 - \frac{x}{b} + \frac{x^2}{2b^2}\right)\sum_{n=3}^{\infty}\frac{s^n x^n}{n!\sigma_\epsilon^n}\right]\mathrm{d}x \\
&\quad +o(1) \\
&= \lambda\int_0^\epsilon \left[\frac{s^2x}{2\sigma_\epsilon^2} + \left(1 - \frac{x}{b} + \frac{x^2}{2b^2}\right)\sum_{n=3}^{\infty}\frac{s^n x^{n-1}}{n!\sigma_\epsilon^n}\right]\mathrm{d}x + o(1) \\
&= \lambda\left(\left[\frac{s^2x^2}{4\sigma_\epsilon^2}\right]_0^\epsilon + \sum_{n=3}^{\infty}\left[\frac{s^n x^n}{n!n\sigma_\epsilon^n}\right]_0^\epsilon\right) = \lambda\left(\frac{s^2\epsilon^2}{4\sigma_\epsilon^2} + \sum_{n=3}^{\infty}\frac{s^n\epsilon^n}{n!n\sigma_\epsilon^n}\right) + o(1) \\
&= \frac{s^2}{2} + \lambda\sum_{n=3}^{\infty}\frac{s^n\epsilon^n}{n!n\sigma_\epsilon^n} + o(1) = \frac{s^2}{2} + \lambda\sum_{n=3}^{\infty}\frac{s^n}{n!n\left(\frac{\lambda}{2}\right)^{\frac{n}{2}}} + o(1). \qquad (2.12)
\end{aligned}
$$

The first term is recognized as the cumulant generating function for a normally distributed random variable. The second term is a series that only contributes with one nonzero term when calculating cumulants, and that is only for cumulants higher than the second. As we excepted, the normal approximation is 'close', and the two first moments actually coincide with those from a normally distributed random variable. By simulating different Gamma processes we will see how well the approximation works for different parameters—see section 3.7.

3 Simulation

In order to preserve our intuition when interpreting the simulations we will focus on the one dimensional case. Recall (2.1) which says that any Lévy process can be written as the independent sum of a drift term at, a Brownian component bW_t and a (compensated) pure jump part $J_t = J_t^{(1)} + J_t^{(2)}$ with Lévy measure ν. The Lévy measures of $J_t^{(1)}$ and $J_t^{(2)}$ are the restrictions $\nu^{(1)}$, $\nu^{(2)}$ of ν to respectively $[-\epsilon, \epsilon]$ and $\{x : |x| > \epsilon\}$. The increments of the Brownian motion are stationary, independent and normally distributed—this makes simulation of a sample path along a discrete skeleton straightforward. $J_t^{(2)}$ is a compound Poisson process with finite intensity $\lambda = \int \nu(\mathrm{d}y)$ and jump size distribution $\frac{\nu(\mathrm{d}y)}{\lambda}$. Simulation of a sample path is fairly simple since the "waiting times" between jumps are exponentially distributed, thus problems only occur if the jump size distribution is hard to simulate. All we are left with is $J_t^{(1)}$—the small jumps. If ν is not finite then $J_t^{(1)}$ has infinitely many jumps, and therefore we need to truncate or perform some limiting procedure.

3.1 Series representation

The (compensated) process J_t of jumps is described by the Lévy measure ν and the Poisson random measure N, recall (1.13) and (1.14). From theo-

rem 1.9 we know that

$$\begin{aligned} J_t(\omega) &= \lim_{\epsilon \to 0} \int_{(0,t]\times D(\epsilon,1]} \{xN(\mathrm{d}(s,x),\omega) - x\widetilde{\nu}(\mathrm{d}(s,x))\} \\ &\quad + \int_{(0,t]\times D(1,\infty)} xN(\mathrm{d}(s,x),\omega), \end{aligned} \tag{3.1}$$

and for $n \in \mathbb{N}$ we then define the limiting jump process

$$\begin{aligned} J_{n,t}(\omega) &= \int_{(0,t]\times D(\frac{1}{n},1]} \{xN(\mathrm{d}(s,x),\omega) - x\widetilde{\nu}(\mathrm{d}(s,x))\} \\ &\quad + \int_{(0,t]\times D(1,\infty)} xN(\mathrm{d}(s,x),\omega). \end{aligned} \tag{3.2}$$

Notice that the limiting jump process is compound Poisson and by use of the order statistics property write it as ($\stackrel{\mathcal{D}}{=}$ refers to identity of the finite dimensional distributions)

$$J_{n,t} \stackrel{\mathcal{D}}{=} \sum_{i\in\Lambda_n} \left(J^i 1_{(U_i \le t)} - tb_n\right), \tag{3.3}$$

where $\Lambda_n = \{i \ge 1 : |J^i| \ge \frac{1}{n}\}$, J^i are the jump sizes, U_i are i.i.d. uniform on [0,1] and $b_n = \int_{(0,t]\times D(\frac{1}{n},1]} x\widetilde{\nu}(\mathrm{d}(s,x))$. Thus in the limit we have

$$\sum_{i\in\Lambda_n} \left(J^i 1_{(U_i \le t)} - tb_n\right) \to J_t \quad a.s. \tag{3.4}$$

as $n \to \infty$. This is almost a series representation—all there is left to do is to define an order of summation. Basically a series representation can be written as

$$J_t = \sum_{i=1}^{\infty} \left(J^i 1_{(U_i \le t)} - tc_i\right), \tag{3.5}$$

for suitable centering constants c_i. There are many ways of building such series, see the article of Rosiński [26] and Samorodnitsky and Taqqu [28] for the special case of stable processes. The following example (corollary 1.4.3 from Samorodnitsky and Taqqu [28]) shows how a series representation for a stable random variable could look like.

Example 3.1 Let $\{\varepsilon_1, \varepsilon_2, \ldots\}$, $\{W_1, W_2, \ldots\}$, $\{\Gamma_1, \Gamma_2, \ldots\}$ be three independent series of random variables. $\varepsilon_1, \varepsilon_2, \ldots$ is an i.i.d. sequence of Rademacher[4] variables, $W_1, W_2, \ldots$ is an i.i.d. sequence of random variables with a finite absolute αth moment and $\Gamma_1, \Gamma_2, \ldots$ is a sequence of arrival times of

[4]Define the Rademacher variables as

$$\varepsilon_i = \begin{cases} 1 & \text{with probability } \frac{1}{2}, \\ -1 & \text{with probability } \frac{1}{2}. \end{cases}$$

a Poisson process with unit arrival rate $\lambda = 1$. Any random variable having the symmetric stable distribution $S_\alpha(\sigma, 0, 0)$, $0 < \alpha < 2$ (see section 3.4 for an explanation of the parameters) has the a.s. series representation

$$\sigma\left(\frac{C_\alpha}{\mathrm{E}|W_1|^\alpha}\right)^{\frac{1}{\alpha}} \sum_{i=1}^{\infty} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i, \tag{3.6}$$

where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin(x) \mathrm{d}x\right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\frac{\pi\alpha}{2})} & if\ \alpha \neq 1, \\ \frac{2}{\pi} & if\ \alpha = 1. \end{cases} \tag{3.7}$$

The representation from example 3.1 is for a symmetric stable distribution—see theorem 3.2 for a representation of the skewed and translated stable. Notice, that (3.6) converges absolutely for $\alpha < 1$. Thus we would expect much faster convergence for $\alpha < 1$ than for $\alpha > 1$.
The connection between the series representation of a random variable and the series representation of a stochastic process is given by Rosiński [26]. We can use example 3.1 to illustrate this. Let the stable random variable X be given a.s. by the series representation (3.6). Rosiński then uses a sequence $\{U_1, U_2, \ldots\}$ to transform the random variable into a stochastic process on $[0, 1]$. Let our sequence $\{U_1, U_2, \ldots\}$ replace the sequence $\{W_1, W_2, \ldots\}$ and for our stable random variable X we then have

$$X(t) = \sigma\left(\frac{C_\alpha}{\mathrm{E}|W_1|^\alpha}\right)^{\frac{1}{\alpha}} \sum_{i=1}^{\infty} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} 1_{(U_i \leq t)}, \quad a.s. \tag{3.8}$$

This may seem straightforward but it requires quite some work to achieve the important fact, that the series (3.8) converges a.s. uniformly on $[0, 1]$.
A series representation could be used for generating/simulating a Lévy jump process. It is of course not possible to sum infinitely many terms and inevitably we need to truncate to a finite number of terms. Since a series representation is not unique, we need to be aware of the error when truncating—there could be differences according to which distributions and representations that are being used, i.e. the speed of convergence varies with the type of distribution and representation. Also notice, that when truncating a representation of the same type as (3.6) we leave all the small jumps out of account! So actually this method can be compared with the method of just simulating the compound Poisson process $J_t^{(2)}$.
Note the completely different approach for simulating a process by the series approximation, contrary to the well known method of the i.i.d. increments. That is, we normally make use of the increment distribution and build up the process by iterating forward on a mesh. But this cannot by done for any Lévy process, because in general we do not know the distribution of the increments.

3.2 The normal approximation in simulation

We now want to test how the normal approximation from theorem 2.1 works in practice. As always it is hard to get results for a sample path $J_t(\omega)$, $t \in [0,1]$, so we will be looking at a fixed time, say $t = 1$. By doing this we also take advantage of being able to simulate J_1 directly—i.e. the distribution of a stable random variable can be simulated precisely (see Chambers, Mallows and Stuck [9] and Nolan [24]), but generation of a sample path is still only done by a limiting procedure. We will later (in section 4) take a closer look at the quality of the pathwise approximation, and for now believe, that the paths of our simulation are close to the real path (ω-wise) whenever they are close at the end of the interval $[0,1]$.
Quantiles are very useful when doing Q-Q plots, and for the stable distribution they can be approximated by a program[5] developed by Nolan. When looking at heavy tailed distributions, we need a lot of points in order to produce useful Q-Q plots. The tails are highly unpredictable, so we should definitely be cautious and not pay too much attention to the far out tails. We will try to overcome this problem a bit by choosing to simulate a very high number of variables. Furthermore we will also have a look at both the theoretical and simulated residuals.
It would be nice to compare a (truncated) series X^ϵ of the form (3.5) to the corresponding "real" random variable X. But in order to do so by simulation, we need to be able to generate both random variables for the same ω—which is impossible. The idea is now to simulate a large number, say K, of truncated series X^ϵ of the form (3.5), and then test different truncation levels ϵ against the "real" distribution—i.e. by calculating residuals and doing Q-Q plots. If we are able to get close enough to the real distribution, we will assume the truncated random variable X^ϵ has got the right (real) distribution. We can then easily get random variables for the same ω just by noting the value of $X^{\epsilon'}, \epsilon' > \epsilon$ at a lower summation level, corresponding to a higher truncation level.
As stated earlier, truncating a representation of the same type as (3.6) can be compared to simulating the compound Poisson process $J_t^{(2)}$. Recall from section 1.3, that the compound Poisson process has intensity $\lambda = \int \nu(\mathrm{d}x)$ and jump size distribution $\frac{\nu(\mathrm{d}x)}{\lambda}$. But the method of simulating a compound Poisson process, cannot be used in this situation, because (as before) we need to have both variables for the same ω.

3.3 Preparing for simulation

We will look at three different Lévy processes at time $t = 1$, i.e. three different distributions: stable, gamma and our own example made for the occasion. First we find the intensity λ_ϵ for (the compound Poisson process) $J_t^{(2)}$ and the quadratic variation of the small jumps σ_ϵ for all three Lévy

[5]The program is called STABLE and can by downloaded from Nolan's homepage http://academic2.american.edu/~jpnolan/home.html.

measures. We need λ_ϵ if we want to use the method of simulating a compound Poisson process—here it is mostly stated in order to have a clue of the actual intensity. We will use σ_ϵ when we later want to make use of the normal approximation. This can be done by simulating to a certain lower level of the series than we would normally do, and then adding an independent normally distributed random variable with the right variance given by σ_ϵ^2—according to theorem 2.1, this is the asymptotic variance of $X_\epsilon(1)$.
For the stable and gamma distributions we explicitly know a series representation, but for our own example we need to use a general result in order to get a representation. For this we will be using the inverse Lévy measure method by Khintchine [19], Ferguson and Klass [12]—this method and several others are clearly described and exemplified by Rosiński [26].
The stable process where $X(1)$ has a stable distribution $S_\alpha(\sigma, \beta, 0)$[6] with a Lévy measure of the form

$$\nu(\mathrm{d}x) = \frac{c_1}{x^{1+\alpha}} 1_{(x>0)} \mathrm{d}x + \frac{c_2}{|x|^{1+\alpha}} 1_{(x<0)} \mathrm{d}x, \tag{3.9}$$

where $c_1 \geq 0$, $c_2 \geq 0$, $c_1 + c_2 > 0$ and c_1, c_2 depend on σ and β. We now find

$$\begin{aligned} \lambda_\epsilon &= \int_{D(\epsilon,\infty)} \nu(\mathrm{d}x) = \int_{-\infty}^{-\epsilon} \frac{c_2}{|x|^{1+\alpha}} \mathrm{d}x + \int_\epsilon^\infty \frac{c_1}{x^{1+\alpha}} \mathrm{d}x \\ &= \left[\frac{c_2}{(-\alpha)x^\alpha}\right]_\epsilon^\infty + \left[\frac{c_1}{(-\alpha)x^\alpha}\right]_\epsilon^\infty = \frac{c_2}{\alpha\epsilon^\alpha} + \frac{c_1}{\alpha\epsilon^\alpha} = \frac{c_1 + c_2}{\alpha\epsilon^\alpha} \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} \sigma_\epsilon^2 &= \int_{-\epsilon}^{\epsilon} x^2 \nu(\mathrm{d}x) = \int_{-\epsilon}^{0} x^2 \frac{c_2}{|x|^{1+\alpha}} \mathrm{d}x + \int_0^\epsilon x^2 \frac{c_1}{x^{1+\alpha}} \mathrm{d}x \\ &= \left[\frac{c_2 x^{2-\alpha}}{2-\alpha}\right]_0^\epsilon + \left[\frac{c_1 x^{2-\alpha}}{2-\alpha}\right]_0^\epsilon = \frac{c_1 + c_2}{2-\alpha} \epsilon^{2-\alpha}. \end{aligned} \tag{3.11}$$

The calculations for the other two are similar and we find for the gamma process (see also example 2.5)

$$\nu(\mathrm{d}x) = a \frac{\mathrm{e}^{-\frac{x}{b}}}{x} \mathrm{d}x \quad a, b > 0,\ x > 0 \tag{3.12}$$

$$\lambda_\epsilon = a \int_\epsilon^\infty \frac{\mathrm{e}^{-\frac{x}{b}}}{x} \mathrm{d}x \tag{3.13}$$

$$\sigma_\epsilon^2 = ab\Big(b - (\epsilon + b)\mathrm{e}^{-\frac{\epsilon}{b}}\Big), \tag{3.14}$$

[6] This parameterization is the one used by Samorodnitsky and Taqqu [28]. It is now the most used in the literature even though the characteristic function is not jointly continuous in all four parameters. There are quite a few different parameterizations of the stable distribution that deal with this problem and further more give a more intuitive meaning to the parameters—these are explained in Nolan [23]. See also Zolotarev [32].

and for our own example

$$\nu(\mathrm{d}x) = \frac{1}{x\sqrt{x}}1_{(0<x<1)}\mathrm{d}x + \frac{\mathrm{e}^{-\log^2(x)}}{x}1_{(x\geq 1)}\mathrm{d}x, \quad x>0 \tag{3.15}$$

$$\lambda = \frac{2-2\sqrt{\epsilon}}{\sqrt{\epsilon}} + \int_1^\infty \frac{\mathrm{e}^{-\log^2(x)}}{x}\mathrm{d}x \tag{3.16}$$

$$\sigma_\epsilon^2 = \frac{2}{3}\epsilon\sqrt{\epsilon}. \tag{3.17}$$

Notice, that the remaining integrals have to be solved numerically.

