Approximating stable densities with Padé approximants and asymptotic series

by

Jiaxi Liang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Statistics

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis, we are interested in using the Padé approximants and asymptotic series to approximate the density functions of the stable distributions. The paper specifically discusses the selection of the optimal degree and central point of Padé approximants as well as how to connect the Padé approximants and asymptotic series as a piecewise function. Based on such approximation, a computational algorithm is developed to estimate the maximum likelihood estimator with confidence interval of the parameters, using quasi-Newton method. Simulations are conducted to evaluate the performance of this algorithm, and comparisons are made to Nolan's integral method to show that the method introduced in the thesis is fast and reliable in approximation and estimation.

Acknowledgements

I would like to express my gratitude to my supervisor Professor Christopher Small, who devoted much time, patient guidance, and constructive advice on my master's work. I am very grateful to have such a wonderful supervisor.

I also want to thank Professor Shoja'eddin Chenouri, and Professor Jock Mackay for their time in reading this thesis, and their helpful suggestions.

Dedication

This is dedicated to my parents and Ying Yang, for their kindly support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction to stable distributions

Stable distributions are a very attractive tool in modeling non-normal probability distributions in physical, economic and financial systems. For example, financial asset returns, which are the cumulative outcome of a large number of i.i.d. random variables, can be modeled by stable distributions rather than normal distributions. In fact, stable distributions can be considered as generalizations of the normal distributions in the Generalized Central Limit Theorem. They are a rich class of heavy-tailed probability distributions, which was first characterized by Paul Lévy [17]. Recent monographs using stable models in finance and economics are Rachev and Mittnik (2000) [27], McCulloch (1996) [19], and Embrechts et al. (1997) [10]. Although widely applied in many areas, the lack of closed form for densities and distribution functions except for three special cases (normal, Cauchy, and Lévy) is the main difficulty for its users.

1.1 Basic definition and characteristic function

A key property of the stable distributions is that the shape of a stable random variable is preserved under convolutions. Basically, this property is similar to the convolutions of normal distributions, and following that, the basic definition of stable distribution can be provided:

Definition 1.1. *Suppose that X, X_1, and X_2 are non-degenerate independent identically distributed random variables. The random variable X is said to be stable (or stable in the broad sense) if for any positive constants a and b, it follows*

$$aX_1 + bX_2 \stackrel{d}{=} cX + d \tag{1.1}$$

for some positive c and some $d \in \mathbb{R}$ (the symbol $\stackrel{d}{=}$ means equality in distribution)

Specifically, X is called strictly stable (or stable in the narrow sense) if (1.1) holds with $d = 0$ for all choices of a and b. And X is symmetric stable if it is stable and symmetrically distributed around a center δ, e.g. $X - \delta \stackrel{d}{=} -X + \delta$

As mentioned before, in only three cases [25] discussed in section 1.2 does the density function of a stable distribution have a closed-form expression. Generally, a stable distribution can be described by its characteristic function. Based on (1.1), one can easily generate its characteristic function. Using this, we can give an equivalent definition of stable distribution [25]:

Definition 1.2. *A non-degenerate random variable X is stable if and only if $X \stackrel{d}{=} \gamma Z + \delta$, where Z is a random variable with characteristic function*

$$E(e^{iuZ}) = \begin{cases} exp(-|u|^{\alpha}[1 - i\beta tan\frac{\pi\alpha}{2}(sign(u))]), & \alpha \neq 1 \\ exp(-|u|[1 + i\beta\frac{2}{\pi}(sign(u))log|u|]), & \alpha = 1 \end{cases} \tag{1.2}$$

and $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma > 0$, $\delta \in \mathbb{R}$. Explicitly, for the case $\alpha = 1$, $0 \cdot log0$ is always interpreted as 0.

1.2 Parameterization and symmetric stable distributions

Definition 1.2 has four parameters in describing the characteristic function of general stable distributions: an index of stability (or exponent) α, which is of most interest in this thesis, reflecting the scaling law for the distribution when the distribution is convolved with itself. Then a skewness parameter β, a scale parameter γ, and a location parameter δ, which are restricted to the range that $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma > 0$ and $\delta \in \mathbb{R}$. These parameters completely determine a stable distribution.

However, the most confusing parts in learning stable laws are caused by many different parameterizations. This vast number of parameterizations come from the historical evolution, and for convenience in different applications. For example, numerical calculation, algebraic properties and analytic properties favor different types of parameterizations. Eleven parameterizations in total have been developed to meet all sorts of need in research [25]. Here the one used in this thesis will be introduced.

Definition 1.3. *Let $S(\alpha, \beta, \gamma, \delta; 1)$ denote the class of stable laws with parameter $(\alpha, \beta, \gamma, \delta)$. A non-degenerate random variable X is $S(\alpha, \beta, \gamma, \delta; 1)$ if*

$$X \stackrel{d}{=} \begin{cases} \gamma Z + \delta, & \alpha \neq 1 \\ \gamma Z + (\delta + \beta\frac{2}{\pi}\gamma log\gamma), & \alpha = 1 \end{cases} \tag{1.3}$$

where $Z = Z(\alpha, \beta)$ is defined by (1.2), and X has characteristic function

$$E(e^{iuX}) = \begin{cases} exp(-\gamma^\alpha |u|^\alpha [1 - i\beta(tan\frac{\pi\alpha}{2})(sign(u))] + i\delta u), & \alpha \neq 1 \\ exp(-\gamma|u|[1 + i\beta\frac{2}{\pi}(sign(u))log|u|] + i\delta u), & \alpha = 1 \end{cases} \tag{1.4}$$

Under this parameterization, we will give two special cases: normal and Cauchy distributions as examples.

Example 1.1. *normal (or Gaussian) distributions. $X \sim N(\mu, \sigma^2)$ if it has a density*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

It is straightforward for one to verify that a normal distribution is $S(2, 0, \frac{\sigma}{\sqrt{2}}, \mu; 1)$

Example 1.2. *Cauchy (or Lorentz) distributions. $X \sim Cauchy(\gamma, \delta)$ if it has a density*

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-\delta)^2}$$

Similar to normal distribution, one can prove that a Cauchy distribution is $S(1, 0, \gamma, \delta; 1)$

Notice that both normal distributions and Cauchy distributions are symmetric. Although there are no explicit form for the density functions, the characteristic functions of symmetric stable distributions have a very simple form, allowing us to do the expansions and approximations (will be discussed in section 3.1.2) on the density functions. The definition of symmetric stable distributions is given below. This thesis will present a new numerical method of using Padé approximants and asymptotic series to approximate the density functions for symmetric stable distributions and correspondingly maximum likelihood estimation.

Definition 1.4. *Let the notation $S(\alpha, \gamma, \delta)$ denote the symmetric stable distribution with parameters (α, γ, δ). A random variable X is symmetric stable distribution with parameter (α, γ, δ), i.e. X is $S(\alpha, \gamma, \delta)$, if X has characteristic function*

$$E(e^{iuX}) = exp(-\gamma^\alpha |u|^\alpha + i\delta u) \tag{1.5}$$

And when the distribution is standardized, i.e. scale $\gamma = 1$, and location $\delta = 0$, the symbol $S(\alpha)$ will be used.

Under this definition, normal distributions in example 1.1 can be expressed as $S(2, \frac{\sigma}{\sqrt{2}}, \mu)$, and Cauchy distributions in example 1.2 as $S(1, \gamma, \delta)$

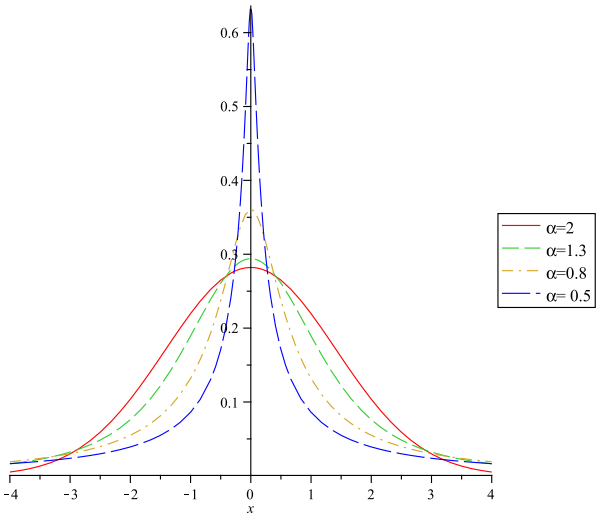

Figure 1.1: Density function for $S(\alpha, 1, 0)$ with different α

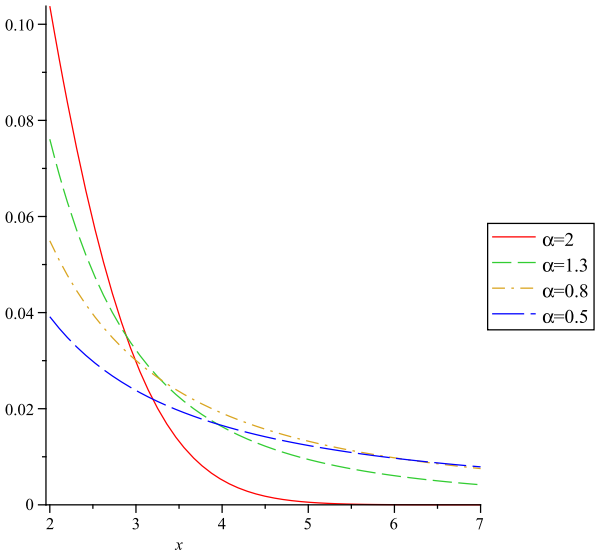

Figure 1.2: Density function for $S(\alpha, 1, 0)$ in tails

Figure 1.1 displays a plot of density functions for the standardized symmetric stable distributions with different values of parameter α (the index), and figure 1.2 shows the same but only on tails. These are calculated using the method of Nolan [23]. The graphs clearly tell that an important difference between each symmetric distribution is the heaviness in the tail, which is controlled by the parameter α. As the index α becomes smaller, the density

function tends to have heavier tails and narrower peak. Unlike the normal distribution, the significantly heavy tail in most symmetric stable laws cannot be ignored in modeling. That explains in some sense the importance of stable distributions in researching certain systems.

Chapter 2

Introduction to Padé approximants

The class of Padé approximants is a set of rational functions used to approximate given functions. While power series and asymptotic series are popular in numerical calculations and approximations, one may note that both of them share a common property that their partial sums are rational functions. Thus, it is intuitive to use a rational function as a generalization of power series in approximations. Based on Taylor's theorem and Taylor series, the corresponding theory for rational functions leads to the class of Padé approximants.