3.4 Simulating stable

For the stable distribution $S_\alpha(\sigma,\beta,\mu)$ we do not explicitly know the two constants from the Lévy measure c_1 and c_2—if needed, the connection to the parameters can be calculated (not very pleasant) or estimated by simulation. We do know, that the process has no positive (respectively, no negative) jumps when $c_1 = 0$ (respectively, $c_2 = 0$). $\beta \in [-1,1]$ is a skewness parameter and we know the relation $\beta = (c_1 - c_2)/(c_1 + c_2)$. In order to keep things simple and restrict the number of simulations we will only look at the symmetric case, i.e. $\beta = 0$ and thus $c_1 = c_2$. Our results from simulating the symmetric case will also be usable for the skewed case—comments on this matter will be made at the end of section 3.5. The parameter $\sigma \geq 0$ is a scale parameter and is also determined by c_1 and c_2—we choose $\sigma = 1$. The last parameter in the triplet, μ, is a translation parameter and we choose $\mu = 0$. The index $\alpha \in (0,2)$ is the most significant parameter. Roughly speaking, a stable process moves mainly by big jumps if α is close to 0, and mainly by small jumps if α is close to 2. By looking at the simulation results in Clausen [10] we see a clear picture of the computational effort required to have (3.6) converging. The convergence becomes significantly slower for α's growing and especially for $\alpha > 1$ it gets very slow—for α's near 2 it might be impossible to get the series converging within a reasonable simulation time. With that in mind we choose the two values $\alpha = 0.75$ and $\alpha = 1.5$, and will have a look at both.

Our first concern is how many variables to generate in order to get a reasonable Q-Q plot for the stable distribution. This can be examined by simulating stable random variables by the method of Chambers, Mallows and Stuck [9], and then doing Q-Q plots against the real quantiles. This generation method is naturally much faster than summing in a series representation. As mentioned previously, we will use the STABLE program by Nolan to approximate[7] the real quantiles. When doing Q-Q plots with non-heavy tailed distributions we normally consider 5,000 observations to be enough to ensure a perfect Q-Q plot, but obviously this is far from enough for

[7]The program STABLE has a preset precision (relative error) when calculating quantiles. The relative error is $0.12 \cdot 10^{-13}$ and all our calculations will be done with this precision.

heavy tailed distributions. On the other hand we need to keep the number of variables to a level where we will be able to do our later simulations of the series representation. We will be using S-PLUS[8] for simulation of these stable random variables (S-PLUS uses the method of Chambers, Mallows and Stuck). Repeated runs indicate that an amount of 1 mill. random variables is adequate for our purpose—the runs still differ in the tails at this level, but we will take this into account when looking at the plots. Two runs are depicted in figure 3.1—it shows the full Q-Q plot and two more Q-Q plots of the same plot with truncated axes. Notice that, when truncating at ± 5 mill. and $\pm 50{,}000$ this corresponds to truncating at respectively the $3.5 \cdot 10^{-5}$th. and the $1.0 \cdot 10^{-4}$th. fractile on the negative side and equivalent values on the positive side—this leaves respectively about 6 and 250 observations out of account. When looking at the Q-Q plots we notice a systematic deviation in the plot truncated at 50,000. It looks as if the stable random variables generated by the method of Chambers, Mallows and Stuck has got a bit lighter tails than they should have! A further research by doing several runs show the same picture, but with both lighter, heavier and the combination of the two at either tail. This indicates, that either the Chambers, Mallows and Stuck method is not "perfect" due to numerical problems, though the method is exact, or the built-in random number generator in S-PLUS leaves room for improvement. Actually the S-PLUS random number generator fails several statistical tests—see McCullough [21].

We now turn to simulating stable random variables by the series representation (see example 3.1). We recall, that the idea is to simulate 1 mill. truncated series $X^{\epsilon}(1)$ of the form (3.6), and then test different truncation levels ϵ against the "real" distribution (for ease we choose W_i iid. uniform). This could be done in S-PLUS, but we would definitely want to avoid this because of the questionable random number generator and the extensive running time. S-PLUS is known to be very slow when doing loops[9], and thus we take an alternative approach by programming the simulation from scratch in the programming language Delphi[10]. We avoid using Delphi's build in random number generator (linear congruential—optimized only for speed) and instead we will use an assembler optimized version of an algorithm called ISAAC[11]—see Flannery et. al. [18] for more information on

[8] S-PLUS is a statistical analysis program by Insightful (former MathSoft)—see http://www.splus.com for more information.

[9] S-PLUS does not release memory after doing each loop and unfortunately (for our purpose) it keeps testing if it can be released. This tedious procedure makes loops not only very slow, but also has the side effect of increasing the running time to worse than linear.

[10] Delphi is a programming language by Borland and is often referred to as Object Pascal. The speed is comparable to C++. See http://www.borland.com for more information.

[11] ISAAC (Indirection, Shift, Accumulate, Add, and Count) is a random number generator developed by Robert J. Jenkins Jr. There are no cycles in ISAAC shorter than 2^{40} (i.e. no bad seeds) values and the expected cycle length is 2^{8295} values. No bias has been detected in this algorithm (contrary to the linear congruential generators) and it passes all the regular statistical tests for random number generators, including DIEHARD. More information on ISAAC can be found at http://burtleburtle.net/bob/rand/isaac.html.

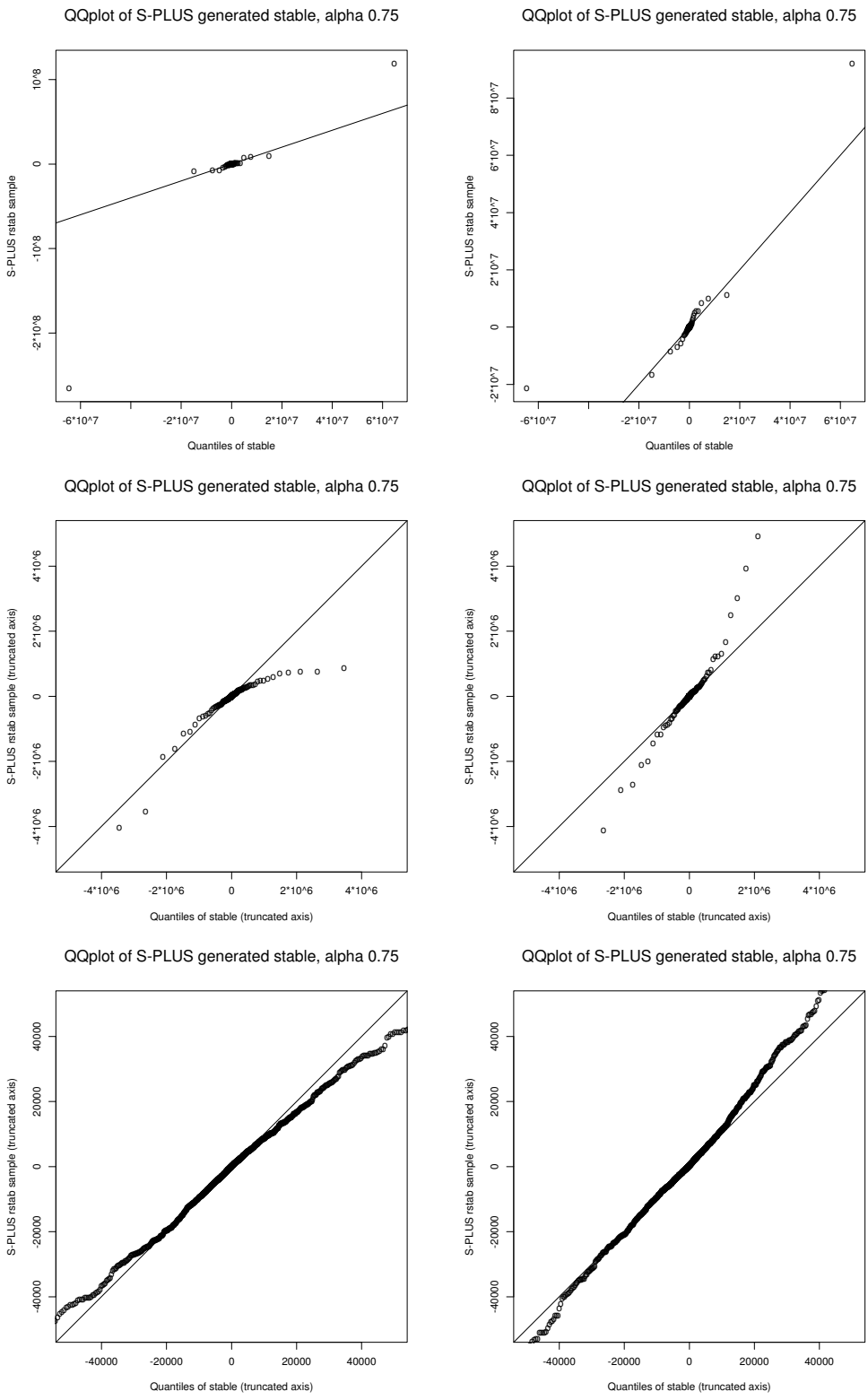

Figure 3.1: Two separate simulations (one in each column) of 1 mill. stable random variables with $\alpha = 0.75$ in S-PLUS with the built-in rstab function.

random number generators. We know, that we should get a simulation of better quality from our own program with the ISAAC generator, and we presume that it is also much faster. We have done a test by running the S-PLUS code by Clausen [10] and our program with the same parameters on the same computer (a 850Mhz P-III). The test generates 100 independent stable random variables with index 0.75 by summing 50,000 terms in the series representation for each variable. The running time for the Clausen program was 65 minutes and for our program 5 seconds—thus our program is 780 times faster! This certainly tells us to avoid S-PLUS when doing very large simulations.

We now want to determine the level, say N, of terms we need to sum in the series representation in order to assume our variable has got the "real" distribution. This can be done by looking at the theoretical variance of the residuals, i.e. the variance of the sum of terms from $N+1$ to $\widetilde{N}$ in the series representation (the mean is of course 0). In the following we will make use of Stirling's formula

$$\Gamma(i) = \sqrt{2\pi} i^{i-\frac{1}{2}} \mathrm{e}^{-i} q(i), \ i > 0, \tag{3.18}$$

where $1 < q(i) < \mathrm{e}^{\frac{1}{12i}}$, and the following moments of the Gamma distributed variable

$$\mathrm{E}\left(\Gamma_i^{-\frac{2}{\alpha}}\right) = \int_0^\infty x^{-\frac{2}{\alpha}} \frac{1}{\Gamma(i)} x^{i-1} \mathrm{e}^{-x} \mathrm{d}x = \frac{\Gamma\left(i - \frac{2}{\alpha}\right)}{\Gamma(i)}, \tag{3.19}$$

for approximating the theoretical variance of the residuals. We follow the same idea as used in Clausen [10] and have

$$\begin{aligned}
\mathrm{Var}\left[\sum_{i=N+1}^{\widetilde{N}} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i\right] &= \mathrm{E}\left(\sum_{i=N+1}^{\widetilde{N}} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i\right)^2 - \left(\mathrm{E}\sum_{i=N+1}^{\widetilde{N}} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i\right)^2 \\
&= \mathrm{E}\left(2 \sum_{j=N+2}^{\widetilde{N}} \sum_{i=N+1}^{j-1} \varepsilon_i \varepsilon_j (\Gamma_i \Gamma_j)^{-\frac{1}{\alpha}} W_i W_j\right) + \\
&\quad \mathrm{E}\left(\sum_{i=N+1}^{\widetilde{N}} \left(\varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i\right)^2\right) - 0 \\
&= \sum_{i=N+1}^{\widetilde{N}} \mathrm{E}\left(\Gamma_i^{-\frac{2}{\alpha}}\right) \mathrm{E}(W_i^2) = \frac{1}{3} \sum_{i=N+1}^{\widetilde{N}} \frac{\Gamma\left(i - \frac{2}{\alpha}\right)}{\Gamma(i)} \\
&= \frac{1}{3} \sum_{i=N+1}^{\widetilde{N}} \frac{\sqrt{2\pi}\left(i - \frac{2}{\alpha}\right)^{i-\frac{2}{\alpha}-\frac{1}{2}} \mathrm{e}^{-\left(i-\frac{2}{\alpha}\right)} q\left(i - \frac{2}{\alpha}\right)}{\sqrt{2\pi} i^{i-\frac{1}{2}} \mathrm{e}^{-i} q(i)} \\
&\sim \frac{1}{3} \sum_{i=N+1}^{\widetilde{N}} \left(1 - \frac{2}{\alpha i}\right)^{i-\frac{1}{2}} \left(i - \frac{2}{\alpha}\right)^{-\frac{2}{\alpha}} \mathrm{e}^{\frac{2}{\alpha}}
\end{aligned}$$

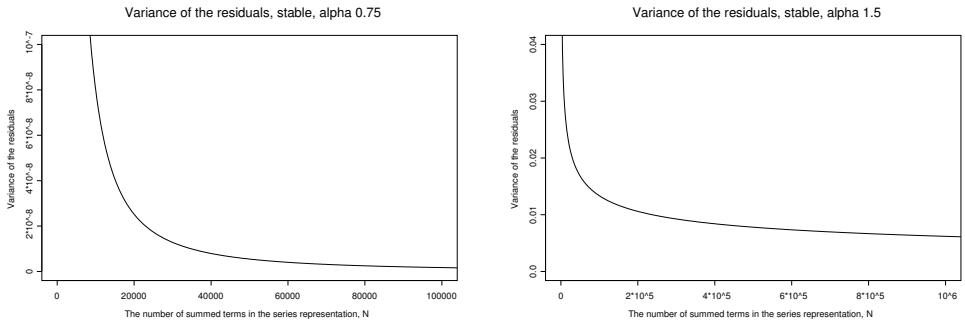

Figure 3.2: Variance of the residuals $X^{\text{res}}_{N,\infty}$ ($\widetilde{N} = \infty$) for our two values of α in the case of stable (beware of the highly different axes).

$$\begin{aligned} &\sim \frac{1}{3}\sum_{i=N+1}^{\widetilde{N}} \left(i - \frac{2}{\alpha}\right)^{-\frac{2}{\alpha}} \sim \frac{1}{3}\sum_{i=N+1}^{\widetilde{N}} i^{-\frac{2}{\alpha}} \\ &\sim \frac{1}{3}\frac{N^{1-\frac{2}{\alpha}} - \widetilde{N}^{1-\frac{2}{\alpha}}}{\frac{2}{\alpha}-1}, \end{aligned} \tag{3.20}$$

and therefore

$$\begin{aligned} \operatorname{Var}\big(X^{\text{res}}_{N,\widetilde{N}}\big) &= \operatorname{Var}\left[\sigma\left(\frac{C_\alpha}{\mathrm{E}|W_1|^\alpha}\right)^{\frac{1}{\alpha}} \sum_{i=N+1}^{\widetilde{N}} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i\right] \\ &\sim \frac{\sigma^2}{3}\left(\frac{C_\alpha}{\mathrm{E}|W_1|^\alpha}\right)^{\frac{2}{\alpha}} \frac{N^{1-\frac{2}{\alpha}} - \widetilde{N}^{1-\frac{2}{\alpha}}}{\frac{2}{\alpha}-1}. \end{aligned} \tag{3.21}$$