2.1 The Padé table

As a generalization of power series, Padé approximants can be developed from a very fundamental start-point: Maclaurin series. Suppose there exists a power series $\sum_{k=0}^{\infty} c_k x^k$ (here, assuming it to be a Maclaurin series), representing a given function $f(x)$, such that

$$f(x) = \sum_{k=0}^{\infty} c_k x^k \tag{2.1}$$

and two non-negative integers m and n. The Padé approximant of degree (m, n) for $f(x)$ (approximated around 0) shall mean the rational function

$$\begin{aligned} R_{[m,n]}(x) &= \sum_{j=0}^{m} a_j x^j \Big/ \sum_{i=0}^{n} b_i x^i \\ &= p(x)/q(x) \end{aligned}$$

which satisfies the condition

$$f(x) - \frac{p(x)}{q(x)} = o(x^{m+n}) \tag{2.2}$$

as $x \to 0$. Using the notation of power series in (2.1), we have

$$\sum_{k=0}^{\infty} c_k x^k - \sum_{j=0}^{m} a_j x^j / \sum_{i=0}^{n} b_i x^i = o(x^{m+n})$$
$$\sum_{j=0}^{m} a_j x^j - (\sum_{k=0}^{\infty} c_k x^k) \cdot (\sum_{i=0}^{n} b_i x^i) = o(x^{m+n})$$

Multiplying the left-side up to order $m+n$, we find that

$$\sum_{j=0}^{m+n} \left(\sum_{i=0}^{min(j,n)} b_i c_{j-i} \right) x^j = \sum_{j=0}^{m} a_j x^j + o(x^{m+n}) \tag{2.3}$$

Equating the coefficients on the left with corresponding coefficients on the right for each value of j, we find that

$$\sum_{i=0}^{min(j,n)} b_i c_{j-i} = a_j \quad \text{for } 0 \le j \le m \tag{2.4}$$

$$\sum_{i=0}^{min(j,n)} b_i c_{j-i} = 0 \quad \text{for } m < j \le m+n \tag{2.5}$$

Solving these equations directly gives the unique Padé approximants with degree (m, n). Because the degree (m, n) can be chosen as any non-negative integers, all the possible Padé approximants for $f(x)$ can be arranged in a table (so-called Padé table) as follow.

$$\begin{array}{cccc} f_{[0,0]}(x) & f_{[0,1]}(x) & f_{[0,2]}(x) & \cdots \\ f_{[1,0]}(x) & f_{[1,1]}(x) & f_{[1,2]}(x) & \cdots \\ f_{[2,0]}(x) & f_{[2,1]}(x) & f_{[2,2]}(x) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array} \tag{2.6}$$

Notice that above argument oversimplifies some of the regularity assumptions [2] which are not major concerns in this thesis. Disregarding such assumptions, we can provide a simple definition of padé approximants, assuming that the function is approximated around 0:

Definition 2.1. *Consider a function $f(x)$ and its corresponding Maclaurin series $\sum_{k=0}^{\infty} c_k x^k$. The Padé approximant for $f(x)$ of degree (m, n)($m \ge 0$, and $n \ge 0$) is given as:*

$$R_{[m,n]}(x) = \sum_{j=0}^{m} a_j x^j / \sum_{i=0}^{n} b_i x^i \tag{2.7}$$

satisfying that

$$f(x) - \sum_{j=0}^{m} a_j x^j / \sum_{i=0}^{n} b_i x^i = o(x^{m+n}) \tag{2.8}$$

as $x \to 0$, and the coefficients a_j, and b_i come from equations (2.4) and (2.5).

If Cramer's rule is used to solve (2.4) and (2.5), the polynomials for numerator and denominator can be expressed in a more convenient way:

$$p(x) = \sum_{j=0}^{m} a_j x^j = \begin{vmatrix} c_{m-n+1} & c_{m-n+2} & \cdots & c_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_m & c_{m+1} & \cdots & c_{m+n} \\ \sum_{j=n}^{m} c_{j-n} x^j & \sum_{j=n-1}^{m} c_{j-n+1} x^j & \cdots & \sum_{j=0}^{m} c_j x^j \end{vmatrix}$$

$$q(x) = \sum_{i=0}^{n} b_i x^i = \begin{vmatrix} c_{m-n+1} & c_{m-n+2} & \cdots & c_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_m & c_{m+1} & \cdots & c_{m+n} \\ x^n & x^{n-1} & \cdots & 1 \end{vmatrix}$$

where $c_n = 0$ for $n < 0$.

More generally, if one wants to get the Padé approximants centered around point x_0, they can be obtained from the power series expanded about x_0.

Definition 2.2. *Consider a function $f(x)$ and its Taylor series about x_0: $\sum_{k=0}^{\infty} d_k (x - x_0)^k$. The Padé approximant for $f(x)$ of degree (m, n)($m \geq 0$, and $n \geq 0$) about x_0 is given as:*

$$\begin{aligned} R_{[m,n]}(x) &= \sum_{j=0}^{m} a_j (x - x_0)^j / \sum_{i=0}^{n} b_i (x - x_0)^i \\ &= p(x)/q(x) \end{aligned}$$

where

$$\begin{aligned} p(x) &= \sum_{j=0}^{m} a_j (x - x_0)^j \\ &= \begin{vmatrix} d_{m-n+1} & d_{m-n+2} & \cdots & d_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ d_m & d_{m+1} & \cdots & d_{m+n} \\ \sum_{j=n}^{m} d_{j-n} (x - x_0)^j & \sum_{j=n-1}^{m} d_{j-n+1} (x - x_0)^j & \cdots & \sum_{j=0}^{m} d_j (x - x_0)^j \end{vmatrix} \end{aligned}$$

$$q(x) = \sum_{i=0}^{n} b_i (x - x_0)^i = \begin{vmatrix} d_{m-n+1} & d_{m-n+2} & \cdots & d_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ d_m & d_{m+1} & \cdots & d_{m+n} \\ (x-x_0)^n & (x-x_0)^{n-1} & \cdots & 1 \end{vmatrix}$$

2.2 Comparison with power series

From Padé table (2.6) one may notice that approximants down the left-hand column of the table are all polynomials, and functions across the diagonal from upper right to lower left have the error term that with the same order. Thus, theoretically speaking, Padé approximants will result at least no worse than Taylor polynomials by choosing proper degree (m, n). However, one may still want to investigate the exact advantages of using Padé approximants rather than Taylor polynomials.

For simplicity, consider a Maclaurin series and Padé approximants about $x = 0$. To approximate some function $f(x)$, error analysis tells us that both polynomial and rational function fit very well when x is closed to 0. For x in the tail however, the polynomial will be in trouble for those "unpolynomial-like "functions, even if the series does converge. Padé approximants, on the other hand, work appropriately for these kind of functions with flexible choice of degree (m, n)

The fact can be illustrated by the following example:

Example 2.1. *Consider the function $f(x) = e^{-x^2}$. Figure 2.1 and Figure 2.2 shows plots of $f(x)$, Taylor polynomial of order 102 and Padé approximant of degree $(12, 26)$ in different intervals.*

Figure 2.1 and 2.2 makes our point clearly. Although the Taylor polynomial does converge in tail, it may be still unable to offer an acceptable result with a very high order. In contrast, the Padé approximant fits amazingly well with the order of only $m + n = 38$. To sum up, Padé approximants bring both computational and algebraic convenience in such "unpolynomial-like"functions.

Another attractive fact to speak of is that, although Padé approximants of degree $(m+n)$ is based on a power series, the terms beyond order $m+n$ do not actually effect the Padé approximants. This fact implies that if the Padé approximants are generated from a divergent series, which "act as convergent" for the first several terms, the approximation may still be well performed.

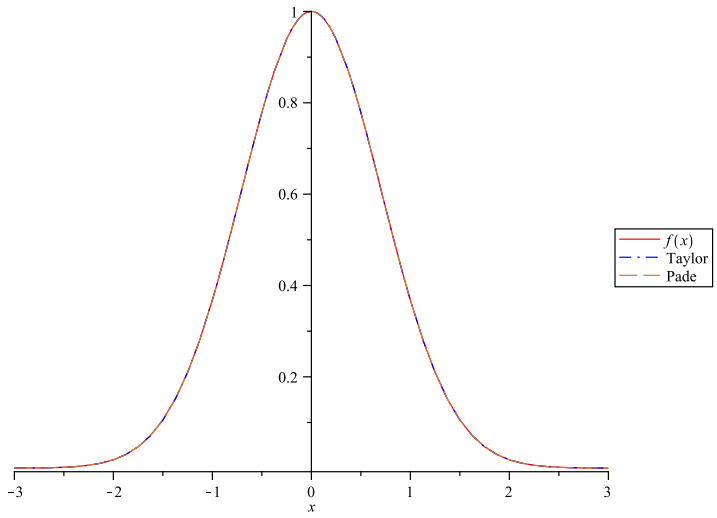

Figure 2.1: Comparison between Taylor polynomial and Padé approximant around 0

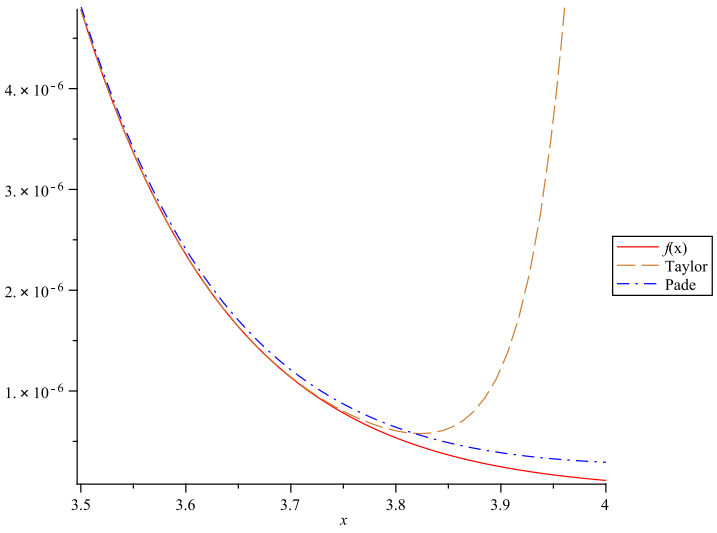

Figure 2.2: Comparison between Taylor polynomial and Padé approximant in tail

Unfortunately, Padé approximants, as rational functions, inevitably suffer from an obvious drawback in practice that singularities appear in some cases. The idea how to deal with this issue will be discussed in Chapter 3.

Chapter 3

Stable density functions

3.1 A brief review of previous results

3.1.1 Numerical integral methods

To evaluate stable density functions in practice, several numerical methods have been developed. DuMouchel [7], Holt and Crow [15] provide procedures to approximate the density functions. Both algorithms are computationally intensive and time consuming. Now the most popular algorithm to compute the stable densities is given by Nolan [24], who directly implemented the integral formulas based on Zolotariov's (M) parameterization [32]:

$$f(x,\alpha,\beta) = \frac{1}{\pi}\int_0^\infty cos(h(x,t;\alpha,\beta))exp(-t^\alpha)dt$$

where

$$h(x,t;\alpha,\beta) = \begin{cases} xt + \beta tan(\frac{\pi\alpha}{2})(t - t^\alpha), & \alpha \neq 1 \\ xt + \beta tan(\frac{\pi\alpha}{2})t + \frac{2\alpha tln(t)}{\pi}, & \alpha = 1 \end{cases}$$

The proof of this result is provided in Nolan's paper [24].

Nolan splits the region of integration into intervals according to the sign of the cosine term. The endpoints of each interval are available analytically when the stable densities are symmetric.

Nolan's formulas are remarkablly faster than the DuMouchel and Holt-Crow algorithms. However, evaluating integral form in numerical calculation is still not efficient enough where too many integral forms would slow down the approximation process. Nolan reports in his paper [23] that when $\alpha < 0.4$ computations are very slow because the exponent $exp(-t^\alpha)$ decreases slowly and the region of integration must get larger to obtain sufficient accuracy.

In this case, there are usually too many subintervals, round-off errors increase quickly and the desired accuracy cannot be always achieved. Besides, the challenging problems of this approach lie in the numerical difficulties in computing the term $tan(\frac{\pi\alpha}{2})(t-t^{\alpha})$ when α is close to 1. Moreover, when $|x|$ is large, the integrand oscillates very fast for some value of α. In this thesis, when $0.4 \leq \alpha < 2$, the true values of $f(x)$ are all evaluated in Nolan's method to compare with our approximations. For $0.4 < \alpha$, we calculate the true values of density functions by evaluating the inverse Fourier transforms directly, which is slower than Nolan's algorithm in computation.

3.1.2 Power series and asymptotic series

As discussed in section 1.2, a symmetric stable distribution has a simple characteristic function, allowing series expansion (first developed independently by Feller [12] and H.Bergström [4]). For simplicity, we consider the standardized symmetric stable distributions here. The general case can be gained by transforming $X = \gamma Z + \delta$, where Z is standardized symmetric stable distributed, with characteristic function

$$E(e^{iuZ}) = exp(-|u|^{\alpha}).$$

The inverse Fourier transform tells us that if a random variable X has characteristic function $\varphi(u)$, the probability density function $f(x)$ of X is determined uniquely as

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\varphi(u)du$$

In the case of standardized symmetric stable distributions, we can plug in the characteristic function defined in equation(1.5) (with $\gamma = 1$ and $\delta = 0$) to obtain

$$f(x) = \frac{1}{\pi}\int_{0}^{\infty} exp(-iux - u^{\alpha})du \tag{3.1}$$

Unfortunately, the idea to expand the integrand in (3.1) will fail because the resulting series cannot be integrated term-by-term. However, we can find a method to deal with this inverse Fourier transform in Small's book [30] to get the asymptotic series and power series:

Theorem 3.1. *The density function $f(x)$ of a symmetric stable distribution, whose characteristic function is $exp(-|u|^{\alpha})$, has an asymptotic series for $x \neq 0$*

$$f(x) \sim \sum_{k=1}^{\infty} \frac{\alpha}{\pi}(-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right)\frac{\Gamma(k\alpha)}{(k-1)!}\frac{1}{|x|^{k\alpha+1}} \tag{3.2}$$

i.e. let $A_n(x)$ denote the sum of the first n terms of this series, for every $n \geq 0$,

$$\lim_{x\to\infty} x^{n\alpha+1}[f(x) - A_n(x)] = 0$$

When $\alpha < 1$, this series is convergent and when $1 < \alpha < 2$, it is divergent. The case $\alpha = 1$ is the Cauchy density, for which the series converges for $|x| > 1$ and diverges elsewhere. The case $\alpha = 2$ is the normal density, for which the series degenerate to 0.

Proof. Consider the integral in equation (3.1) as a curve integral in the complex plane, and let the contour C be the positive real axis. Using a stereographic projection of the complex plane onto the Riemann sphere, where the positive real axis and the negative imaginary axis are two contours with the same start point and end point. Thus, integral the integrand along these two contours will the same. i.e.

$$f(x) = \frac{1}{\pi}\int_0^{\infty} exp(-iux - u^{\alpha})du = \frac{1}{\pi}\int_0^{-i\infty} exp(-iux - u^{\alpha})du$$

let $t = ixu$, we get

$$f(x) = \frac{-i}{\pi x}\int_0^{\infty} exp(-t)exp(-(-it/x)^{\alpha})dt$$

expand the second exponential term into power series, and note the integral of each term reduces to the gamma function

$$\begin{aligned} f(x) &= \frac{-i}{\pi x}\int_0^{\infty} exp(-t)exp(-(-it/x)^{\alpha})dt \\ &= \frac{-i}{\pi x}\int_0^{\infty} exp(-t)\left[\sum_{k=0}^{\infty}(-1)^k\frac{(-it/x)^{k\alpha}}{k!}\right]dt \\ &\sim \frac{-i}{\pi x}\left[\sum_{k=0}^{\infty}(-1)^k\left(\frac{-i}{x}\right)^{k\alpha}\frac{\Gamma(k\alpha+1)}{k!}\right] \end{aligned}$$

Note that this series must contain only real part. By Euler's identity, we have:

$$\begin{aligned} \Re\left[(-i)^{k\alpha+1}\right] &= \Re\left[(exp(\frac{-i\pi}{2}))^{k\alpha+1}\right] \\ &= cos\left(-\frac{\pi}{2}\cdot(k\alpha+1)\right) \\ &= -sin\left(\frac{k\pi\alpha}{2}\right) \end{aligned}$$

Thus, taking the real part of this series, we have:

$$\begin{aligned} f(x) &\sim \Re\left[\frac{-i}{\pi x}\sum_{k=0}^{\infty}(-1)^k\left(\frac{-i}{x}\right)^{k\alpha}\frac{\Gamma(k\alpha+1)}{k!}\right] \\ &\sim \frac{1}{\pi}\sum_{k=0}^{\infty}(-1)^{k-1}\cdot sin\left(\frac{k\pi\alpha}{2}\right)\cdot\frac{\Gamma(k\alpha+1)}{k!}\cdot\frac{1}{x^{k\alpha+1}} \\ &\sim \sum_{k=1}^{\infty}\frac{\alpha}{\pi}(-1)^{k-1}sin\left(\frac{k\pi\alpha}{2}\right)\frac{\Gamma(k\alpha)}{(k-1)!}\frac{1}{x^{k\alpha+1}} \end{aligned}$$

Note that this result requires x to be positive because we did the integral along the positive real axis. Since $f(x)$ is symmetric, we can replace x with $|x|$ to get a more general formula, i.e. equation (3.2). □

Theorem 3.2. *The density function $f(x)$ of a symmetric stable distribution, whose characteristic function is $exp(-|u|^\alpha)$, has a power series for $x \neq 0$ that*

$$f(x) = \frac{1}{\alpha\pi}\sum_{k=0}^{\infty}(-1)^k\frac{\Gamma[(2k+1)/\alpha]}{(2k)!}x^{2k} \tag{3.3}$$

When $1 \leq \alpha \leq 2$, this series is convergent and when $\alpha < 1$, it is divergent.

Proof. In the integral in equation (3.1), let $t = u^\alpha$, we have

$$f(x) = \frac{1}{\pi\alpha}\int_0^{\infty} exp(-t)exp(-ixt^{1/\alpha})t^{(1-\alpha/\alpha)}dt$$

expand the second exponential part into a power series, and reversing the order of integration and summation. We have

$$\begin{aligned} f(x) &= \frac{1}{\pi\alpha}\int_0^{\infty} exp(-t)t^{(1-\alpha/\alpha)}\left[\sum_{k=0}^{\infty}\frac{(-1)^k\left(ixt^{1/\alpha}\right)^k}{k!}\right]dt \\ &= \frac{1}{\pi\alpha}\sum_{k=0}^{\infty}\left[\frac{(-ix)^k}{k!}\int_0^{\infty} e^{-t}t^{(1-\alpha/\alpha)}\left(t^{1/\alpha}\right)^k dt\right] \\ &= \frac{1}{\pi\alpha}\sum_{k=0}^{\infty}(-i)^k\frac{\Gamma[(k+1)/\alpha]}{(k)!}x^k \end{aligned}$$

Also note that this series must contain only real part so we take the real part and get the

result:

$$\begin{aligned} f(x) &= \Re\left[\frac{1}{\pi\alpha}\sum_{k=0}^{\infty}(-i)^k\frac{\Gamma[(k+1)/\alpha]}{k!}x^k\right] \\ &= \frac{1}{\pi\alpha}\sum_{k=0}^{\infty}(-1)^k\frac{\Gamma[(2k+1)/\alpha]}{(2k)!}x^{2k} \end{aligned}$$

□

Two facts should be noted here: First, it is not really important the series (3.2) and (3.3) are convergent or not while using them for computing the density functions. Take series (3.2) as example. In practice, even if series (3.2) is divergent for $1 < \alpha < 2$, one would still sum the terms as long as the terms are decreasing in absolute value. This could lead to a quite satisfactory result when $|x|$ is large, because the error analysis tells that after summing N terms, the error term $o\left(\frac{1}{|x|^{N\alpha+1}}\right)$ is negligibly small for large $|x|$. On the other hand, when $|x|$ is small, the series works poorly even if it converges because one needs to sum up a large number of terms before getting the desirable accuracy. In contrast, the power series (3.3) has the properties mirror those of asymptotic series, with computational advantages for small $|x|$ and disadvantages for large $|x|$. Second, the asymptotic series supplies a more accurate approximation than Padé approximants in the tail, because it follows the correct power law in the tail. The above facts may suggest that, disregarding the convergence, we may use the asymptotic series to approximate the density functions in tails, and Padé approximants otherwise.

3.2 Approximating stable densities using Padé approximants

In this section, we will present in details how to approximate symmetric stable densities using Padé approximants and asymptotic series. Consider a random variable X that $X \stackrel{d}{=} S(\alpha)$ (see Definition 1.4), and its probability density function $f(x)$. Using the power series expansion of section 3.1.2 to obtain the form of Padé approximants of $f(x)$ will not be a tricky task. However, we still have some important issues. First of all, we should determine the "best" degree (m, n) of Padé approximants. Following that, it is necessary to explore the "connecting point" between two approximations, since we are going to use asymptotic series in the tails and Padé approximants around zero. Last but definitely not least, we must find out how to deal with the possible singularities.