If we let $\widetilde{N} \to \infty$ we have

$$\operatorname{Var}\big(X^{\text{res}}_{N,\infty}\big) \sim \frac{\sigma^2}{3}\left(\frac{C_\alpha}{\mathrm{E}|W_1|^\alpha}\right)^{\frac{2}{\alpha}} \frac{N^{1-\frac{2}{\alpha}}}{\frac{2}{\alpha}-1}. \tag{3.22}$$

For computational reasons we choose W_i as iid uniform(0,1), as it is the fastest to generate. We can set the levels of N for our two choices of α, and in order to clearly see the dependence on N, we choose to plot $X^{\text{res}}_{N,\infty}$ against N in figure 3.2. For $\alpha = 0.75$ the variance quickly gets very small and for $N = 20,000$ it is as little as $2.5 \cdot 10^{-8}$. This is negligible and we can definitely assume our variables to have the stable distribution when summing 20,000 terms in the series representation. The picture is somewhat different for $\alpha = 1.5$, as the variances are much bigger. We do not have the same kind of convergence and will just choose the highest possible number which can be simulated in a reasonable amount of time. By running a couple of tests with our program, we can see that $N = 500,000$ is by far the highest value we can consider—we need to run the program several times, so larger values of N are not an option. For our value of $N = 500,000$ the variance is $7.8 \cdot 10^{-3}$. On our 850Mhz P-III this simulation would take just over 6 days, so we will

stable, $\alpha = 0.75$			stable, $\alpha = 1.5$		
M	$\mathrm{Var}(X^{\mathrm{res}}_{M,N})$	$\mathrm{Var}(X^{\mathrm{res}}_{N,\infty})$	M	$\mathrm{Var}(X^{\mathrm{res}}_{M,N})$	$\mathrm{Var}(X^{\mathrm{res}}_{N,\infty})$
10	$8.0 \cdot 10^{-3}$	$8.0 \cdot 10^{-3}$	100	$1.3 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$
25	$1.7 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	1,000	$5.4 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$
100	$1.7 \cdot 10^{-4}$	$1.7 \cdot 10^{-4}$	10,000	$2.1 \cdot 10^{-2}$	$2.9 \cdot 10^{-2}$
500	$1.2 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$	100,000	$5.5 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$
5,000	$2.3 \cdot 10^{-7}$	$2.5 \cdot 10^{-7}$	250,000	$2.0 \cdot 10^{-3}$	$9.8 \cdot 10^{-3}$

Table 3.1: Theoretical variance of the residuals at the levels $M < N$. For the two values of $\alpha = 0.75$ and $\alpha = 1.5$, N is respectively 20,000 and 500,000.

be lending a few faster computers for this.
The next step is to choose the levels $M < N$ for which the current value of the sum in the series representation is noted. These levels need to be considerably lower, as we expect the normal approximation to do a good job for us. We do not have a clue of how these levels should be chosen, as this is one of the reasons for doing this thesis. By looking at the theoretical variance of the residuals at some lower levels $M < N$ and doing a few test runs, we choose 5 values of M and will compare these—see table 3.1.
After several runs of our program with both α's we compare the runs by looking at Q-Q plots. The Q-Q plots of the random variables generated by S-PLUS (the method of Chambers, Mallows and Stuck) showed systematic deviations in the plot truncated at $\pm$50,000—the deviation changed from one run to another but was clearly systematic. We are of course looking as to whether this is also the case in our simulations. Some of our simulations also show some deviations from the straight line through the Q-Q plot, but they are much smaller than for the ones generated by S-PLUS. The deviations in our plots do not seem to be systematic and they are not growing bigger towards the edges of the plot (at least not in the plot with the smallest truncation) but are rather fluctuating around the straight line. Two runs are depicted in figure 3.3—it shows the same types of Q-Q plots as in figure 3.1. The truncation levels for the plots of $\alpha = 0.75$ in figure 3.3 are chosen to match those from figure 3.1, i.e. truncation is done at the same fractiles (the actual fractiles are stated earlier in this section). We will not pay too much attention to the tails in the Q-Q plots as they are highly fluctuating due to the heavy tails in the stable distribution—we did not expect them to match and they certainly do not. All in all we are pleased with our simulations, as they show good signs of matching the real stable distribution, that is, a good convergence of the series representation. But we will keep the highly different variance of the residuals in mind, and watch if they affect the final outcome when we test the normal approximation.

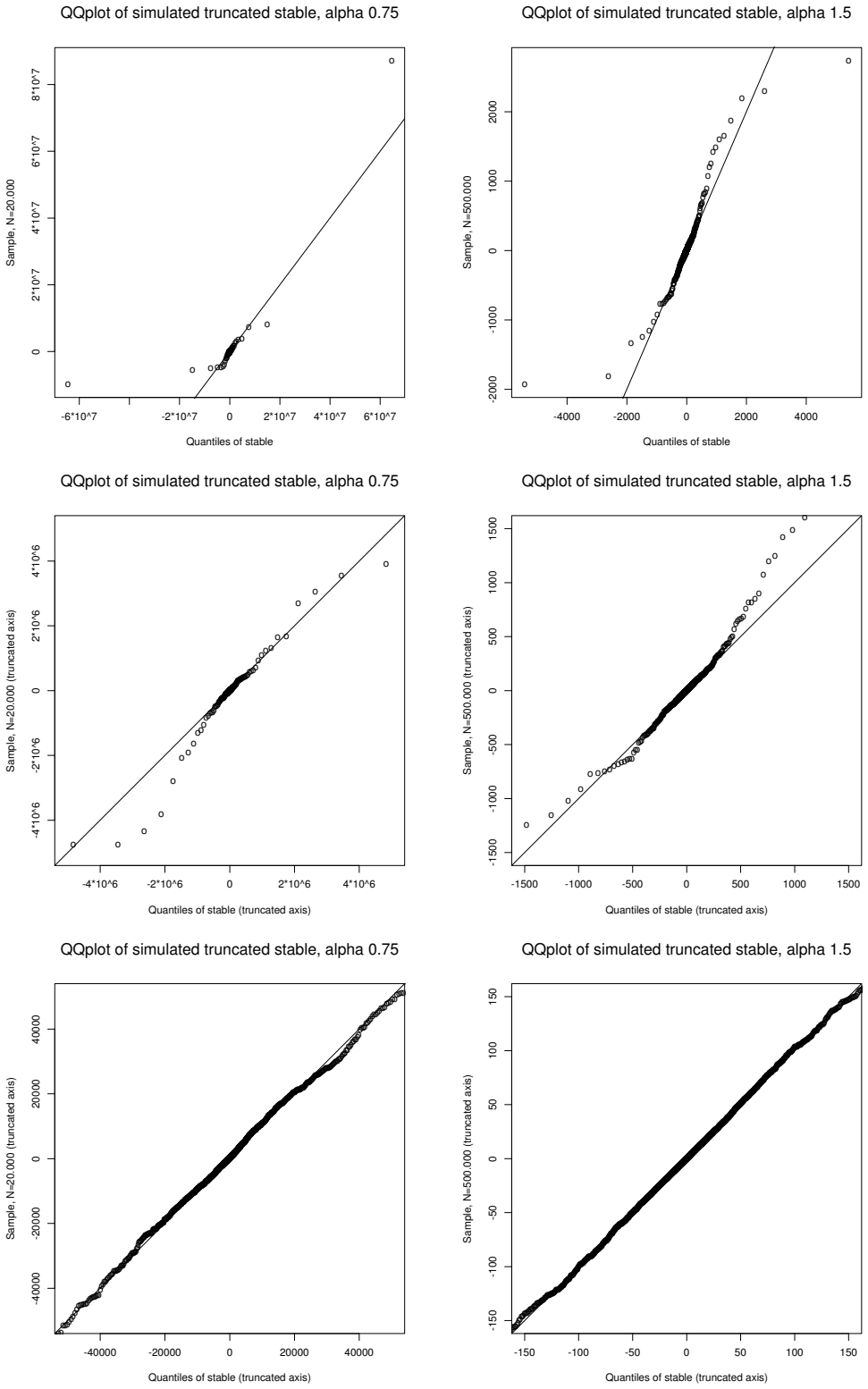

Figure 3.3: Two separate simulations (one in each column) of 1 mill. stable random variables (truncated series representation)—by our own program.

3.5 The normal approximation in the case of stable

From example 2.4 we know that the normal approximation holds for the one sided strictly $\frac{1}{2}$-stable process. Now recall $\sigma_\epsilon^2 = \frac{c_1+c_2}{2-\alpha}\epsilon^{2-\alpha}$ from (3.11)—ν from (3.9) does not have any atoms and therefore by proposition 2.2, the normal approximation holds. We will now have a closer look at the behaviour of the normal approximation in practice.

We previously concluded, that the variance of the residuals (towards infinity) was small enough to assume, that we had a real stable random variable by summing 20,000 terms in the series representation with $\alpha = 0.75$. We were more skeptical in the case of $\alpha = 1.5$ as the residuals were much bigger. We will now test the normal approximation by doing Q-Q plots (versus standard Normal) of the difference between summing N terms and summing M terms, i.e. the residuals $X_{M,N}^{\text{res}}$. We have previously used a different notation for this difference, which expressed the difference in terms of the truncation level ϵ—that is, in these terms we would say that the difference $X_\epsilon \sim X^{\epsilon'} - X^\epsilon,\ \epsilon' < \epsilon$ between the two dependent random variables should follow a normal distribution. There is, of course, a unique correspondence between these two notations and more on this topic will be given at the end of section 3.6. The main idea is to find out how many terms are needed in order to make full use of the normal approximation, i.e. by summing M terms in the series representation and adding a Normal random variable. The Q-Q plots are depicted in figure 3.4.

The Q-Q plots for the case of $\alpha = 0.75$ show how summing additional terms improves the normal approximation. For $M = 10$ we easily see that the residuals are not Normal. The improvements over $M = 25$ and $M = 100$ are indeed noticeable, and from $M = 500$ the residuals are almost indistinguishable from the Normal distribution. This is indeed very useful as we know now, that generating a stable random variable with index $\alpha = 0.75$ can be done by summing 500 terms in the series representation and adding a Normal random variable with mean 0 and variance $1.2 \cdot 10^{-5}$—the variance of $X_{N,\infty}^{\text{res}}$, see (3.22). Because the running time of our program is linear, we are now able to generate a stable random variable (by the method of series representation) $20,000/500 = 40$ times faster.

For $\alpha = 1.5$ the picture is completely different—all plots are showing residuals ($X_{M,N}^{\text{res}}$) that are close to Normal. We recall, that the theoretical residuals (table 3.1) for the case of $\alpha = 1.5$ are much bigger than the ones for the case of $\alpha = 0.75$, i.e. we know that the convergence of the series representation for $\alpha = 1.5$ is much slower. That is, we expect that M should be much larger in this case in order to have small (and Normal) residuals towards $N = \infty$ (i.e. $X_{M,\infty}^{\text{res}}$). By comparison to the case of $\alpha = 0.75$ we can then rule out, that the residuals from any $M < N$ to infinity are Normal—we can only conclude, that our $N = 500,000$ is not high enough. This means, that the residuals $X_{N,\infty}^{\text{res}}$ are too big for us to assume, that summing N terms in the series representation is enough to have a "real" stable random variable. In order to use the normal approximation in the case of $\alpha = 1.5$ we could

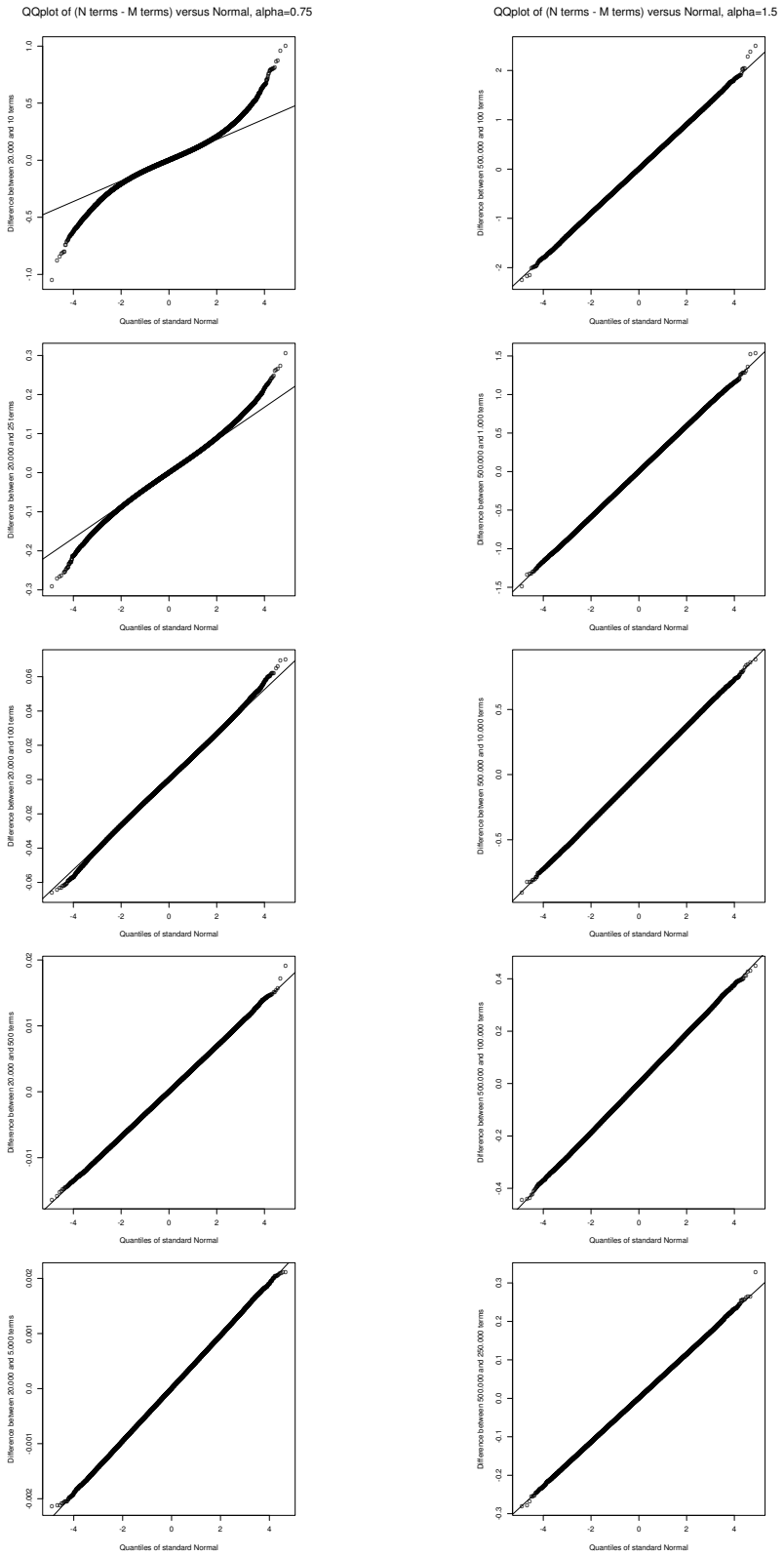

Figure 3.4: Q-Q plots of the difference between summing N terms and summing M terms in the series representation for stable.

argue, that we should have the same small variance of residuals ($X^{\text{res}}_{M,\infty}$) as for $\alpha = 0.75$. We concluded, that for $\alpha = 0.75$ we needed $M = 500$, i.e. variance of the residuals $X^{\text{res}}_{M,\infty}$ around $1.9 \cdot 10^{-5}$—some backward calculations then gives us, that we need $M = 4.0 \cdot 10^{15}$ for $\alpha = 1.5$ in order to have residuals $X^{\text{res}}_{M,\infty}$ of the same order. Obviously, we do not have that kind of enormous computer power available—actually it would take 8 billion times the time we used for running the case $\alpha = 1.5$ with $N = 500,000$. This also explains why our Q-Q plots all showed Normal residuals— they were only showing a tiny residual in terms of the "full" sum.

3.6 Further comments on stable

Here we have only looked at the symmetric stable case, but the normal approximation is valid for any stable, i.e. also the skewed and translated stable. If we follow the lines from above and argue, that we should have the same small variance of residuals ($X^{\text{res}}_{M,\infty}$) for any stable, as for the symmetric stable with $\alpha = 0.75$, then we are able to calculate the magnitude of M. We first state a series representation for any stable random variable, $S_\alpha(\sigma, \beta, \mu)$.

Theorem 3.2 (Samorodnitsky and Taqqu [28] theorem 1.4.5) *The series*

$$\sum_{i=1}^{\infty} \left(\Gamma_i^{-\frac{1}{\alpha}} W_i - k_i^{(\alpha)} \right) \tag{3.23}$$

with

$$k_i^{(\alpha)} = \begin{cases} 0 & if\ 0 < \alpha < 1, \\ \mathrm{E}\Big(W_1 \int_{|W_1|/i}^{|W_1|/(i-1)} x^{-2} \sin x \mathrm{d}x \Big) & if\ \alpha = 1, \\ \frac{\alpha}{\alpha - 1} \Big(i^{\frac{\alpha-1}{\alpha}} - (i-1)^{\frac{\alpha-1}{\alpha}} \Big) \mathrm{E} W_1 & if\ \alpha > 1, \end{cases}$$

$i = 1, 2, \ldots$, *converges a.s. to a* $S_\alpha(\sigma, \beta, 0)$ *random variable with*

$$\sigma^\alpha = \frac{\mathrm{E}|W_1|^\alpha}{C_\alpha} \qquad and \qquad \beta = \frac{\mathrm{E}|W_1|^\alpha \mathrm{sign} W_1}{\mathrm{E}|W_1|^\alpha},$$

where C_α *is defined as in (3.6).*

We can now use (3.23) to generate any stable random variable, because if $X \sim S_\alpha(\sigma, \beta, 0)$ then $\mu + X \sim S_\alpha(\sigma, \beta, \mu)$. In order to calculate the variance of the residuals

$$X^{\text{res}}_{N,\widetilde{N}} = \sum_{i=N+1}^{\widetilde{N}} \left(\Gamma_i^{-\frac{1}{\alpha}} W_i - k_i^{(\alpha)} \right) \tag{3.24}$$

in this general case, we need moments of the following type, $j > i$

$$\begin{aligned}\mathrm{E}\big[\Gamma_i\Gamma_j\big]^{-\frac{1}{\alpha}} &= \mathrm{E}\left[\sum_{k=1}^{i} E_k \sum_{k=1}^{j} E_k\right]^{-\frac{1}{\alpha}} \\ &= \mathrm{E}\left[\left(\sum_{k=1}^{i} E_k\right)^2 + \sum_{k=1}^{i} E_k \sum_{k=i+1}^{j} E_k\right]^{-\frac{1}{\alpha}} = \mathrm{E}\big[\Gamma_i^2 + \Gamma_i\Gamma'_{j-i}\big]^{-\frac{1}{\alpha}},\end{aligned}$$

where $E_1, E_2, \ldots$ is an i.i.d. sequence of standard exponential variables and Γ, Γ' are independent. These non-trivial moments can only be found by numerical integration—see appendix A.1 for the distribution of $[\Gamma_i^2 + \Gamma_i\Gamma'_{j-i}]$. This is very tedious and in practice one would choose an easier approach. Having the results for the symmetric case in mind, one could run some simulations and then try to get the same small (empirical) variance $\mathrm{Var}\big(X^{\mathrm{res}}_{M,N}\big)$, while making sure the series really converges by trying different values of M and N. This is the same procedure we will use for simulating our own example, see section 3.7.
We still need to state the connection between the truncation level ϵ of the Lévy measure (the notation used in section 2.1) and the number of summed terms in the series representation. Let us truncate the sum at the Nth term, and find the corresponding level of ϵ. We will use the symmetric stable case in order to keep our intuition. We recall the terms of the series representation from (3.6) and have

$$\begin{aligned}N(\epsilon) &= \#\left\{i : \left|\sigma\left(\frac{C_\alpha}{\mathrm{E}|W_1|^\alpha}\right)^{\frac{1}{\alpha}} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i\right| > \epsilon\right\} \\ &= \#\left\{i : \Gamma_i \le \left[\frac{\epsilon}{\sigma\left(\frac{C_\alpha}{\mathrm{E}|W_1|^\alpha}\right)^{\frac{1}{\alpha}} W_i}\right]^{-\alpha}\right\}. \qquad (3.25)\end{aligned}$$

This connection can be used for getting an intuitive idea about the level of N for fixed ϵ and vice versa, but because of the W_i's we cannot say much about the actual distribution of $N(\epsilon)$. This is a bit disappointing, as this connection would be very useful for determining the real variance σ_ϵ^2 of the small jump part. Instead we will have to be content with the empirical variance of the residuals.

3.7 Simulating Gamma

For the Gamma distribution the job will be a lot easier—we know from Clausen [10] that the series representation for the Gamma case converges much faster than for the stable, and that the convergence gets slower when increasing the shape parameter. The Gamma distribution has two parameters, a shape parameter a and a scale parameter b. We choose the two values $a = 10$ and $a = 500$ in order to have one small parameter together with a

bigger parameter with a slower convergence. In both cases the scale parameter is kept fixed at $b = 1$. We recall from example 2.5, that the Gamma distribution did not meet the requirements for the normal approximation—but later in section 2.1 we also saw, that in some sense it was very close indeed. In this section we will see whether the approximation can be used. The Gamma distribution is not heavy tailed, so we will save a lot of computer power compared to the case of stable. We consider 5,000 observations to be enough to determine the distribution completely, so this is the number of Gamma random variables we will simulate. We will use the same programming language and the same random number generator as for stable. Rosiński [26] has listed several series representations for the Gamma process and we will use the one based on Bondesson's idea (obtaining a Gamma random variable as a shot noise variable, see Bondesson [8]). Simulation of Gamma processes (and fields) by the inverse Lévy measure, as will be done with our own example in section 3.9, is treated by Ickstadt and Wolpert [13]—they focus on computing posterior distributions in nonparametric hierarchical Bayes statistics by extensive use of S-PLUS. A Gamma process $X(t)$ at time $t = 1$ can be represented as

$$X(1) = \sum_{i=1}^{\infty} b\mathrm{e}^{-\frac{\Gamma_i}{a}} E_i, \tag{3.26}$$

where $E_1, E_2, \ldots$ is an i.i.d. sequence of standard exponential random variables that are independent of Γ_i (Γ_i is defined the same way as for the representation of stable, see example 3.1). We will again use the theoretical residuals as a measure of what levels of N and M to choose. We assume $b = 1$ and from Clausen [10] we have

$$\begin{aligned} \mathrm{Var}\big(X^{\mathrm{res}}_{N,\widetilde{N}}\big) &= 2\sum_{j=N+2}^{\widetilde{N}} \sum_{i=N+1}^{j-1} \left(1+\frac{1}{a}\right)^{-j} \left(\frac{2+a}{1+a}\right)^{-i} + \\ &\quad 2\sum_{i=N+1}^{\widetilde{N}} \left(1+\frac{2}{a}\right)^{-i} - \left(\sum_{i=N+1}^{\widetilde{N}} \left(1+\frac{1}{a}\right)^{-i}\right)^2. \end{aligned} \tag{3.27}$$

As we (will) see in table 3.2, the variances quickly get small and we will not bother to find the variances as $\widetilde{N} \to \infty$. This is no surprise, as (3.26) only has positive terms. When choosing the different levels of M, we will have the corresponding levels of $\mathrm{Var}\big(X^{\mathrm{res}}_{M,N}\big)$ from the case of stable (with $\alpha = 0.75$) in mind and try to find similar levels.

Some test calculations with different levels of M and N give us an idea of how big the level of N should be for our two choices of a. For $a = 10$ we have a very fast convergence and find $N = 1,000$ is more than enough (e.g. $\mathrm{Var}\big(X^{\mathrm{res}}_{1,000;1,100}\big) = 6.915 \cdot 10^{-78}$ and $\mathrm{Var}\big(X^{\mathrm{res}}_{1,000;5,000}\big) = 6.916 \cdot 10^{-78}$). Similarly we find that $N = 10,000$ is enough for the case $a = 500$. We have plotted the two simulations of the truncated series representation in figure 3.5. As we expected (from the negligible variances in table 3.2), the Q-Q plots

Gamma, $a = 10$		Gamma, $a = 500$	
M	$\mathrm{Var}\big(X^{\mathrm{res}}_{M,N}\big)$	M	$\mathrm{Var}\big(X^{\mathrm{res}}_{M,N}\big)$
40	$2.3 \cdot 10^{-2}$	3,000	$2.0 \cdot 10^{-2}$
60	$7.8 \cdot 10^{-4}$	4,000	$4.9 \cdot 10^{-4}$
80	$2.5 \cdot 10^{-5}$	5,000	$1.1 \cdot 10^{-5}$
100	$7.4 \cdot 10^{-7}$	6,000	$2.4 \cdot 10^{-7}$

Table 3.2: Theoretical variance of the residuals at the levels $M < N$. For the two values of $a = 10$ and $a = 500$ N is respectively 1,000 and 10,000.

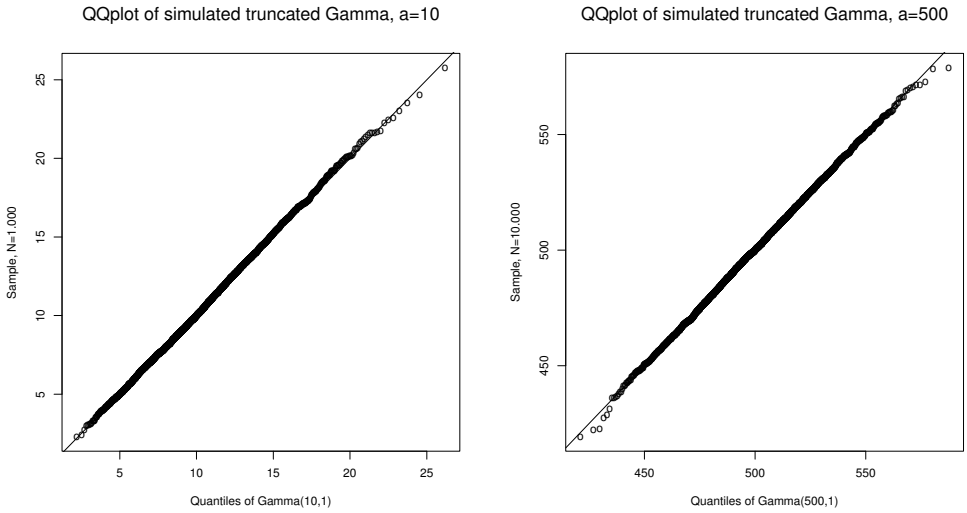

Figure 3.5: Two separate simulations of 5,000 Gamma random variables (truncated series representation)—by our own program.

show distributions that are almost indistinguishable from the Gamma, and therefore our simulations have a very good match to the Gamma distribution. Thus we can assume that we have a Gamma distributed variable by summing respectively $N = 1,000$ and $N = 10,000$ terms in the series representation for our two different parameters ($a = 10$, $a = 500$). This also means, that we have optimal conditions for determining whether the normal approximation is a useful tool in the Gamma case, even though the Gamma distribution did not meet the requirements of the normal approximation.

3.8 The normal approximation in the case of Gamma

We have, as for the case of stable, plotted the differences between summing N terms and summing M terms, i.e. the residuals $X^{\mathrm{res}}_{M,N}$, against the corresponding Normal quantiles. The Q-Q plots are depicted in figure 3.6.
All the residuals are strictly positive due to the series representation (3.26) only consisting of strictly positive terms. Theoretically this does not match

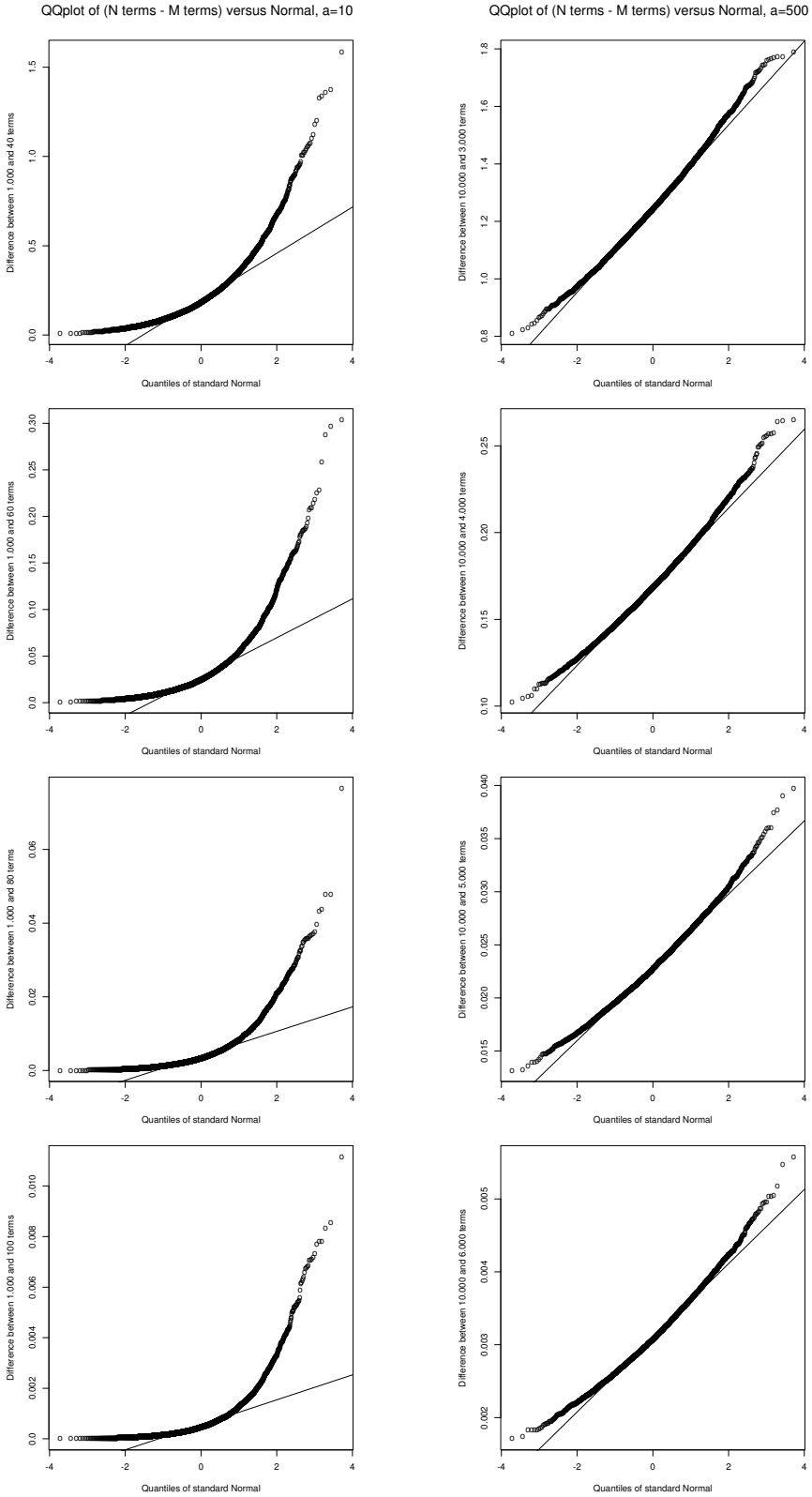

Figure 3.6: Q-Q plots of the difference between summing N terms and summing M terms in the series representation for Gamma.

the properties of the Normal distribution, but in practice we could still have a good approximation, because of the light tails of the Normal distribution. None of the Q-Q plots shows residuals that are Normal, and neither of our two parameters are showing improvement to a Normal behavior when we increase M. Even though the values get smaller and smaller when increasing M, the shape of the Q-Q plots is still kept. It is now obvious, that the normal approximation is not a good tool in the Gamma case—even though we saw in section 2.1, that the residuals and the Normal distribution had the first two moments in common. When compared to the Normal distribution, the simulated distributions have lighter tails at the lower end and heavier tails at the upper end of the distribution. The larger parameter ($a = 500$) has a shape closer to Normal than the smaller parameter ($a = 10$). Intuitively, this is because the Gamma distribution approaches the Normal distribution when the shape parameter increases.
Even though the Normal approximation is not a good tool in the Gamma case, we can still use the series representation in this case—the convergence of the series representation is quite fast and due to the Gamma distribution not being heavy tailed, we often do not need to generate a lot of random variables when simulating this distribution. For reference we state, that our program generated the 5,000 'Gamma' random variables (summing 10,000 terms in the series representation for each variable) in about half a minute.