3.2.1 Padé approximations for stable densities

Now, let's start from the form of Padé approximants of $f(x)$. Theorem 3.2 provides a power series of $f(x)$ (it is in fact the Maclaurin series of $f(x)$), and we can directly calculate the corresponding Padé approximants from definition 2.1:

Theorem 3.3. *Let $f(x)$ be the probability density function of a stable distribution $S(\alpha)$, and the Padé approximation to $f(x)$ with degree (m,n) based on the power series (3.3) is given by:*

$$R_{[m,n]}(x) = p(x)/q(x) \tag{3.4}$$

$$p(x) = \begin{vmatrix} a_{m-n+1} & a_{m-n+2} & \cdots & a_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_m & a_{m+1} & \cdots & a_{m+n} \\ \sum_{j=n}^{m} a_{j-n}x^j & \sum_{j=n-1}^{m} a_{j-n+1}x^j & \cdots & \sum_{j=0}^{m} a_j x^j \end{vmatrix} \tag{3.5}$$

$$q(x) = \begin{vmatrix} a_{m-n+1} & a_{m-n+2} & \cdots & a_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_m & a_{m+1} & \cdots & a_{m+n} \\ x^n & x^{n-1} & \cdots & 1 \end{vmatrix} \tag{3.6}$$

where

$$a_k = \begin{cases} \frac{(-1)^{\frac{k}{2}} \cdot \Gamma[(k+1)/\alpha]}{\alpha \cdot \pi \cdot (k!)}, & k \text{ is even} \\ 0, & k \text{ is odd} \end{cases}$$

Having a Padé table for $f(x)$, we want to choose the "best" one from the table, i.e. to determine the degree (m,n) of Padé approximants. As discussed in the beginning of section 2.2, the error analysis implies that the Padé approximants $R_{[m,n]}(x)$ across the diagonal of the Padé table from upper right to lower left will fit the given function $f(x)$ to the same order when $x \to 0$. However, when $|x|$ grows large, these approximants start to work differently, and one of these functions will fit $f(x)$ better than any others.

In section 3.1.2, we introduced an asymptotic series of $f(x)$ ($A_n(x)$ stands for the asymptotic series up to term n). It suggests that the order of $f(x)$ in the tails is $O(|x|^{-(\alpha+1)})$ since $x^{\alpha+1} \cdot A_n(x) \to sin(\frac{\pi\alpha}{2})\Gamma(\alpha)$, as $x \to \infty$. Therefore, the approximants with order closer to $O(|x|^{-(\alpha+1)})$ in the tails will fit $f(x)$ globally better. Let's verify this idea by following example.

Example 3.1. *Consider a stable distribution $S(\alpha)$ with index $\alpha = 1.25$, and its density function $f(x)$. From Theorem 3.1, the order of $f(x)$ in tail is $O(|x|^{-2.25})$. Let's apply three Padé approximants with different degrees to fit $f(x)$: $R_{[8,10]}(x)$, $R_{[6,12]}(x)$, and $R_{[8,12]}(x)$. Using formula (3.4)-(3.6), we can easily obtain these approximants. Figure 3.1 plots them separately in the interval $[5, 13]$.*

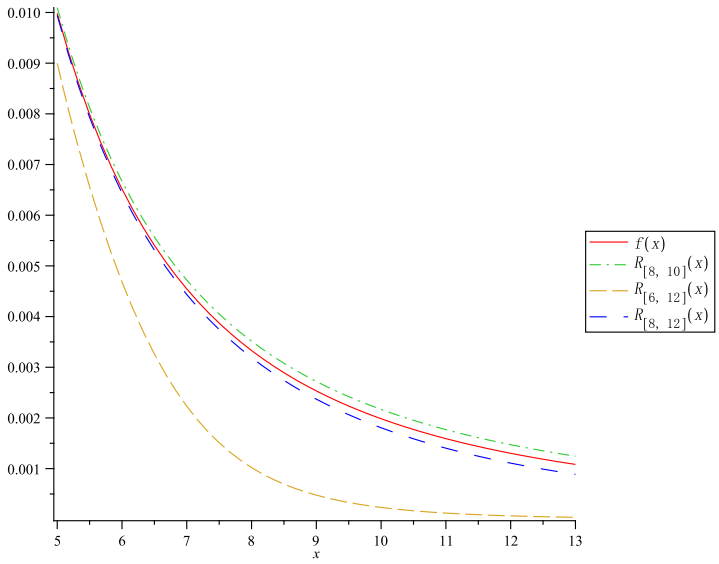

Figure 3.1: Approximate density function with different degrees

As can be seen in Figure 3.1, although the order of $R_{[8,10]}(x)$ and $R_{[6,12]}(x)$ are the same $(8 + 10 = 6 + 12 = 18)$, there is a huge difference in the tail. The order of $R_{[8,10]}(x)$ in the tail is $O(|x|^{-2})$, which is closer to $O(|x|^{-2.25})$ compared to $R_{[6,12]}(x)$, whose order is $O(|x|^{-6})$ in the tail. Since $O(|x|^{-6})$ is much higher than $O(|x|^{-2.25})$, it is not surprising to observe that $R_{[6,12]}(x)$ goes down too rapidly when $|x|$ becomes large.

Note that for symmetric stable densities, the power series contains only even power terms. Thus, the degree of Padé approximants (m, n) can be only even integers. It worth mentioning that using the formula (3.4)-(3.6), one is able to calculate the Padé approximants with odd powers, such like $R_{[7,10]}(x)$ or $R_{[8,11]}(x)$. However, we can prove that they are actually equivalent to the even powers approximants.

Lemma 3.1. *Let $R_{[m,n]}(x)$ denote the Padé approximant of the stable density function with degree (m, n), which is calculated from equations (3.4) to (3.6). For a positive integer d, it follows that*

$$R_{[2d,2d+2]}(x) = R_{[2d+1,2d+2]}(x) = R_{[2d,2d+3]}(x) = R_{[2d+1,2d+3]}(x)$$

Proof. see Appendices. □

Since $\alpha \in (0, 2]$, the order of $f(x)$ in the tail is between $O(|x|^{-1})$ and $O(|x|^{-3})$. We can conclude from all above facts that the "best" degree for approximating a symmetric stable density function $f(x)$ is $(2d, 2d + 2)$ for some positive integer d. Besides, there is another unexpected advantage for using degree $(2d, 2d + 2)$ to approximate stable densities, which will be discussed in section 3.2.2. Note that when the value of α is close to 1, the order of $f(x)$ is closer to $O(|x|^{-2})$ compared to the cases when α is close to the boundary. Thus, it is not surprising to see that the Padé approximants offer a better approximation when α is close to 1 than those when α is close to the boundary.

Another important issue is to find the "connecting point" between Padé approximants and asymptotic series. The notation x_c is used to denote such points. In standardized cases, the Padé approximants will be used in the interval $(-x_c, x_c)$ while asymptotic series in $(-\infty, -x_c) \cup (x_c, \infty)$. To determine x_c, let's consider two different situations according to the convergence of corresponding power series and asymptotic series.

First, for $1 < \alpha < 2$, as discussed in section 3.1.2, the asymptotic series will be divergent. However, for a relatively large x, if we truncate the asymptotic series at the Nth term where the terms begin to increase in their absolute values and sum up the first N terms, the approximation will be still accurate. From another aspect, if we fix N, by solving the inequality

$$\left| \frac{\Gamma((N-1)\alpha)}{(N-2)!} \cdot \frac{1}{|x|^{(N-1)\alpha+1}} \right| \geq \left| \frac{\Gamma(N\alpha)}{(N-1)!} \cdot \frac{1}{|x|^{N\alpha+1}} \right|$$

we can find a critical point $x_c = \left| \frac{\Gamma(N\alpha)}{(N-1)\cdot\Gamma((N-1)\alpha)} \right|^{\frac{1}{\alpha}}$ such that for $x > x_c$, the terms will decrease in absolute value till nth term. The choice of proper N should not be big according to two facts. First, the Padé approximants perform not as well as asymptotic series in the tail. Second, x_c is an increasing function of n. Thus, the Padé approximants will be overused when $|x|$ becomes relatively large if N is too big, and we can expect that the approximation would not be always satisfactory in this case. For example, for $\alpha = 1.8$, choosing $N = 30$ and the degree of Padé approximants as $(20, 22)$, x_c can be calculated as $x_c = 8.1$. In this case, the fractional error in the interval $(6.5, 8.1)$ will be bigger than 0.01.

In practice, choosing $10 \leq N \leq 15$ allows the asymptotic series to supply a good approximation and the Padé approximants as well. For example, to approximate the density function of $S(1.8)$, we choose $N = 12$ and calculate $x_c = 5.33$. The fractional error of the asymptotic series is plotted in Figure 3.2 to show how well this divergent series works.

At the same time, since the power series is convergent, the Padé approximants with a high enough degree will fit fairly well until x is very large. In other words, assuming the

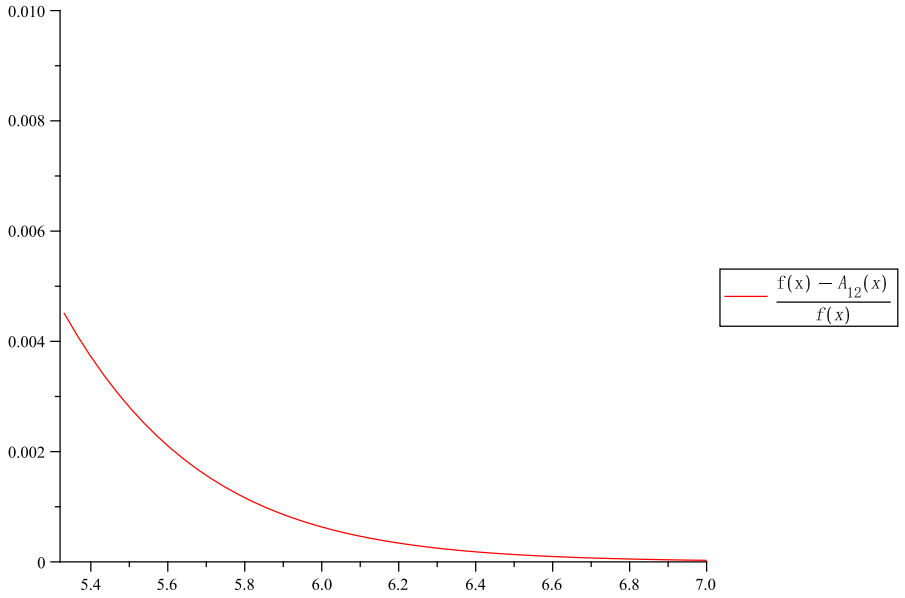

Figure 3.2: Fractional error of the asymptotic series in the interval $[5.33, 7]$

Padé approximants are suitable within the interval $(-x_p, x_p)$ in which the fractional error of the Padé approximants is less than 0.01, then for most of $\alpha \in (1, 2)$ we find in practice that $x_p > x_c$ (see Figure 3.3 and 3.4). Therefore, it is not difficult in this case to pick a "connecting point".

Notice that x_c is a increasing function of α for fixed N. Thus, for simplicity, we can choose a unified x_c which is calculated from the largest value of α, i.e. $\alpha = 2$. Then, for other $\alpha \in (1, 2)$, the asymptotic series will perform a better job. Based on the above facts, let's give the approximation formula for the density function $f(x)$ of a symmetric stable distribution $S(\alpha)$.

Lemma 3.2. *For $1 < \alpha < 2$, the density function $f(x)$ of a standardized symmetric stable distribution $S(\alpha)$ can be approximate by following formula:*

$$f(x) \approx G(x) = \begin{cases} R_{[2d,2d+2]}(x), & |x| < x_c \\ \frac{\alpha}{\pi} \sum_{k=1}^{N} (-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right) \frac{\Gamma(k\alpha)}{(k-1)!} \frac{1}{|x|^{k\alpha+1}}, & |x| \geq x_c \end{cases} \tag{3.7}$$

where d is a positive integer, and $x_c = \left| \frac{\Gamma(N\alpha_0)}{(N-1)\cdot\Gamma((N-1)\alpha_0)} \right|^{\frac{1}{\alpha_0}}$

Theoretically, we should set $\alpha_0 = 2$ to find the unified x_c. However, x_c calculated from $\alpha_0 = 2$ is too large for the Padé approximants to perform a good approximation when

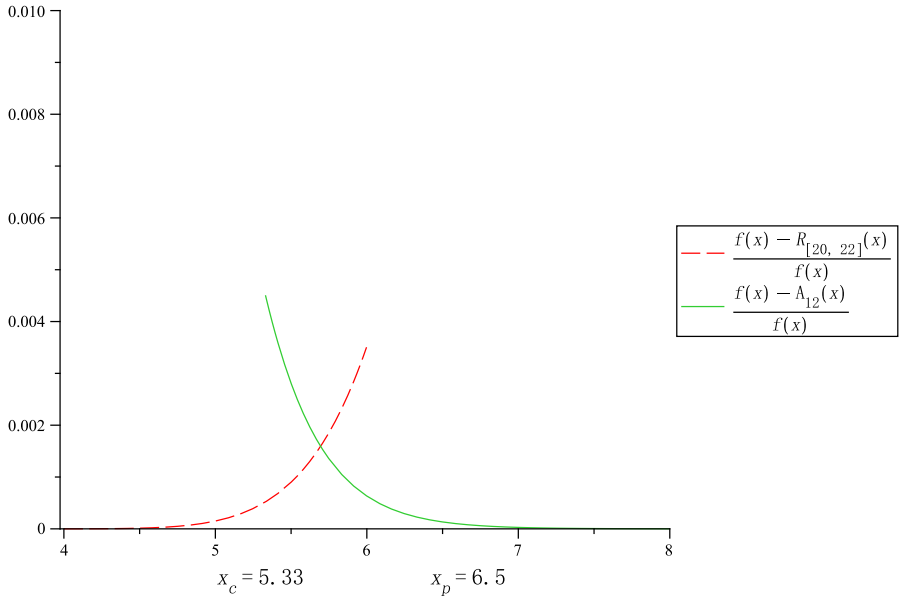

Figure 3.3: Fractional error of Padé approximants and asymptotic series when $\alpha = 1.8$

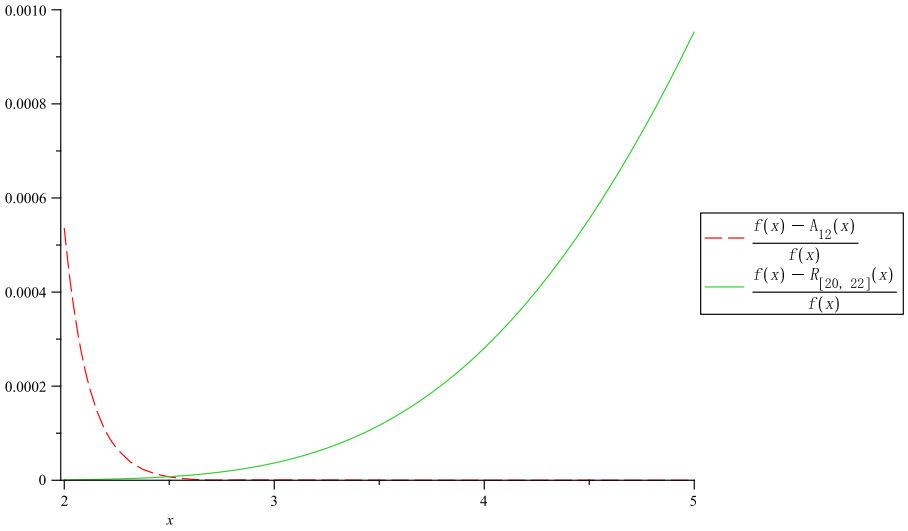

Figure 3.4: Fractional error of Padé approximants and asymptotic series when $\alpha = 1.1$

α is very close to 2, i.e. the Padé approximants will work poorly near the connecting points (the fractional error of Padé approximants will be bigger than 0.1) in these cases. As discussed before, the asymptotic series supplies a more accurate approximation than

Padé approximants in the tail, so we want x_c to be a little less for these cases. Thus, in practice, we will use a relatively large α_0 such as 1.9 instead of 2 to calculate x_c. Actually, the approximation will still be good enough (as shown in section 3.2.3).

For $0 < \alpha < 1$, we adopt a similar approach. Although the asymptotic series is convergent for $|x| > 0$, the Padé approximants are still useful, because when $|x|$ is very close to zero, the asymptotic series needs too many terms to be close to the limit, and it is challenging numerically to calculate it accurately. In this case, we have

$$f(x) = \frac{\alpha}{\pi} \sum_{k=1}^{\infty} (-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right) \frac{\Gamma(k\alpha)}{(k-1)!} \frac{1}{|x|^{k\alpha+1}}$$

$$f(x) - A_n(x) = \frac{\alpha}{\pi} \sum_{k=n+1}^{\infty} (-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right) \frac{\Gamma(k\alpha)}{(k-1)!} \frac{1}{|x|^{k\alpha+1}}$$

Note that $\lim_{n\to\infty} \frac{\alpha}{\pi} sin\left(\frac{k\pi\alpha}{2}\right) \frac{(-1)^{k-1}\Gamma(k\alpha)}{(k-1)!} \frac{1}{|x|^{k\alpha+1}} = 0$. Thus, if we fix a large N, and set a error bound err, by solving the following inequality:

$$err \geq \left| \frac{\alpha}{\pi} \cdot \frac{\Gamma((N+1)\alpha)}{N!} \cdot \frac{1}{|x|^{(N+1)\alpha+1}} \right| \geq \left| \frac{\alpha}{\pi} \cdot \frac{(-1)^N sin\left(\frac{(N+1)\pi\alpha}{2}\right) \Gamma((N+1)\alpha)}{N!} \cdot \frac{1}{|x|^{(N+1)\alpha+1}} \right|$$

we can calculate x_c for $0 < \alpha < 1$: $x_c = \left| \frac{\alpha}{\pi} \cdot \frac{\Gamma((N+1)\alpha)}{N!} \cdot \frac{1}{err} \right|^{\frac{1}{(N+1)\alpha+1}}$. Then, we can provide the approximation formula:

Lemma 3.3. *For $0 < \alpha < 1$, the symmetric stable density functions can be approximate by following formula:*

$$f(x) \approx G(x) = \begin{cases} R_{[2d,2d+2]}(x), & |x| < x_c \\ \frac{\alpha}{\pi} \sum_{k=1}^{N} (-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right) \frac{\Gamma(k\alpha)}{(k-1)!} \frac{1}{|x|^{k\alpha+1}}, & |x| \geq x_c \end{cases} \tag{3.8}$$

where d is a positive integer, and $x_c = \left| \frac{\alpha}{\pi} \cdot \frac{\Gamma((N+1)\alpha)}{N!} \cdot \frac{1}{err} \right|^{\frac{1}{(N+1)\alpha+1}}$, which is determined by N, α and err.

In this case, the choice of N should be big because the asymptotic series is convergent while the Padé approximants work well only in a narrow interval round zero. So we want to use asymptotic series as much as possible. Since x_c is a decreasing function of N in this case, theoretically, summing up more terms, i.e. choosing bigger value of N, results in better approximation by asymptotic series. However, since the term $\frac{1}{|x|^{n\alpha+1}}$ becomes smaller when n goes bigger, choosing a extremely big N will not significantly improve the approximation but slow down the computation.

For example, consider the density function of $S(0.6)$. The fractional error of the Padé approximant $R_{[20,22]}(x)$ will be bigger than 0.01 when $|x| > 0.42$, while the fractional error of the asymptotic series will be bigger than 0.5 when $|x| < 0.04$ even if we sum up more than 1000 terms. This example explains the necessity of using Padé approximants and choosing a relatively big N when $0 < \alpha < 1$.

In practice, choosing $50 \leq N \leq 60$ will be appropriate in approximation. For example, for $\alpha = 0.4$, if we set $N = 55$ and $err = 0.0001$, x_c can be calculated as $x_c = 0.00775$. The performance of asymptotic series after the connecting point is plotted in the figure below.

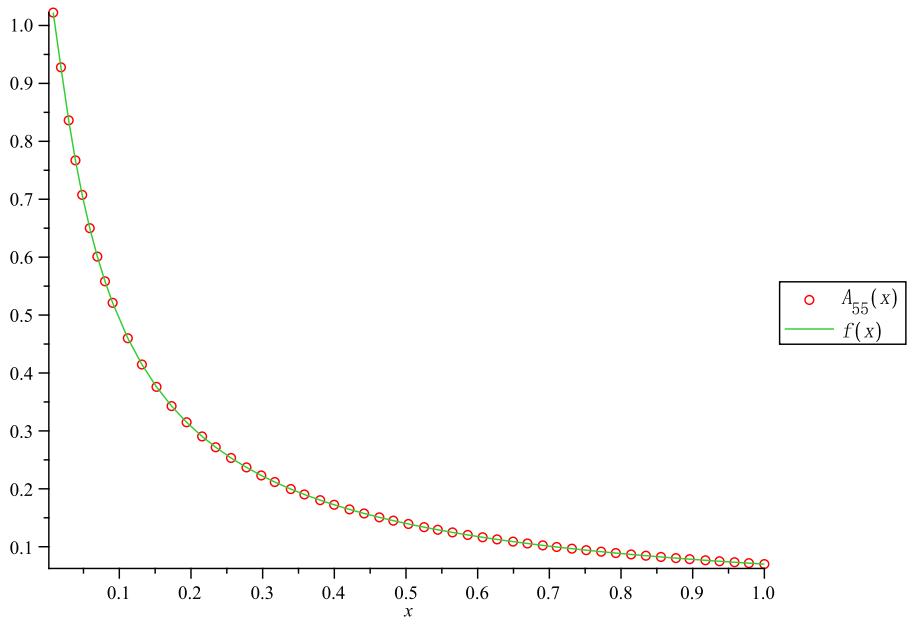

Figure 3.5: Asymptotic series in the interval $[0.00775, 1]$

For the cases when $\alpha = 1$ and $\alpha = 2$, $S(\alpha)$ becomes Cauchy and normal distributions which have closed-form density functions. Thus, they are not problems in approximation.

3.2.2 Singularities and numerical challenges

As a major drawback of Padé approximants, the possible existence of singularities is undesirable in approximating smooth functions. However, an interesting phenomenon in practice is observed: in some special cases while certain degrees are used, the singularities disappear even when the denominator of the Padé approximant has real roots. For

example, when $\alpha = 1.25$, the Padé approximants with degree $(8, 10)$ is

$$\frac{0.29646866 - 0.3647251267x^2 - 0.1109846x^4 - 0.016916375x^6 - 0.000658404x^8}{1 - 0.69675902x^2 - 0.96079136x^4 - 0.34656557x^6 - 0.05432847x^8 - 0.00329134x^{10}}$$

solve the roots of the numerator and denominator numerically and we can find that both of them have two real roots with multiplicity 1: -0.81518073 and 0.81518073. The values of the Padé approximant at these two points are $R_{[8,10]}(x)|_{x=-0.81518073} = R_{[8,10]}(x)|_{x=0.81518073} = 0.1993102$, and one can also see from the figures below that the there are no singularities in this case.