3.9 Simulating our own example

The last of our three Lévy processes from section 3.3 is our own example. The distribution of $X(1)$ is not well known, as for the previous two processes, and we have only specified the process by its Lévy measure (3.15). This means, that we will have to use a general result in order to obtain a series representation. As mentioned in section 3.3, we will be using the inverse Lévy measure method (see section 3.3 for references). We recall, that ν from (3.15) does not have any atoms, and by proposition 2.2 the Normal approximation holds.
We define the generalized inverse of the tail of ν by

$$\nu^{\leftarrow}(u) = \inf\{x > 0 : \nu([x, \infty[) < u\}, \quad u > 0, \tag{3.28}$$

and Rosinski [26] gives us the series representation

$$J_t = \sum_{i=1}^{\infty} \nu^{\leftarrow}(\Gamma_i) 1_{(U_i \leq t)} \quad a.s., \quad t \in [0, 1], \tag{3.29}$$

where Γ_i and U_i are defined the same way as in section 3.1. Define the modified exponential integral $\mathrm{E}_{\log^2,k}(x) = \int_{\max\{x,1\}}^{\infty} \frac{\mathrm{e}^{-\log^2(u)}}{u^k} \mathrm{d}u$, and we have

$$\nu([x, \infty[) \quad = \quad \int_x^1 \frac{1}{u\sqrt{u}} \mathrm{d}u 1_{(x<1)} + \int_{\max\{x,1\}}^{\infty} \frac{\mathrm{e}^{-\log^2(u)}}{u} \mathrm{d}u$$

$$= \left[\frac{1}{-\frac{1}{2}\sqrt{u}}\right]_x^1 1_{(x<1)} + \mathrm{E}_{\log^2,1}(\max\{x,1\})$$
$$= 2\left(\frac{1}{\sqrt{x}} - 1\right) 1_{(x<1)} + \mathrm{E}_{\log^2,1}(\max\{x,1\}). \quad (3.30)$$

The inverse of $x = 2\left(\frac{1}{\sqrt{x}} - 1\right)$ is $y = 1/\left(\frac{x}{2} + 1\right)^2$, and thus the (generalized) inverse of ν is

$$\nu^{\leftarrow}(y) = \frac{1_{\left(y>\mathrm{E}_{\log^2,1}(1)\right)}}{\left(\frac{y-\mathrm{E}_{\log^2,1}(1)}{2}+1\right)^2} + \mathrm{E}^{-1}_{\log^2,1}(y) 1_{\left(y\leq \mathrm{E}_{\log^2,1}(1)\right)}. \quad (3.31)$$

By inserting (3.31) in (3.29) we end up with our a.s. series representation of J_t, $t \in [0,1]$

$$J_t = \sum_{i=1}^{\infty} \left(\frac{1_{\left(\Gamma_i>\mathrm{E}_{\log^2,1}(1)\right)}}{\left(\frac{\Gamma_i-\mathrm{E}_{\log^2,1}(1)}{2}+1\right)^2} + \mathrm{E}^{-1}_{\log^2,1}(\Gamma_i) 1_{\left(\Gamma_i\leq \mathrm{E}_{\log^2,1}(1)\right)} \right) 1_{(U_i\leq t)}. \quad (3.32)$$

The obvious difficulty in the series representation (3.32) is to calculate the inverse of our modified exponential integral. Not only will we need to approximate the integral by numerical integration, but we also need to approximate in order to find its inverse. This is very time consuming, and luckily, the number of times it needs to be done is indeed very limited for the generation of each random variable—notice $\mathrm{E}_{\log^2,1}(1) \simeq 0.89$.
As for the previous two processes, we will only look at time $t = 1$. We will again use the same programming language and the same random number generator as for stable. The integral approximation is programmed using Simpson's rule—see e.g. [1].
In this general case we do not assume anything about the simulation levels in advance. We start out by simulating J_1 from (3.32) with different levels of truncation N, and noting the values for different (much lower) levels $M < N$. We will try to get similar variances of the residuals, as for the case of stable random variables with $\alpha = 0.75$, see table 3.1. Our previous results show, that it would be wise to choose the largest value of M to be ten times smaller than N—by Q-Q plotting the residuals against the normal distribution, we can then see how large we need N, in order to have normal residuals. This can be done with a fairly small set of generated variables, but we still need an idea of how many variables to generate. This, of course, depends on whether the distribution of J_1 is heavy tailed or not. There are many degrees of heaviness, and by doing Q-Q plots against different known distributions with different amounts of variables, we are able to have an idea about the heaviness of the tail. In figure 3.7 we have Q-Q plotted 5,000 variables against the corresponding exponential quantiles—a solid (auxiliary) line with slope 2 is also drawn in the figure. The figure shows, that the heaviness of the tail of J_1 is comparable to the tail of the

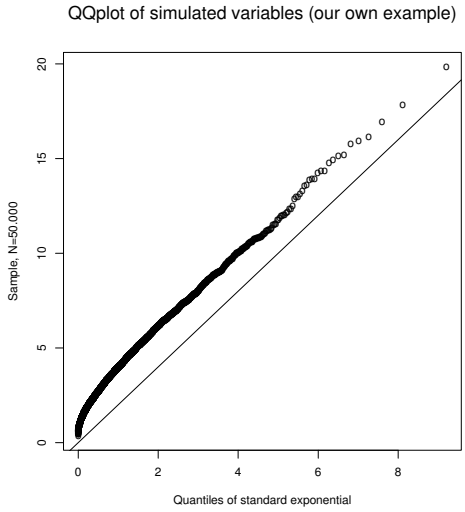

Figure 3.7: Q-Q plot of our own example against standard exponential.

M	$\mathrm{E}(X_{M,N}^{\mathrm{res}})$	$\mathrm{Var}(X_{M,N}^{\mathrm{res}})$
10	$3.9 \cdot 10^{-1}$	$2.1 \cdot 10^{-2}$
25	$1.6 \cdot 10^{-1}$	$1.4 \cdot 10^{-3}$
100	$4.0 \cdot 10^{-2}$	$2.2 \cdot 10^{-5}$
500	$7.9 \cdot 10^{-3}$	$1.7 \cdot 10^{-7}$
5,000	$7.2 \cdot 10^{-4}$	$1.7 \cdot 10^{-10}$

Table 3.3: Empirical mean and variance of the residuals at the levels $M < N$. The value of N is 50,000.

exponential distribution. Thus, generating 5,000 variables should be enough for this distribution.
In table 3.3 the empirical mean and variance of the residuals for our 5 chosen levels of M are shown (for 5,000 variables). When we compare the levels in table 3.3 with our two previous processes, we are convinced, that the distribution is well approximated at $N = 50,000$ to assume we have the right distribution. Therefore we choose $N = 50,000$.

3.10 The normal approximation for our own example

We have again plotted the differences between summing N terms and summing M terms in the series representation, i.e. the residuals $X_{M,N}^{\mathrm{res}}$, against the corresponding Normal quantiles. The Q-Q plots are depicted in figure 3.8.
All the residuals are (as for the case of Gamma) strictly positive due to the series representation (3.32) only consisting of strictly positive terms. We see, that summing more terms in the series representation gives a better fit

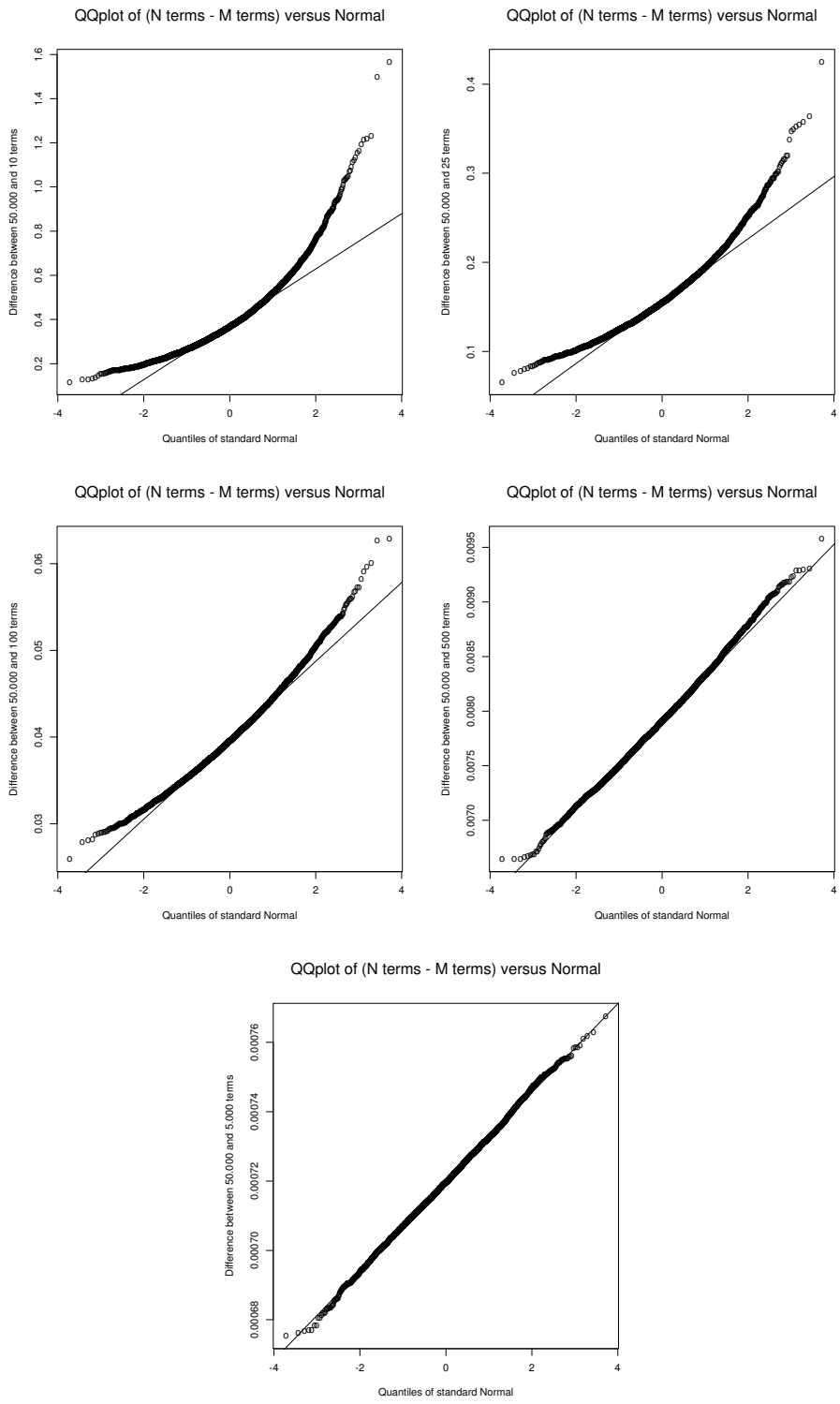

Figure 3.8: Q-Q plots of the difference between summing N terms and summing M terms in the series representation for our own example.

of the residuals to the Normal distribution—the same pattern as we saw for the case of stable with $\alpha = 0.75$. For our lowest choice of M, $M = 10$, the residuals are definitely not Normal, but the improvement when M increases is obvious. For $M = 500$ the distribution is close to the normal, and for $M = 5,000$ the distribution of the residuals is almost indistinguishable from Normal.
We notice a quite large difference to the case of stable, as in our example we needed the variance of the residuals to be of order 10^{-10} to have Normal residuals, whereas for stable with $\alpha = 0.75$ the order was only 10^{-5}. Of course, there will always be differences between distributions, and we are not able to use the level of the residuals from the case of stable directly for other distributions—but it can be used as an opening guess as we did in this example.
Again, this knowledge can be used for generating the distribution of J_1 by summing 5,000 terms in the series representation and adding a Normal random variable—the variance can be assessed by the method used in section 3.6. We have

$$\begin{aligned} N(\epsilon) &\approx \#\left\{ i : \frac{1}{\left(\frac{\Gamma_i - \mathrm{E}_{\log^2,1}(1)}{2} + 1\right)^2} > \epsilon \right\} \\ &= \#\left\{ i : \frac{\Gamma_i - \mathrm{E}_{\log^2,1}(1)}{2} + 1 \leq \frac{1}{\sqrt{\epsilon}} \right\} \\ &= \#\left\{ i : \Gamma_i \leq 2\left(\frac{1}{\sqrt{\epsilon}} - 1\right) + \mathrm{E}_{\log^2,1}(1) \right\}. \end{aligned} \tag{3.33}$$

This connection between N and ϵ can be used with (3.17) in order to calculate the variance—but as it is difficult to make any use of (3.33) in practice, we also here take a large short cut and use the empirical values from table 3.3. Thus, we are able to generate J_1 by summing 5,000 terms in the series representation and adding a Normal random variable with mean $7.2 \cdot 10^{-4}$ and variance $1.7 \cdot 10^{-10}$. We note, that the mean of the residuals is not 0 because we have not compensated our process as assumed in theorem 2.1.
For reference we state, that our program generated the 5,000 random variables (summing 50,000 terms in the series representation for each variable) in little under half an hour.