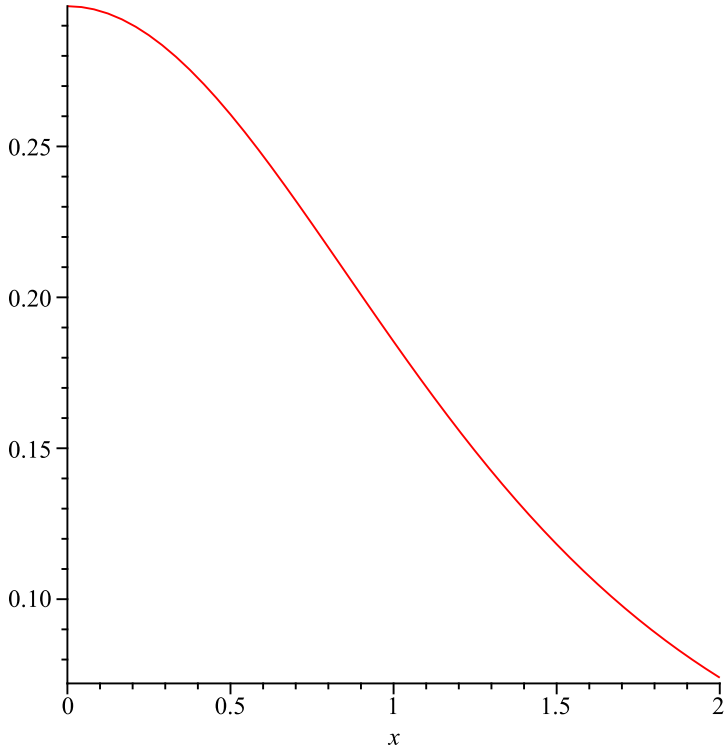

Figure 3.6: Padé approximant with degree $[8, 10]$ for $\alpha = 1.25$

Recall the "best" degree $(2d, 2d+2)$ we mentioned in last section, and we find in practice that it is one of these special cases where no singularities appear. When $d \leq 3$, it is not difficult to show that the denominator contains no real roots because the roots can be found analytically. In other cases, this conclusion of the disappearance of the singularities is based on the simulation result, and it is difficult to prove. Thus, the following conjecture is presented.

Conjecture 3.1. *Let $f(x)$ be the density function of a symmetric stable distribution. The Padé approximant for $f(x)$ with degree $(2d, 2d+2)$ has the following property: When $d > 3$,*

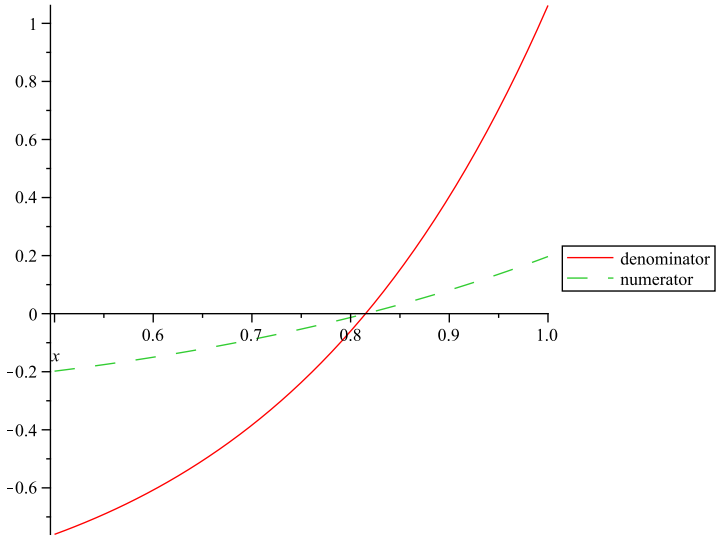

Figure 3.7: The denominator and numerator of $R_{[8,10](x)}$ for $\alpha = 1.25$

if the denominator $q(x)$ has some real root x_0 with multiplicity k, then x_0 is also the root of $p(x)$ with the same multiplicity.

In practice, the singularities do disappear using some math software (for example *Matlab*), but still happen occasionally in some others (for example *Maple*). The reason is that the coefficients of the higher power terms become smaller and smaller in absolute values. *Maple* is not as good as *Matlab* in numerical calculations, so it is difficult to compute the coefficients very accurately in *Maple* when they get too small. Thus, if the denominator of the Padé approximants has a relatively large root x_0, for example $x_0 > 4$, the singularity would not vanish. To deal with it, one can simply check the existence of large root of the denominator, and choose a different degree if the root exists.

The difficulty of accurate calculation for coefficients leads to another numerical problem. Theoretically we can improve the approximations by increasing the power of both denominator and numerator without changing the order in the tail, i.e. increasing the value of d in Lemma 3.2 and 3.3. However, this may not be true in practice. Since it is harder to compute the coefficients of the higher power terms very accurately in some software which is not proficient in numerical calculation, increasing powers would not necessarily improve the approximation given current degree is high enough. In most math softwares, choosing $6 \leq d \leq 12$ allows the Padé approximants to provide a satisfactory result, and The choice of $d > 15$ would only slow down the computation. Besides, it is better to choose a relatively big d when α is close to the boundary because the performance of Padé approximants is getting worse when α gets closer to the boundary as we discussed in section 3.2.1.

3.2.3 Numerical results

In this section, we will present the result of our approximation to symmetric stable density functions with comparison to the true value.

Let X be a random variable that follows a symmetric stable distribution with parameters (α, γ, δ), and let $f(x|\alpha, \gamma, \delta)$ denote the probability density function of X. Notice that we can transform $f(x|\alpha, \gamma, \delta)$ into $f(x|\alpha)$, the probability density function of a standardized stable distribution, by $f(x|\alpha, \gamma, \delta) = \frac{1}{\gamma} f(\frac{x-\delta}{\gamma}|\alpha)$, and then the approximation of density functions of the standardized symmetric stable distribution will automatically give a density approximation in general cases.

For $0 < \alpha < 1$, use formula (3.8), choosing $d = 10$, $N = 55$, and $err = 0.0001$:

$$f(x|\alpha) \approx G(x|\alpha) = \begin{cases} R_{[20,22]}(|x|), & |x| < x_c \\ \frac{\alpha}{\pi} \sum_{k=1}^{55} (-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right) \frac{\Gamma(k\alpha)}{(k-1)!} \frac{1}{x}, & |x| \geq x_c \end{cases}$$

$$x_c = \left| \frac{\alpha}{\pi} \cdot \frac{\Gamma(56\alpha)}{55!} \cdot \frac{1}{0.0001} \right|^{\frac{1}{56\alpha+1}}$$

$$f(x|\alpha, \gamma, \delta) \approx \frac{1}{\gamma} G(\frac{x-\delta}{\gamma}|\alpha)$$

In the general cases instead of standard, the connecting points x_{c1} and x_{c2} can be obtained by solving $|\frac{x-\delta}{\gamma}| = x_c$: $x_{c1} = -\gamma x_c + \delta$ and $x_{c2} = \gamma x_c + \delta$. In this formula, the approximation around the connecting points is the worst comparing to the tails or central point. Table 3.1 lists the result of approximation calculated from above formula according to different parameters (α, γ, δ). For each set of parameters, we choose three values of x respectively from around the central point, around the connecting points, and in the extreme tail. For each x, we provide the approximate value and fractional error (using *Maple*).

For $1 < \alpha < 2$, use formula (3.7), choosing $d = 10$, $N = 12$. To choose the connecting points, setting $\alpha_0 = 1.9$, we have $x_c = 6$ for all $\alpha \in (1, 2)$. For the values of $f(x)$ and $G(x)$ list in the tables we only keep four significant digits.

$$f(x|\alpha) \approx G(x|\alpha) = \begin{cases} R_{[20,22]}(|x|), & |x| < 6 \\ \frac{\alpha}{\pi} \sum_{k=1}^{12} (-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right) \frac{\Gamma(k\alpha)}{(k-1)!} \frac{1}{x}, & |x| \geq 6 \end{cases}$$

$$f(x|\alpha, \gamma, \delta) \approx \frac{1}{\gamma} G(\frac{x-\delta}{\gamma}|\alpha)$$

Table 3.2 picks four different value of x for each set of parameters, and present the approximate value and fractional error. The connecting points in this case are all $(-6, 6)$.

As can be seen, for most α and x, our approximation is almost as good as evaluating the integral directly. The only unsatisfactory part is when $\alpha < 0.2$ and $\alpha > 1.95$, the fractional error will be relatively large ($> 10^{-2}$) for x very close to the connecting points. However, it takes much shorter time to calculate $G(x)$ than $f(x)$. The time required for calculating $f(x)$ with parameter $0.4 \leq \alpha < 2$ using *Maple* is about 1.37 seconds and it is even longer for $\alpha < 0.4$, while for $G(x)$ with any possible α it needs less than 0.01 seconds. Thus, we can conclude that the Padé approximants and asymptotic series provide a fast and reliable approximation to stable densities, allowing us to do the maximum likelihood estimation.