4 Stochastic differential equations driven by general Lévy processes

In this section we look at approximations to the solution of stochastic differential equations (SDE). Solutions to SDEs are often used to describe prices of financial objects of interest such as shares, foreign exchange rates and interest rates. But only in a few exceptional cases, such as the Black–Scholes setup, one is able to give an explicit solution. This fact is one of the main

reasons why the Black–Scholes method of pricing is still being heavily used in practice. In other cases one is often dependent on simulations for approximating sample paths and distributions—which is often difficult and time consuming.
Returns over long time intervals (sums of returns over finer time intervals) tend to be closer to a Gaussian distribution than ones over short horizons. But this does not indicate that other Lévy processes, besides the Brownian motion, are unfit for use in financial mathematics. The use of jump processes has actually expanded in the last couple of years and thus general Lévy processes are called for. Lévy jump processes are now considered as more realistic models for applications in finance, mainly because the jumping sample paths are closer to real-life financial data than the paths of Brownian motion—e.g. we all know that jumps are very common when the exchange opens in the morning. The class of Lévy processes is indeed very versatile when it comes to modelling tails and jumps, and the opportunity for fitting your own model is indeed present. The most popular Lévy processes in finance are the normal inverse Gaussian process (Ole E. Barndorff–Nielsen), the generalized inverse Gaussian process (Ole E. Barndorff–Nielsen), hyperbolic Lévy motion (Ernst Eberlein) and stable processes (Samorodnitsky and Taqqu [28] and John P. Nolan)—the references given are to large contributors in the development of theory and practical use of the mentioned processes, and a search on the given names will reveal a lot of articles.
In the last section we saw how to simulate Lévy processes by a series representation, and if the process fulfilled the normal approximation (theorem 2.1) we were able to simulate it much faster. We will in this section have a look at how to simulate solutions to SDEs, which are driven by a general Lévy process, when we make use of the normal approximation. By making use of the normal approximation we will save a lot of simulation time—this can then be used for simulating more sample paths or perhaps in cases that used to be too time consuming.

4.1 The (generalized) Vasiček interest rate model

It is not the purpose of this section to show every aspect of simulating solutions to SDEs that are driven by Lévy processes. This section is meant to show how the results from the last section could be used in practice, thus focusing on a simple application to illustrate the method. We will be looking at one particular example—a linear equation with additive noise

$$X_t = X_0 + \int_0^t [c_1(s)X_s + c_2(s)]\,\mathrm{d}s + \int_0^t \sigma(s)\mathrm{d}W_s, \quad t \in [0,1], \tag{4.1}$$

where W_s is a Brownian motion, and then choose $-c_1(s) = c > 0$, $c_2(s) = c\mu > 0$, $\sigma(s) = \sigma > 0$, $X_0 = x_0 > 0$ as constants. This gives us the well known Vasiček interest rate model (proposed by Vasiček [30])

$$X_t = x_0 + c\int_0^t [\mu - X_s]\,\mathrm{d}s + \sigma\int_0^t \mathrm{d}W_s, \quad t \in [0,1]. \tag{4.2}$$

X_t is called the instantaneous interest rate. This is a standard (and very easy) model for describing a fluctuating interest rate. X_t fluctuates around the value μ—whenever it deviates from μ it will be drawn back with speed according to the parameter c. σ determines the volatility. Notice, that (4.2) is just a "shifted" Ornstein–Uhlenbeck process, i.e. setting $\mu = 0$ gives us an Ornstein–Uhlenbeck process. A large drawback of the Vasiček model is that it can assume negative values. Thus, in practice the parameters will have to be chosen, so that the probability of this undesirable event is very small. We also mention that the solution to (4.2) is a Gaussian process.
If we replace the Brownian motion W_s in the Vasiček model (4.2) with a general Lévy process L_s (other than Brownian motion) we have a new linear equation

$$X_t = x_0 + c\int_0^t [\mu - X_s]\,\mathrm{d}s + \sigma\int_0^t \mathrm{d}L_s, \quad t \in [0,1], \tag{4.3}$$

with a different type of additive "noise". The "noise" is not Gaussian anymore and the Lévy process (of course) has jumps. The solution to (4.3) is known as

$$X_t = x_0\mathrm{e}^{-ct} + \mu(1-\mathrm{e}^{-ct}) + \sigma\mathrm{e}^{-ct}\int_0^t \mathrm{e}^{cs}\mathrm{d}L_s, \quad t \in [0,1]. \tag{4.4}$$

Notice, that the integrand in (4.4) is continuous and obviously has bounded variation, so if L has bounded p-variation[12] for some $p > 0$, then the integral in (4.4) exists in the Riemann–Stieltjes sense. This result is due to a theorem of Young [31]. He proved that, if f has bounded p-variation and g has bounded q-variation with $p^{-1} + q^{-1} > 1$, then the integral $\int_a^b f\mathrm{d}g$ exists in the Riemann–Stieltjes sense whenever f and g have no discontinuities at the same point. He also proved two smaller extensions to overcome the continuity problem—for more properties of this type of integration see Dudley and Norvaiša [11]. We still need to know the p-variation of the general Lévy process L, and luckily it has been shown by Lépingle [17] (sharpen your French!) that every semimartingale has bounded p-variation for $p > 2$. This is just what we needed because every Lévy process is a semimartingale—see e.g. Jacobsen [14].
Thus the integral in (4.4) can be approximated by Riemann–Stieltjes sums

$$\int_0^t f(s)\mathrm{d}L_s = \lim_{n\to\infty}\sum_{i=1}^{n} f(\xi_i^{(n)})\Delta_i^{(n)}L, \tag{4.5}$$

[12]The p-variation, $0 < p < \infty$, of a real-valued function f on $[a,b]$ is defined as

$$v_p(f) = v_p(f;[a,b]) = \sup_{\kappa}\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|^p,$$

where the supremum is taken over all subdivisions κ of $[a,b]$. If $v_p(f) < \infty$ then f is said to have bounded p-variation on $[a,b]$. Note, that 2-variation is not the same as the quadratic variation of a stochastic process, which is used in Itô calculus.

where $\Delta_i^{(n)} L = L(t_i^{(n)}) - L(t_{i-1}^{(n)})$, $\xi_i^{(n)} \in]t_{i-1}, t_i]$ and the limit is taken along any sequence of partitions τ_n of the interval $[0, t]$ with $\text{mesh}(\tau_n) \to 0$. For convenience we normally choose the partition to be $t_i^{(n)} = i/n$ for $i = 0, \ldots, n$ and $\xi_i^{(n)}$ as the interval midpoints. The Riemann–Stieltjes approach saves us a lot of energy because we are able to calculate (approximate) the solution to (4.3) by Riemann–Stieltjes sums.
This setup is indeed very exciting as we now approach a situation where we will be able to use our results from the previous section. We basically want to compare the solution we get by using the normal approximation to the real solution. This can be done ω-wise by comparing the path of the solution (4.4), based on our simulation of the Lévy process at summation level M with added normal variable, to the path based on the higher simulation level N. The meaning of 'comparing' is definitely not trivial—this notion will be discussed in section 4.3.
In our simple example we know the explicit solution and this is of course not the case in general applications. We will make use of knowing the explicit solution by approximating the real solution at the high level N by Riemann sums and approximating the solution at the lower level M with added normal variable with the Euler scheme. We will approximate the real solution as good as possible by choosing a very fine partition for the Riemann sums and then hopefully we will be able to say whether the normal approximation (and the Euler scheme) performs well. We will of course experiment with the partitions, but it is our intention to keep the partition for Riemann sums finer than for the Euler scheme in order to approximate the real solution as exact as possible.

4.2 The Euler approximation

The Euler scheme is a natural way of approximating the solution to an SDE. For better understanding we write (4.3) in the language of differentials

$$\mathrm{d}X_t = c(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}L_t, \quad t \in [0, 1], \tag{4.6}$$

with $X_0 = x_0$. The idea of the Euler scheme is to replace the differentials $\mathrm{d}t$ and $\mathrm{d}L_t$ by differences on a discrete partition τ_n. This indicates the following iterative scheme

$$\begin{aligned} X_0^{(n)} &= X_0 \\ X_{t_i}^{(n)} &= X_{t_i-1}^{(n)} + c(\mu - X_{t_i-1}^{(n)})\Delta_i + \sigma \Delta_i L, \quad i = 1, \ldots, n, \end{aligned}$$

where $\Delta_i = t_i - t_{i-1}$ and $\Delta_i L = L_{t_i} - L_{t_i-1}$. This determines the Euler approximation $X^{(n)}$ at the points t_i of the partition τ_n—notice that t_i actually depends on n which is suppressed in the notation. In practice one often chooses an equidistant partition $\tau_n : t_i = i/n$. It is common to linearly interpolate between the points t_i, which defines a continuous process. This is indeed not what we were looking for because our process can have jumps, and therefore we will keep the process constant between jumps, i.e.

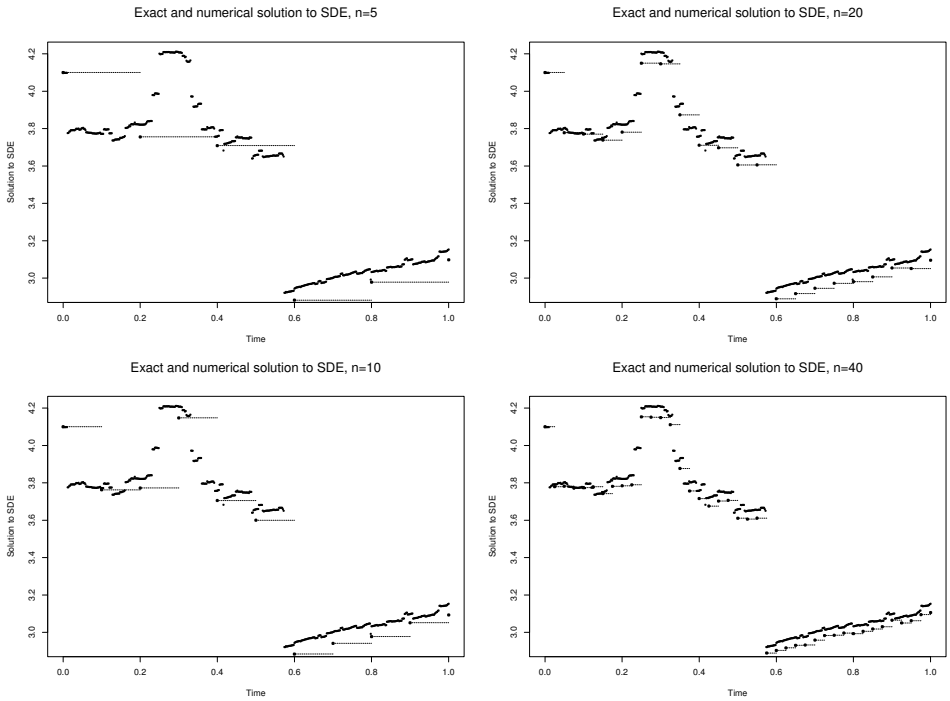

Figure 4.1: The Euler scheme at work for the SDE (4.6) with the parameters from section 4.3. The numerical solutions are showed as dashed lines.

$X_t^{(n)} = X_{t_i}^{(n)}$ for $t \in [t_i, t_{i+1}[$. As n increases the outcome is not obvious to be what we hoped for, i.e. a convergence in some sense. This problem is in the framework considered by Mikosch and Norvaiša [22], and has been generally treated by Jacod and Protter [15] in the case of a multidimensional SDE

$$X_t = x_0 + \int_0^t f(X_{s-})\mathrm{d}Y_s, \quad t \in [0,1], \tag{4.7}$$

where f denotes a matrix $f = (f^{ij})$ of functions that are all continuous differentiable and satisfy the linear growth condition $\|f^{ij}(x)\| \leq K(1 + \|x\|)$ for some constant K and where the driving process Y is a vector of semimartingales. This indeed covers the case of our linear SDE. They prove that the Euler scheme does work in the outlined situation and find the rate of convergence. Thus we are able to use the Euler scheme for approximating (4.4)—the solution to our SDE.

In figure 4.1 we have depicted the Euler scheme at work for a solution of the SDE (4.6) to appear in section 4.4. In order to see the improvement in the approximation as the step size $1/n$ decreases, we have plotted the Euler approximation of the same solution for $n \in \{5, 10, 20, 40\}$. This is not the same step sizes as will be used in section 4.4, but these small choices enable us to easily see how the scheme works in practice. We can see from the graphs, that the Euler approximation is struggling in order to keep up with

the jumping solution to the SDE. But as n increases we see a significant improvement and it is not hard to imagine, that the approximation would get even better if we choose an even larger value of n, i.e. a smaller step size $1/n$.

4.3 Simulation of solution to our SDE

We have now concluded our setup and are almost ready for simulating the solution (4.4)—but we still need to specify the Lévy process L_s. Notice, that in the light of the results from section 2, it is again only relevant to look at the Lévy jump process J_s.
In the previous section we focused a lot on stable processes and therefore it is natural to stay in this class during our further analysis. We recall the simulation results from section 3.5 and (again) choose $\alpha = 0.75$ because the series representation converged nicely in this case. By choosing a stable distribution with such a small parameter α, we are aware of the fact, that our model definitely is not suited for describing interest rates. Stable processes could perhaps be used if the parameter α was chosen close to 2 in order to avoid big jumps. However it is not the purpose of this section to describe interest rates but to show how the results on the normal approximation could be used for simulating solution to SDEs, which are driven by a non-Gaussian Lévy process. Therefore in this section we have chosen to work with the same stable process which previously proved a successful simulation. Unfortunately we are not able to simulate stable paths with an α high enough to consider modelling interest rates, so we will leave the idea of interest rates behind and just think of the process as (shifted) Ornstein–Uhlenbeck.
We also recall, that the summation levels in the series representation could be chosen as $M = 500$ and $N = 20,000$ in order to have nice results with this particular α. Therefore we will reuse these parameters for our simulations in this section.
We furthermore choose the following parameters when simulating the solution (4.4): $x_0 = 4.1$, $\mu = 4$, $c = 0.5$ and $\sigma = 1$. The parameters are chosen with the levels of interest rates in mind even though we know the driving process is not well suited for this—as explained above.
We have treated the connection between the series representation of a random variable and the series representation of a stochastic process in section 3.1. There we already specified a series representation for the generation of a stable process in (3.8), and this is the one we now will be using for generation of paths. We also need to define a partition τ_n for the approximation of Riemann–Stieltjes sums, and not surprisingly we choose $t_i^{(n)} = i/n$ for $i = 0, \ldots, n$ and $\xi_i^{(n)}$ as the interval midpoints—furthermore we choose $n = 500$ for a fine partition. For the case with Euler approximation we are only summing 500 terms in the series representation, so there we will start out with $n = 125$.
For the generation of paths we will again use the same programming lan-

guage and random number generator as in the previous section. Our intention is to generate a lot of paths and then compare the path of the solution (4.4), based on our simulation of the Lévy process at summation level M with added normal variable, to the path based on the higher simulation level N. When comparing paths it is common to assume that the two paths X_t and $X_t^{(n)}$ are 'close' whenever they are close at the interval endpoint $t = 1$. Therefore one often looks at moments of the endpoint differences, i.e. the quantity

$$e_s(\delta_n) = \mathrm{E}|X_1(\omega) - X_1^{(n)}(\omega)|^{0.25}, \tag{4.8}$$

where the s in e_s indicates that we are looking at a strong numerical solution—that is, a pathwise approximation $X^{(n)}$ which converges to X when $\delta_n = \mathrm{mesh}(\tau_n) \to 0$. The exponent 0.25 is chosen in order to make sure the second moment of $X_1 - X_1^{(n)}$ exists—we know from Samorodnitsky and Taqqu [28] that for an α-stable Y, $\mathrm{E}|Y|^p < \infty$ for all $0 < p < \alpha$ whereas $\mathrm{E}|Y|^p = \infty$ for all $p \geq \alpha$.

We know from Jacod and Protter [15] that the Euler scheme in our setup yields a strong numerical solution. We have however chosen a setup where the real solution is known, so we are able to take a step further and look at the differences over the whole path and not just only at the endpoint differences. As a measure of pathwise closeness it is common to use the difference at the point where the processes are most apart

$$\mathrm{E} \sup_{t\in[0,1]} |X_t(\omega) - X_t^{(n)}(\omega)|^{0.25}. \tag{4.9}$$

Note, that (of course) we are not able to look at every $t \in [0, 1]$ and therefore the sup in (4.9) in practice will be discretized, i.e. the same as max over i/n for $i = 0, \ldots, n$.

For completeness we should add, that a weak numerical solution aims at approximating the moments of the solution. That is, it is not important how close $X_t(\omega)$ and $X_t^{(n)}(\omega)$ really are, because one only looks at the difference of the moments—typically one would look at $|\mathrm{E}f(X_1) - \mathrm{E}f(X_1^{(n)})|$ while paying attention to the existence of the moments used. This approach should only be chosen in situations when there is no strong solution available—situations where we do not know the path. Protter and Talay [25] deal with the rate of convergence of the Euler scheme for Lévy driven SDEs in the case of weak solutions.

For the investigation of the large number of reiterations of (4.9) we use the same idea as Kloeden and Platen [16] in their section 9.3. They carry out a practical example where they use an error estimate based on the same type of differences as (4.8)—we will do the same, just for (4.9). The idea is to estimate the absolute error by the statistic

$$\hat{\epsilon} = \frac{1}{K} \sum_{i=1}^{K} \sup_{t\in[0,1]} |X_{t,i}(\omega) - X_{t,i}^{(n)}(\omega)|^{0.25}. \tag{4.10}$$

We know from the Central Limit Theorem that the error $\hat{\epsilon}$ asymptotically becomes a Gaussian random variable and converges in distribution to the (true) expectation ϵ as $K \to \infty$. We cannot generate an infinite number of paths, so if we furthermore arrange our simulations into B batches of K simulations, we can estimate the variance σ_ϵ^2 of $\hat{\epsilon}$. We denote by the subscripts i, j the ith path in the jth batch and thus for the average errors write

$$\hat{\epsilon}_j = \frac{1}{K} \sum_{i=1}^{K} \sup_{t \in [0,1]} |X_{t,i,j}(\omega) - X_{t,i,j}^{(n)}(\omega)|^{0.25} \tag{4.11}$$

for the B batches $j = 1, \ldots, B$. The B batches are then independent and approximately Gaussian for large K. We can estimate the mean of the batch averages as

$$\hat{\epsilon} = \frac{1}{B} \sum_{j=1}^{B} \hat{\epsilon}_j = \frac{1}{KB} \sum_{j=1}^{B} \sum_{i=1}^{K} \sup_{t \in [0,1]} |X_{t,i,j}(\omega) - X_{t,i,j}^{(n)}(\omega)|^{0.25}, \tag{4.12}$$

and then use the common formula

$$\hat{\sigma}_\epsilon^2 = \frac{1}{B-1} \sum_{j=1}^{B} (\hat{\epsilon}_j - \hat{\epsilon})^2 \tag{4.13}$$

to estimate the variance $\hat{\sigma}_\epsilon^2$ of the batch averages. The estimate of the variance can be used for e.g. calculating confidence intervals. For $B > 3$ we know that the random variable

$$T_B = \frac{\hat{\epsilon} - \epsilon}{\sqrt{\frac{\hat{\sigma}_\epsilon^2}{B}}} \tag{4.14}$$

is approximated by the Student t-distribution with $B - 1$ degrees of freedom, mean zero and variance $(B - 1)/(B - 3)$. Some calculations then show that a confidence interval for ϵ has the form $(\hat{\epsilon} - \Delta\hat{\epsilon}, \hat{\epsilon} + \Delta\hat{\epsilon})$ where $\Delta\hat{\epsilon} = t_{B-1,1-a/2}\sqrt{\hat{\sigma}_\epsilon^2/B}$. $t_{B-1,1-a/2}$ is the $(1 - a/2)$th quantile of a Student t-distribution (two sided) with $B - 1$ degrees of freedom and a is the significance level. That is, the value of $t_{B-1,1-a/2}$ can be found by lookup in a quantile table.

4.4 The normal approximation when simulating solutions to SDEs

According to Kloeden and Platen [16] the batch averages can be interpreted as Gaussian for batch sizes $K \geq 15$. After doing some test runs we quickly see that this value of K is too small—we will start out by using a batch size of $K = 100$ and simulate $B = 100$ batches. For comparison we do two similar simulations, where the first will be done with $B = 150$ batches with a batch size of $K = 1,000$ and the last one with $B = 200$ batches with a batch

	$\hat{\epsilon}$	$\hat{\sigma}_\epsilon^2$
$B = 100,\ K = 100,\ n = 200$	$4.1 \cdot 10^{-1}$	$5.4 \cdot 10^{-4}$
$B = 100,\ K = 100,\ n = 500$	$3.5 \cdot 10^{-1}$	$2.9 \cdot 10^{-4}$
$B = 100,\ K = 100,\ n = 1,000$	$3.3 \cdot 10^{-1}$	$1.9 \cdot 10^{-4}$
$B = 100,\ K = 100,\ n = 2,000$	$3.1 \cdot 10^{-1}$	$1.2 \cdot 10^{-4}$
$B = 150,\ K = 1,000,\ n = 200$	$4.1 \cdot 10^{-1}$	$6.7 \cdot 10^{-5}$
$B = 150,\ K = 1,000,\ n = 500$	$3.5 \cdot 10^{-1}$	$3.6 \cdot 10^{-5}$
$B = 150,\ K = 1,000,\ n = 1,000$	$3.3 \cdot 10^{-1}$	$2.2 \cdot 10^{-5}$
$B = 150,\ K = 1,000,\ n = 2,000$	$3.1 \cdot 10^{-1}$	$1.5 \cdot 10^{-5}$
$B = 200,\ K = 1,500,\ n = 200$	$4.1 \cdot 10^{-1}$	$5.2 \cdot 10^{-5}$
$B = 200,\ K = 1,500,\ n = 500$	$3.5 \cdot 10^{-1}$	$2.9 \cdot 10^{-5}$
$B = 200,\ K = 1,500,\ n = 1,000$	$3.3 \cdot 10^{-1}$	$1.7 \cdot 10^{-5}$
$B = 200,\ K = 1,500,\ n = 2,000$	$3.1 \cdot 10^{-1}$	$1.1 \cdot 10^{-5}$

Table 4.1: Values of the estimators $\hat{\epsilon}$ and $\hat{\sigma}_\epsilon^2$ for different choices of B, K and n.

size of $K = 1,500$. Hopefully we will be able to see the effect of increasing the total number of paths. We recall, that we decided to use a partition $n = 500$ for the Riemann sums and $n = 125$ for the Euler approximation, $n = (500, 125)$. We would also like to see the effect of decreasing the step size $1/n$, and after a few test runs (to determine the simulation time) we settle on using both our chosen partition of $n = (500, 125)$, a coarser partition $n = (200, 50)$ and some finer partitions $n = (1,000; 250)$, $n = (2,000; 500)$ for the same paths in the two simulations. In this way we are definitely able to see the effect of decreasing the step size, because we use the same simulation paths for all choices of partition.
We notice from (4.14) and the formula for $\Delta\hat{\epsilon}$ later in section 4.3, that we need to increase the number of batches fourfold (that is, increase the number of simulations fourfold) in order to halve the length of the confidence interval. We will be able to roughly test this with our three simulation runs. After running the three simulations we decide to use simulation data for the two smaller simulations ($B = 100$, $K = 100$ and $B = 150$, $K = 1,000$) by using intermediate values from the largest simulation ($B = 200$, $K = 1,500$). In this way we are capable of comparing the graphs as we diminish the deviations between different simulations. We have depicted the Q-Q plots of the absolute error of ϵ in figure 4.2—we have only included the graphs for the finest partition $n = (2,000; 500)$ for each of the three choices of B and K, as the plots for the other partitions shows similar shapes with larger variance for increasing step size. The first thing we notice from the graphs is that the estimator of the absolute error $\hat{\epsilon}$ is not quite normal, as we see deviations in the tails. The distribution of the absolute error $\hat{\epsilon}$ has a lighter tail at the lower end of the distribution and a heavier tail at the upper end, but the shape improves when we increase the number of simulations. The values of

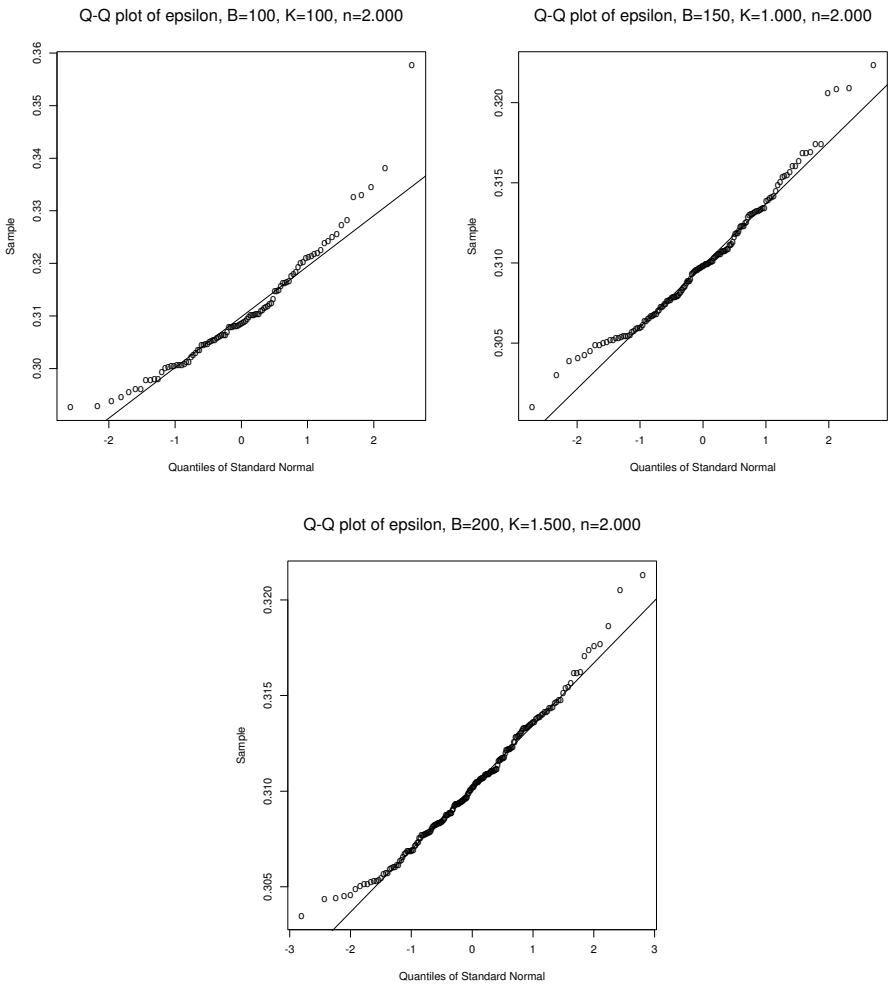

Figure 4.2: Q-Q plots of the absolute error $\hat{\epsilon}_j$.

the estimators $\hat{\epsilon}$ and $\hat{\sigma}_\epsilon^2$ (also for the other choices of partition τ) are shown i table 4.1. The values show, that there is a gain in the approximation when choosing a finer partition—the estimator of the absolute error $\hat{\epsilon}$ decreases when choosing a finer partition. Furthermore the estimator of the variance $\hat{\sigma}_\epsilon^2$ also decreases when choosing a finer partition, making our error estimate more reliable. If we look across the three different values of the total number of paths (10,000, 150,000 and 300,000) we also notice, that the estimators of the absolute error $\hat{\epsilon}$ are equal in the three cases. This tells us, that the mean value ϵ is well determined by simulating to the lower level. The precision can be seen from the variance estimates. The estimators of the variance $\hat{\sigma}_\epsilon^2$ also decrease when increasing the total number of paths—thus, we are also getting a better estimate of the mean by simulating more paths. We would also like to mention, that the graphs still remind us of the heavy tails of the

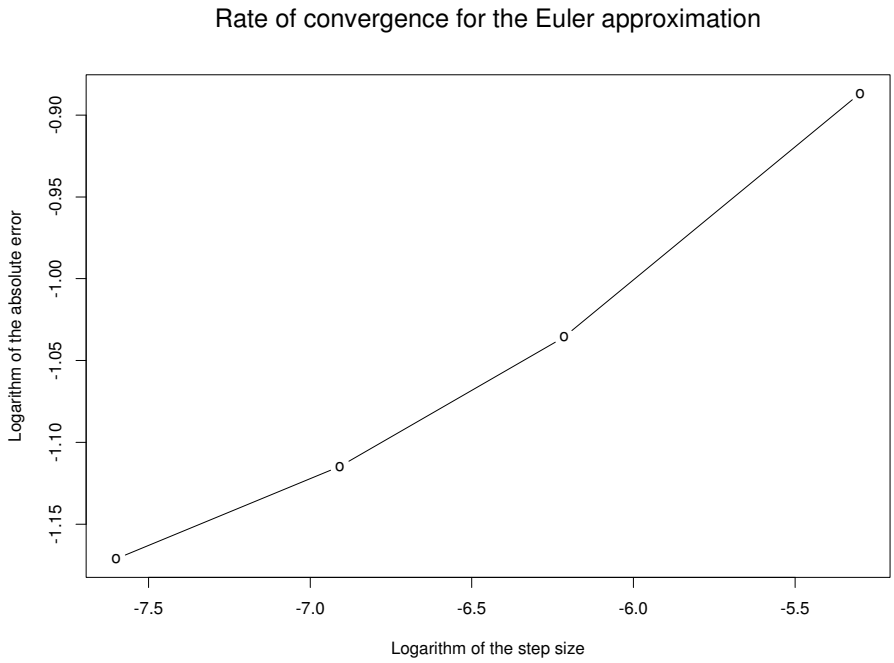

Figure 4.3: Rate of convergence for the Euler approximation in the case where $B = 200$, $K = 1,500$.

stable distribution. It is evident that the approximation does face a hard challenge in keeping up with the very big jumps, but this can be minimized by choosing a sufficiently fine partition.
The simulation time is linear in both the number of paths and the step size of the Euler approximation—simulating the double number of paths will take twice as long and halving the step size will also take twice as long. We are interested in what is the more profitable solution for reducing the variance—simulating more paths or decreasing the step size. There is no unique answer as this matter is quite complicated, and if we furthermore decrease the step size we also have the 'side effect' of decreasing the error estimate. But we will have a closer look at the two different approaches in order to get more familiar with them. In figure 4.3 we have plotted our four choices of step size against the absolute error in the case where $B = 200$, $K = 1,500$. Notice, that we have taken the logarithm to both the step size and the absolute error before plotting. This is done in order to see whether the absolute error decays like a power function (i.e. like $y = 1/x^p$, $p > 0$) when n increases. If this is true we denote the power function as the rate of convergence. We know from Jacod and Protter [15] that (under mild conditions) this rate exists and is $1/\sqrt{n}$ when we are looking at the traditional strong quantity (4.8)—but is that also the case for our sup-strong quantity (4.9)? As we can see from figure 4.3 the rate is not constant and this indicates that there is no convergence rate in the usual (power function) sense. The slope between $1/n = 1/200$ and $1/n = 1/500$ is approximately 0.16 and decreases down to