(α, γ, δ)	x_c	x	$f(x)$	$G(x)$	$\frac{f(x)-G(x)}{f(x)}$
(0.99,1,0)	± 1.1084	0	0.31965	0.31965	$< 10^{-10}$
		1.1084	0.14163477	0.14156	$5.315 \cdot 10^{-4}$
		100	0.00320064	0.00320064	$-1.875 \cdot 10^{-9}$
(0.8,1,0)	± 0.4564	0	0.3606	0.3606	$-5.545 \cdot 10^{-10}$
		0.4564	0.2502	0.2502	$-1.199 \cdot 10^{-9}$
		100	$6.997 \cdot 10^{-5}$	$6.997 \cdot 10^{-5}$	$< 10^{-10}$
(0.6,1,0)	± 0.1076	0	0.4789	0.4789	$-4.176 \cdot 10^{-10}$
		0.1076	0.4238	0.4238	$1.864 \cdot 10^{-8}$
		100	$1.386 \cdot 10^{-4}$	$1.386 \cdot 10^{-4}$	$< 10^{-10}$
(0.4,1,0)	± 0.007476	0	1.058	1.058	$< 10^{-10}$
		0.007476	1.024	1.024	$7.178 \cdot 10^{-5}$
		100	$2.298 \cdot 10^{-4}$	$2.298 \cdot 10^{-4}$	$< 10^{-10}$
(0.3,1,0)	$\pm 6.24 \cdot 10^{-4}$	0	2.948	2.948	$-3.392 \cdot 10^{-10}$
		$6.24 \cdot 10^{-4}$	2.892	2.892	$8.851 \cdot 10^{-10}$
		100	$2.606 \cdot 10^{-4}$	$2.606 \cdot 10^{-4}$	$< 10^{-10}$
(0.2,1,0)	$\pm 6.198 \cdot 10^{-6}$	0	38.197	38.197	$-2.61 \cdot 10^{-10}$
		$6.198 \cdot 10^{-6}$	37.22	37.54	$-8.456 \cdot 10^{-3}$
		100	$2.495 \cdot 10^{-4}$	$2.495 \cdot 10^{-4}$	$< 10^{-10}$
(0.1,1,0)	$\pm 3.75 \cdot 10^{-11}$	0	$1.155 \cdot 10^{6}$	$1.155 \cdot 10^{6}$	$< 10^{-10}$
		$3.75 \cdot 10^{-11}$	$6.924 \cdot 10^{5}$	$6.048 \cdot 10^{5}$	0.1264
		100	$1.639 \cdot 10^{-4}$	$1.639 \cdot 10^{-4}$	$< 10^{-10}$

Table 3.1: Approximating stable densities for $0 < \alpha < 1$

(α, γ, δ)	x	$f(x)$	$G(x)$	$\frac{f(x)-G(x)}{f(x)}$
(1.95,1,0)	0.1	0.2815	0.2815	$< 10^{-10}$
	5.5	0.0006594	0.0006645	$-7.689 \cdot 10^{-3}$
	6.5	0.0002826	0.0002814	$4.556 \cdot 10^{-3}$
	100	$6.016 \cdot 10^{-8}$	$6.016 \cdot 10^{-8}$	$< 10^{-10}$
(1.9,1,0)	0.1	0.2817	0.2817	$-3.54 \cdot 10^{-10}$
	5.5	0.001175	0.001171	$3.356 \cdot 10^{-3}$
	6.5	0.0005731	0.0005734	$-6.102 \cdot 10^{-4}$
	100	$1.444 \cdot 10^{-7}$	$1.444 \cdot 10^{-7}$	$3.462 \cdot 10^{-9}$
(1.8,1,0)	0.1	0.2823	0.2823	$< 10^{-10}$
	5.5	0.002219	0.002217	$8.998 \cdot 10^{-4}$
	6.5	0.001198	0.001198	$1.330 \cdot 10^{-4}$
	100	$4.150 \cdot 10^{-7}$	$4.150 \cdot 10^{-7}$	$-7.229 \cdot 10^{-10}$
(1.6,1,0)	0.1	0.2844	0.2844	$< 10^{-10}$
	5.5	0.004340	0.004338	$5.530 \cdot 10^{-4}$
	6.5	0.002605	0.002605	$1.069 \cdot 10^{-6}$
	100	$1.692 \cdot 10^{-6}$	$1.692 \cdot 10^{-6}$	$< 10^{-10}$
(1.4,1,0)	0.1	0.2889	0.2889	$6.923 \cdot 10^{-10}$
	5.5	0.006462	0.006463	$-1.355 \cdot 10^{-4}$
	6.5	0.004175	0.004175	$5.897 \cdot 10^{-10}$
	100	$5.088 \cdot 10^{-6}$	$5.088 \cdot 10^{-6}$	$< 10^{-10}$
(1.3,1,0)	0.1	0.2926	0.2926	$3.418 \cdot 10^{-10}$
	5.5	0.007491	0.007493	$4.133 \cdot 10^{-4}$
	6.5	0.004994	0.004994	$8.2099 \cdot 10^{-9}$
	100	$8.342 \cdot 10^{-6}$	$8.342 \cdot 10^{-6}$	$3.596 \cdot 10^{-10}$
(1.2,1,0)	0.1	0.2977	0.2977	$3.3595 \cdot 10^{-10}$
	5.5	0.008473	0.008480	$-9.184 \cdot 10^{-4}$
	6.5	0.005814	0.005814	$-1.204 \cdot 10^{-9}$
	100	$1.332 \cdot 10^{-5}$	$1.332 \cdot 10^{-5}$	$-7.506 \cdot 10^{-10}$
(1.01,1,0)	0.1	0.291572	0.291572	$-1.0289 \cdot 10^{-9}$
	5.5	0.0101113	0.0101201	0.1264
	6.5	0.0072882	0.0072882	$< 10^{-10}$
	100	$3.05308 \cdot 10^{-5}$	$3.05308 \cdot 10^{-5}$	$8.9745 \cdot 10^{-8}$

Table 3.2: Approximating stable densities for $1 < \alpha < 2$

Chapter 4

Maximum likelihood estimation

In this chapter, an overview of previous work on methods of parameter estimation is presented in section 4.1, with a preference on the quantile estimation in our approach. Section 4.2 discusses in detail how to perform maximum likelihood estimation (MLE) using Padé approximants and asymptotic series. Simulations and numerical result are included in section 4.3.

4.1 Quantile estimator and other estimators

The lack of closed-form density functions makes the parameter estimation for stable distributions not an easy task. Several methods to estimate the stable index α are developed. The earliest and simplest way is to plot the empirical distribution function of sample on a log-log scale. The slope of the linear regression for large values of x yields the estimate of α, according to the asymptotic tail behavior of stable distributions [28]. Similar methods focusing on the tail behavior was proposed independently by Hill (1975) [14] and DuMouchel (1983) [9]. However, McCulloch (1997) [20], and Fofack and Nolan (1999) [13] have pointed out that this method is not reliable in practice because it is not known when the Pareto tail behavior actually occurs.

Another approach is based on characteristic functions and moments. Press is the first to try this method. He proposed a simple estimation method in 1972, based on transformations of the characteristic function [26]. Many modifications have been suggested after that, Koutrouvelis (1980) [16] presented a regression-type method, and Nikias and Shao (1995) [22] used fractional and negative moments to estimate parameters for symmetric stable distributions.

Besides, a Bayesian approach using Monte Carlo Markov chain methods was proposed by Buckle (1995) [5]. Although there is no closed-form density for X which is stable

distributed, Buckle managed to find an available density function $f(x, y|\alpha)$ jointly with some extra random variable Y. Then the posterior density was obtained through Bayes' theorem by integrating out Y.

Quantile estimation is a simple method based on quantiles of stable distributions. It was first provided by Fama and Roll in 1971 [11] for symmetric stable distributions. McCulloch generalized and improved this method in 1986 [18], obtaining consistent estimators in terms of five population quantiles (the 5th, 25th, 50th, 75th and 95th percentiles) for general stable distributions with restriction $\alpha > 0.6$. McCulloch defined two statistics v_α and v_β:

$$v_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}}$$
$$v_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.5}}{x_{0.95} - x_{0.05}}$$

which are both functions of α and β, and independent of γ and δ. x_p denotes the $p - th$ population quantile. Parameters α and β can be calculated through invert function:

$$\alpha = \phi_1(v_\alpha, v_\beta)$$
$$\beta = \phi_2(v_\alpha, v_\beta)$$

Substituting v_α and v_β by their sample values $\tilde{v}_\alpha$ and $\tilde{v}_\beta$, and applying linear interpolation between values found in tables provided by McCulloch yields estimators $\tilde{\alpha}$ and $\tilde{\beta}$:

$$\tilde{\alpha} = \phi_1(\tilde{v}_\alpha, \tilde{v}_\beta) \tag{4.1}$$
$$\tilde{\beta} = \phi_2(\tilde{v}_\alpha, \tilde{v}_\beta) \tag{4.2}$$

Scale and location parameters γ and δ can be estimated in a similar way. The tables of the functions $\phi_3(\alpha, \beta)$ and $\phi_4(\alpha, \beta)$ defined by McCulloch are provided to calculate the quantile estimators. Searching ϕ_3 and ϕ_4 with the estimators $\tilde{\alpha}$ and $\tilde{\beta}$ in the tables gives $\tilde{\gamma}$ and $\tilde{\delta}$:

$$\tilde{\gamma} = \frac{\tilde{x}_{0.75} - \tilde{x}_{0.25}}{\phi_3(\tilde{\alpha}, \tilde{\beta})} \tag{4.3}$$
$$\tilde{\delta} = \tilde{\gamma} \cdot \phi_4(\tilde{\alpha}, \tilde{\beta}) + \tilde{x}_{0.5} \tag{4.4}$$

Quantile estimates are simple to compute, and reliable for most α when the sample size is very large. Thus, we will use these estimates as initial values of the iteration in our algorithms.

Akgiray and Lamoureux (1989) [1] and Kogon and Williams (1997) [21] give simulation-based comparison of some of the methods mentioned in this section.

4.2 Maximum likelihood estimation

The maximum likelihood estimate of the parameters of the stable distribution can be obtained by maximizing the log-likelihood function:

$$l(\alpha, \gamma, \delta) = \sum_{i=1}^{n} log(f(x_i|\alpha, \gamma, \delta))$$

where $x_1, \ldots, x_n$ is an i.i.d. sample from a symmetric stable distribution. Using Padé approximants and asymptotic series, we can evaluate the log-likelihood function without difficulties. To maximize this function, a constrained (by parameter space) quasi-Newton method is used, and the quantile estimate [18] is used as the initial point in iteration.

4.2.1 Approximate score functions

To maximize the log-likelihood function using quasi-Newton method, the first order derivative (score function) is needed. The score function is difficult to evaluate because of the lack of closed-form density functions, and in this section we will introduce how to approximate the score function by Padé approximants and asymptotic series. The score function is:

$$s(\alpha, \gamma, \delta) = \nabla l(\alpha, \gamma, \delta) = \left(\frac{\partial l}{\partial \alpha}, \frac{\partial l}{\partial \gamma}, \frac{\partial l}{\partial \delta} \right)^T$$

Note that we can compute the score functions as:

$$\begin{aligned} \frac{\partial l}{\partial \theta} &= \sum_{i=1}^{n} \frac{\partial log(f(x_i|\theta))}{\partial \theta} \\ &= \sum_{i=1}^{n} \frac{\partial f(x_i|\theta)}{\partial \theta} \cdot \frac{1}{f(x_i|\theta)} \end{aligned}$$

That means once we have the derivative of the density function, we can compute the score functions.