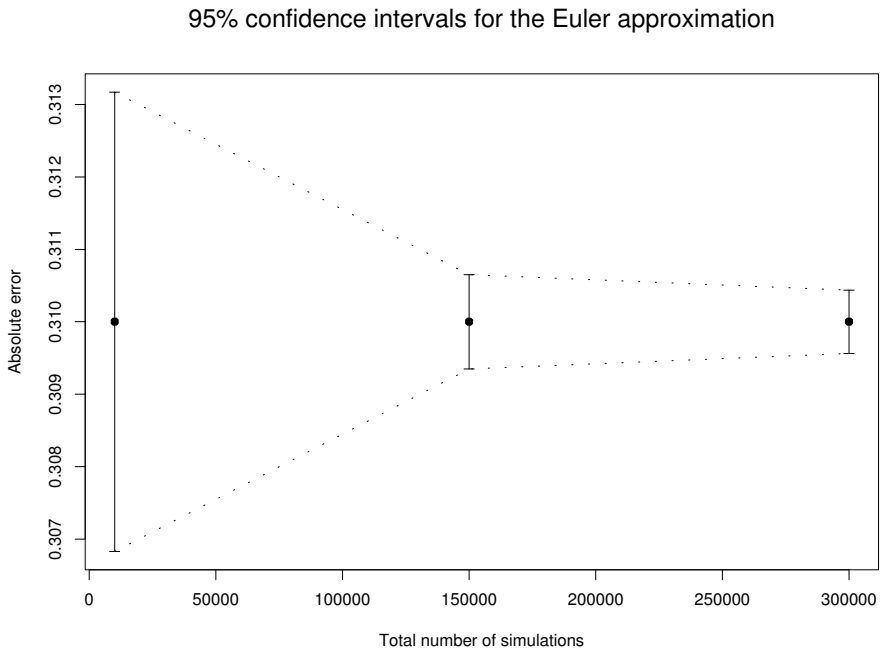

Figure 4.4: 95% confidence intervals for the Euler approximation as B increases and $K = 100$ is kept fixed.

approximately 0.08 between $1/n = 1/1,000$ and $1/n = 1/2,000$. Thus we can see that the convergence is slower than a power function. This confirms that there is quite a difference between the strong quantity (4.8) and our sup-strong quantity (4.9) with respect to convergence.
Previously in this section we mentioned, that in order to halve the length of the confidence interval for the absolute error, we need to increase the number of simulations fourfold. In figure 4.4 we have pictured the calculated 95% confidence intervals for our largest simulation when we keep $K = 100$ fixed. Thus we are again dividing our largest simulation into new batches—this time in order to see how the increasing number of simulations effect the confidence intervals when the batch size is kept fixed. The figure confirms our expectations as the confidence interval with 150,000 paths is approximately 1/4 of the one with 10,000 paths and the confidence interval with 300,000 paths is approximately $1/\sqrt{2}$ of the one with 150,000 paths.
The simulations gives us an idea of what is needed in order to simulate solutions to SDEs driven by general Lévy processes. The series representation of the Lévy process and the normal approximation is what makes these simulations possible. If we were not able to use the normal approximation, we could not simulate solutions with the kind of precision. We have simulated a very high number of processes, and by looking at our simulation results, we have not decreased the step size to the same extent. By looking at the variances in table 4.1 we notice, that in order to get a better approximation we should definitely use our given simulation time in order to decrease the

step size further and thus not simulate that many processes. Furthermore it would of course always be a good idea to use a better approximation than the Euler scheme, such as the Milstein scheme. The Milstein approximation adds an additional correction term to the Euler approximation and usually leads to a substantial improvement of the numerical approximation with Brownian Motion as driving process. Nothing similar is known in the case of an SDE driven by Lévy jump processes.
We state, that our largest simulation with 300,000 paths and step size $1/2,000$ took 5 days to run on our 850Mhz P-III.

5 Conclusion

The theory of Lévy processes has been thoroughly examined in order to present this thesis. We rely extensively on the Lévy–Khintchine representation and the Lévy–Itô decomposition—these are the backbones for splitting a Lévy process into a drift, a Brownian term and two jumping terms as we recall from (2.1). Starting from these building blocks, we have focused on simulating the jump part, as the other two parts are well treated by standard simulation techniques. A particular property for a wide class of these jumping motions has been presented by Asmussen and Rosiński [4]—the approximation of small jumps of Lévy processes by use of a Normal random variable. We have examined how this property behaves in simulations by series representations, and have on the way considered three different Lévy jumping motions.

For the symmetric α-stable Lévy motion we found, that we were able to simulate the case of $\alpha = 0.75$ 40 times faster by making use of the normal approximation. Instead of summing 20,000 terms in the series representation we could be content with just 500 terms and adding a normal random variable, because we showed, that the residual between the two was indistinguishable from normal. In fact, one might accuse us for being a little conservative by choosing the level 500, as the level of 100 was already very close to normal residuals. It could indeed be argued, that it was close enough to be used without hesitation. This would give an additional factor of one fifth to the simulation time, and thus improving the previous factor of 40 times faster to 200 times faster. This is indeed noticeable.

We also tried to simulate a symmetric α-stable Lévy motion with $\alpha = 1.5$, but did not come to near the same success as in the case of $\alpha = 0.75$. The series representation showed a drastic change in convergence to the worse, and we were not able to simulate to a lower level in order to use the normal approximation. As an experiment we argued, that we should have the same small theoretical variance of residuals in this case, as for $\alpha = 0.75$. This meant, that we needed a lower level of $4.0 \cdot 10^{15}$ in order to use the normal approximation for $\alpha = 1.5$—this is of course not even near something we are able to simulate. Even though we only examined symmetric stable processes, our results can without difficulties be applied for any skewed and translated stable.

We also looked at the Gamma process, as the process did not fulfill the requirements for the normal approximation. In some sense the Gamma process was very close to meeting the requirements, but our simulations showed that it definitely would be a bad idea to use the normal approximation in connection with Gamma processes. The residuals did not show normal behavior and they did not approach the normal distribution when we summed more terms in the series expansion, and thus decreased the residuals. But on the other hand, the series representation did converge fast enough in order to be used without the normal approximation. Furthermore the Gamma distribution is not heavy tailed and therefore one often does not need that

many variables in typical applications.
The last Lévy jumping motion we considered was our so-called own example, as we wanted to use the theory on a general Lévy jumping motion. We defined our own Lévy measure and used the inverse Lévy measure method to obtain a series representation—thus we were able to show convergence of the residuals towards a normal distribution. This meant, that we were able to choose simulation levels without prior knowledge of the particular Lévy motion, and thus showing how to make use of the normal approximation in a general setting.
As an application of our results we considered an SDE of the (shifted) Ornstein–Uhlenbeck type, driven by a non-Gaussian Lévy process. This process was chosen because the solution is known explicitly, and in this way we were able to compare the explicit solution to a numerical solution obtained by the Euler scheme. In the light of our nice results with stable we chose to use a stable process with $\alpha = 0.75$ as driving process. In the sections regarding the normal approximation we only looked at endpoint differences in order to determine closeness of processes, but in the SDE section we tried to go a little further and looked at the sup-strong quantity (4.9). We knew from our previous results, that the normal approximation would perform well, so we concentrated on approximating the explicit solution. We were able to approximate the solution to the SDE satisfactory, and showed how the error estimate between the explicit solution and our numerical approximation decreased when choosing a finer mesh for the approximation. In contrast to using a traditional strong quantity (4.8), which by Jacod and Protter [15] has a (power) convergence rate of $1/\sqrt{n}$, we showed that when using the sup-strong quantity (4.9) no (power) rate exists. This fact is notable because we are not aware of any theory on this issue.
This thesis leaves us with a method and an intuition on how to determine simulation levels when using the normal approximation. It shows, that the normal approximation can be a useful tool for simulating Lévy jump motions—in some situations it might be the only way to simulate a process in a given amount of time.

Acknowledgements

I would like to thank Johannes Müller for our fruitful discussions on random number generators—we learned a lot from each other! I would also like to express my gratitude towards Thomas Mikosch for your excellent supervision and for your many research ideas during the progress of the present thesis.

A Appendix

A.1 Distribution of random variable from section 3.6

We will now find the distribution of $\left[\Gamma_i^2 + \Gamma_i \Gamma'_{j-i}\right]$ by transformation—the distribution is used to calculate the variance of (3.24) in section 3.6. The transformation is done by use of theorem 1, section 18, from Andersson and Tolver [2].
Let Γ_i and Γ'_{j-i} be independent and define $\lambda_1 = 1$, $\lambda_2 = j - i$. The transformation

$$h: \ ]0,\infty[\to]0,\infty[, \quad (x_1, x_2) \ \to \ (x_1^2 + x_1 x_2, x_1)$$

is bijective and continuous, and the inverse

$$h^{-1}(y_1, y_2) = (y_2, (y_1 - y_2^2)/y_2),$$

is continuous and differentiable. We then have

$$\mathrm{D}h^{-1}(y_1, y_2) = \begin{bmatrix} 0 & 1 \\ \frac{1}{y_2} & \frac{-1}{y_2^2} - 1 \end{bmatrix}$$

and thus

$$\left|\det \mathrm{D}h^{-1}(y_1, y_2)\right| = \left| -\frac{1}{y_2} \right| = \frac{1}{y_2}.$$

We then have

$$g(y_1, y_2) = \frac{1_{]0,\infty[^2}(y_1, y_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} y_2^{\lambda_1 - 1} \left(\frac{y_1 - y_2^2}{y_2} \right)^{\lambda_2 - 1} \mathrm{e}^{-\left(\frac{y_1 - y_2^2}{y_2} + y_2\right)} \frac{1}{y_2},$$

and by integrating over y_2 we finally arrive at

$$\begin{aligned} g(y_1) &= \frac{1_{]0,\infty[}(y_1)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \int_0^\infty y_2^{\lambda_1 - 2} \left(\frac{y_1 - y_2^2}{y_2} \right)^{\lambda_2 - 1} \mathrm{e}^{-\left(\frac{y_1 - y_2^2}{y_2} + y_2\right)} \mathrm{d}y_2 \\ &= \frac{\mathrm{e}^{-y_1}}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \int_0^\infty y_2^{\lambda_1 - \lambda_2 - 1} \left(y_1 - y_2^2\right)^{\lambda_2 - 1} \mathrm{e}^{y_2} \mathrm{d}y_2 1_{]0,\infty[}(y_1). \end{aligned}$$

References

[1] Adams, Robert Alexander (1995) *Calculus: A Complete Course* 3rd edition, Addison–Wesley publishers.

[2] Andersson, Steen Arne; Jensen, Søren Tolver (1996) *Sandsynlighedsregning* 6th edition, Lecture Notes from Department of Statistics, University of Copenhagen.

[3] Asmussen, Søren (1999) *Stochastic Simulation with a View Towards Stochastic Processes* Lecture Notes from MaPhySto, Aarhus University.

[4] Asmussen, Søren; Rosiński, Jan (2001) *Approximations of small jumps of Lévy processes with a view towards simulation* Research Report 2000-2, MaPhySto, Aarhus University.

[5] Barndorff–Nielsen, Ole E. (1998) *Processes of normal inverse Gaussian type* Finance Stoch. **2**, 41–68.

[6] Bertoin, Jean (1996) *Lévy Processes* Cambridge University Press, Cambridge.

[7] Billingsley, Patrick (1968) *Convergence of Probability Measures* Wiley, New York.

[8] Bondesson, L. (1982) *On simulation from infinitely divisible distributions* Adv. in Appl. Probab. **14**, 855–869.

[9] Chambers, J. M.; Mallows, C. L.; Stuck, B. W. (1976). *A method for simulating stable random variables* J. Amer. Statist. Assoc. **71**, 340–344.

[10] Clausen, Trine L. (2001) *Visualization of Lévy Jump Processes* Master Thesis at Laboratory of Actuarial Mathematics, University of Copenhagen.

[11] Dudley, R. M.; Norvaiša, R. (1999) *Differentiability of Six Operators on Nonsmooth Functions and p-Variation* Lecture Notes in Mathematics 1703, Springer–Verlag, Berlin.

[12] Ferguson, T. S.; Klass, M. J. (1972) *A representation of independent increment processes without Gaussian components* Ann. Math. Statist. **43**, 1634–1643.

[13] Ickstadt, Katja; Wolpert, Robert L. (1999) *Simulation of Lévy random fields* In Dey, D. D.; Muller, M.; Sinha, D. (Eds.) *Practical Nonparametric and Semiparametric Bayesian Statistics* Springer–Verlag, New York, pp. 227–242.

[14] Jacobsen, Martin (1999) *Lévy Processes and Martingales* part of the lecture notes for the course Stochastic Integrals, Dept. of Stat., University of Copenhagen.

[15] Jacod, Jean; Protter, Philip (1998) *Asymptotic error distributions for the Euler method for stochastic differential equations* Ann. Probab. **26**, 267–307.

[16] Kloeden, Peter E.; Platen, Eckhard (1992) *Numerical Solution of Stochastic Differential Equations* Springer–Verlag, Berlin.

[17] Lépingle, D. (1976) *La variation d'ordre p des semi-martingales* Z. Wahrscheinlichkeitstheorie Verw. Geb. **36**, 295–316.

[18] Flannery, Brian P., Press, William H., Teukolsky, Saul A., Vetterling, William T. (Eds.) (2002) *Numerical Recipes in C++: The Art of Scientific Computing* Cambridge University Press, Cambridge.

[19] Khintchine, A. Ya. (1937) *Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze* Mat. Sb. **44**, 79–119.

[20] Kolmogoroff, Andrei Nikolaevich (1950) *Foundations of the Theory of Probability* Chelsea Pub., New York.

[21] McCullough, B. D. (1998) *Assessing the reliability of statistical software: part II* Amer. Statist. **53** vol. 2, 149–159.

[22] Mikosch, Thomas; Norvaiša, Rimas (2000) *Stochastic integrals without probability* Bernoulli **6** (3), 401–434.

[23] Nolan, John P. (1998) *Parameterizations and modes of stable distributions* Statistics and Probability Letters **38**, 187–195.

[24] Nolan, John P. (1998) *Univariate Stable Distributions: Parametrizations and Software* In Adler, Robert J.; Feldman, Raisa E.; Taqqu, Murad S. (Eds.) *A Practical Guide to Heavy Tails* Birkhäuser, Boston, pp. 527–533.

[25] Protter, Philip; Talay, Denis (1997) *The Euler scheme for Lévy driven stochastic differential equations* Ann. Probab. **25**, 393–423.

[26] Rosiński, Jan (2001) *Series representations of Lévy processes from the perspective of point processes* in Barndorff–Nielsen, Ole E.; Mikosch, Thomas; Resnick, Sidney I. (Eds.) *Levy Processes—Theory and Applications* Birkhäuser, Boston.

[27] Rydberg, Tina (1997) *The normal inverse Gaussian Lévy process: simulation and approximation* Stoch. Models **13**, 887–910.

[28] Samorodnitsky, Gennady; Taqqu, Murad S. (1994) *Stable Non-Gaussian Random Processes* Chapman & Hall, London.

[29] Sato, Ken–iti (1999) *Lévy Processes and Infinitely Devisible Distributions* Cambridge University Press, Cambridge.

[30] Vasiček, O. (1977) *An equilibrium characterization of the term structure* Journal of Financial Economics **5**, 177–188.

[31] Young, L. C. (1936) *An inequality of Hölder type, connected with Stieltjes integration* Acta Math. **67**, 251–282.

[32] Zolotarev, V. M. (1986) *One-Dimensional Stable Distributions* Amer. Math. Soc. **65**, Providence, RI. (Translation of 1983 Russian original)

[33] Zolotarev, V. M. (1994) *On the representation of densities of stable laws by special functions* Theory Probab. Appl. **39**, 354–362.