To find the derivative of the density function, we start with the power and asymptotic series (4.5) and (4.6):

$$f(x|\alpha, \gamma, \delta) \sim \sum_{k=1}^{\infty} b_k^* \left| \frac{x - \delta}{\gamma} \right|^{-(k\alpha+1)} \tag{4.5}$$

$$f(x|\alpha, \gamma, \delta) = \sum_{k=1}^{\infty} a_k^* \left(\frac{x - \delta}{\gamma} \right)^k \tag{4.6}$$

where

$$b_k^* = \frac{\alpha}{\pi\gamma}(-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right)\frac{\Gamma(k\alpha)}{(k-1)!} \tag{4.7}$$

$$a_k^* = \begin{cases} \frac{(-1)^{\frac{k}{2}}\cdot\Gamma[(k+1)/\alpha]}{\alpha\cdot\pi\cdot\gamma\cdot(k!)}, & k \text{ is even} \\ 0, & k \text{ is odd} \end{cases} \tag{4.8}$$

Taking the derivative of (4.5) and (4.6) provides us the power series and asymptotic series of $\partial f/\partial\gamma$:

$$\frac{\partial}{\partial\gamma}\left[b_k^*\left(\frac{x-\delta}{\gamma}\right)^{-(k\alpha+1)}\right] = \frac{\alpha}{\pi\gamma^2}\frac{(-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right)\Gamma(k\alpha)k\alpha}{(k-1)!}\left|\frac{x-\delta}{\gamma}\right|^{-(k\alpha+1)} \tag{4.9}$$

$$\frac{\partial}{\partial\gamma}\left[a_k^*\left(\frac{x-\delta}{\gamma}\right)^{k}\right] = \begin{cases} \frac{(-1)^{\frac{k}{2}+1}\cdot\Gamma[(k+1)/\alpha](k+1)\cdot\left(\frac{x-\delta}{\gamma}\right)^k}{\alpha\cdot\pi\cdot\gamma^2\cdot(k!)}, & k \text{ is even} \\ 0, & k \text{ is odd} \end{cases} \tag{4.10}$$

Now calculate the Padé approximants of $\partial f/\partial\gamma$ based on (4.10):

$$R_{[m,n]}^{\gamma}(x) = p_\gamma(x)/q_\gamma(x) \tag{4.11}$$

$$p_\gamma(x) = \begin{vmatrix} c_{m-n+1}^* & c_{m-n+2}^* & \cdots & c_{m+1}^* \\ \vdots & \vdots & \ddots & \vdots \\ c_m^* & c_{m+1}^* & \cdots & c_{m+n}^* \\ \sum_{j=n}^m c_{j-n}^*\left(\frac{x-\delta}{\gamma}\right)^j & \sum_{j=n-1}^m c_{j-n+1}^*\left(\frac{x-\delta}{\gamma}\right)^j & \cdots & \sum_{j=0}^m c_j^*\left(\frac{x-\delta}{\gamma}\right)^j \end{vmatrix} \tag{4.12}$$

$$q_\gamma(x) = \begin{vmatrix} c_{m-n+1}^* & c_{m-n+2}^* & \cdots & c_{m+1}^* \\ \vdots & \vdots & \ddots & \vdots \\ c_m^* & c_{m+1}^* & \cdots & c_{m+n}^* \\ \left(\frac{x-\delta}{\gamma}\right)^n & \left(\frac{x-\delta}{\gamma}\right)^{n-1} & \cdots & 1 \end{vmatrix} \tag{4.13}$$

where c_k^* is defined as:

$$c_k^* = \begin{cases} \frac{(-1)^{\frac{k}{2}+1}\cdot\Gamma[(k+1)/\alpha](k+1)}{\alpha\cdot\pi\cdot\gamma^2\cdot(k!)}, & k \text{ is even} \\ 0, & k \text{ is odd} \end{cases} \tag{4.14}$$

To combine the Padé approximants and asymptotic series as a piecewise function to approximate the derivative of the density function, we apply the same idea in chapter 3 to

choose the degree of the Padé approximants and the connecting points. The degree of the Padé approximants can be chosen according to the order of $\partial f/\partial\gamma$ in the tails, and the connecting points by solving the following inequalities.

For $0 < \alpha < 1$, the connecting points x_c^γ can be calculated from:

$$err \geq \left| \frac{\alpha}{\gamma^2 \pi} \cdot \frac{\Gamma((N+1)\alpha)(N+1)\alpha}{N!} \cdot \frac{1}{|\frac{x-\delta}{\gamma}|^{(N+1)\alpha+1}} \right|$$

and for $1 < \alpha < 2$ from:

$$\left| \frac{\Gamma((N-1)\alpha)(N-1)}{(N-2)!} \cdot \frac{1}{|\frac{x-\delta}{\gamma}|^{(N-1)\alpha+1}} \right| \geq \left| \frac{\Gamma(N\alpha)N}{(N-1)!} \cdot \frac{1}{|\frac{x-\delta}{\gamma}|^{N\alpha+1}} \right|$$

Lemma 4.1. *The derivative of density functions with respect to the scale parameter γ can be approximated by the following function:*

$$\frac{\partial f(x)}{\partial \gamma} \approx g_\gamma(x) = \begin{cases} R^\gamma_{[2d,2d+2]}(x), & |\frac{x-\delta}{\gamma}| < x_c^\gamma \\ \frac{\alpha}{\pi\gamma^2} \sum_{k=1}^{N} \frac{(-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right)\Gamma(k\alpha)k\alpha}{(k-1)!} |\frac{x-\delta}{\gamma}|^{-(k\alpha+1)}, & |\frac{x-\delta}{\gamma}| \geq x_c^\gamma \end{cases} \tag{4.15}$$

where d is a positive integer, and

$$x_c^\gamma = \begin{cases} \left| \frac{\alpha}{\pi\gamma^2} \cdot \frac{\Gamma((N+1)\alpha)\cdot(N+1)\cdot\alpha}{N!} \cdot \frac{1}{err} \right|^{\frac{1}{(N+1)\alpha+1}}, & 0 < \alpha < 1 \\ \left| \frac{\Gamma(N\alpha_0)}{\Gamma((N-1)\alpha_0)\cdot(N-1)} \cdot \frac{N}{N-1} \right|^{\frac{1}{\alpha_0}}, & 1 < \alpha < 2 \end{cases}$$

As discussed in 3.2.1, when $0 < \alpha < 1$, the connecting point x_c^γ is determined by (α, γ, δ), err and N, and when $1 < \alpha < 2$ by (γ, δ), α_0 and N. The choice of the value of N ,d and α_0 is similar to what we did in chapter 3 for the same reason. It is better to choose big N when $0 < \alpha < 1$, for example $50 \leq N \leq 60$, and small N when $1 < \alpha < 2$ for example $10 \leq N \leq 15$. Choosing $6 \leq d \leq 12$ and $\alpha_0 = 1.9$ will result in fast and accurate computation in practice. The same N ,d and α_0 applies in Lemma 4.2 and 4.3.

For the derivative with respect to the location parameter δ and stable index parameter α, we use the same method, starting from the derivative with respect to the location parameter of the power and asymptotic series (4.5) and (4.6):

$$\frac{\partial}{\partial\delta}\left[b_k^* \left(\frac{x-\delta}{\gamma} \right)^{-(k\alpha+1)} \right] =$$
$$\frac{\alpha}{\pi\gamma^2} \frac{(-1)^{k-1} sin\left(\frac{k\pi\alpha}{2}\right)\Gamma(k\alpha)\cdot(k\alpha+1)\cdot sign\left(\frac{x-\delta}{\gamma}\right)}{(k-1)!} \left| \frac{x-\delta}{\gamma} \right|^{-(k\alpha+2)} \tag{4.16}$$

$$\frac{\partial}{\partial\delta}\left[a_k^*\left(\frac{x-\delta}{\gamma}\right)^k\right]=\begin{cases}0, & k \text{ is even}\\ \frac{(-1)^{\frac{k+3}{2}}\cdot\Gamma[(k+2)/\alpha](k+1)\cdot\left(\frac{x-\delta}{\gamma}\right)^k}{\alpha\cdot\pi\cdot\gamma^2\cdot(k+1)!}, & k \text{ is odd}\end{cases} \tag{4.17}$$

$$d_k^*=\begin{cases}0, & k \text{ is even}\\ \frac{(-1)^{\frac{k+3}{2}}\cdot\Gamma[(k+2)/\alpha](k+1)}{\alpha\cdot\pi\cdot\gamma^2\cdot(k+1)!}, & k \text{ is odd}\end{cases} \tag{4.18}$$

Then compute the Padé approximants of $\partial f/\partial\delta$ based on (4.17):

$$R^{\delta}_{[m,n]}(x)=p_\delta(x)/q_\delta(x) \tag{4.19}$$

$$p_\delta(x)=\begin{vmatrix} d^*_{m-n+1} & d^*_{m-n+2} & \cdots & d^*_{m+1}\\ \vdots & \vdots & \ddots & \vdots\\ d^*_m & d^*_{m+1} & \cdots & d^*_{m+n}\\ \sum_{j=n}^m d^*_{j-n}\left(\frac{x-\delta}{\gamma}\right)^j & \sum_{j=n-1}^m d^*_{j-n+1}\left(\frac{x-\delta}{\gamma}\right)^j & \cdots & \sum_{j=0}^m d^*_j\left(\frac{x-\delta}{\gamma}\right)^j\end{vmatrix} \tag{4.20}$$

$$q_\delta(x)=\begin{vmatrix} d^*_{m-n+1} & d^*_{m-n+2} & \cdots & d^*_{m+1}\\ \vdots & \vdots & \ddots & \vdots\\ d^*_m & d^*_{m+1} & \cdots & d^*_{m+n}\\ \left(\frac{x-\delta}{\gamma}\right)^n & \left(\frac{x-\delta}{\gamma}\right)^{n-1} & \cdots & 1\end{vmatrix} \tag{4.21}$$

where d_k^* is defined as (4.18). Applying the same idea as before allows us to choose the optimal degree and connecting points.

Lemma 4.2. *The derivative of density functions with respect to the location parameter δ can be approximated by the following function:*

$$\frac{\partial f(x)}{\partial\delta}\approx g_\delta(x)=\begin{cases}R^{\delta}_{[2d+1,2d+4]}(x), & |\frac{x-\delta}{\gamma}|<x_c^\delta\\ \frac{\alpha}{\pi\gamma^2}\sum_{k=1}^N\frac{(-1)^{k-1}sin\left(\frac{k\pi\alpha}{2}\right)\Gamma(k\alpha)(k\alpha+1)sign(\frac{x-\delta}{\gamma})}{(k-1)!}\left|\frac{x-\delta}{\gamma}\right|^{-(k\alpha+2)}, & |\frac{x-\delta}{\gamma}|\geq x_c^\delta\end{cases} \tag{4.22}$$

where d is a positive integer, and

$$x_c^\delta=\begin{cases}\left|\frac{\alpha}{\pi\gamma^2}\cdot\frac{\Gamma((N+1)\alpha)\cdot((N+1)\alpha+1)}{N!}\cdot\frac{1}{err}\right|^{\frac{1}{(N+1)\alpha+2}}, & 0<\alpha<1\\ \left|\frac{\Gamma(N\alpha_0)}{\Gamma((N-1)\alpha_0)\cdot(N-1)}\cdot\frac{N\alpha_0+1}{(N-1)\alpha_0+1}\right|^{\frac{1}{\alpha_0}}, & 1<\alpha<2\end{cases}$$

Now for the stable index parameter α:

$$\frac{\partial}{\partial\alpha}\left[b_k^*\left(\frac{x-\delta}{\gamma}\right)^{-(k\alpha+1)}\right]=$$
$$\frac{(-1)^k\Gamma(k\alpha)}{(k-1)!}\cdot\left[\frac{sin(\frac{k\pi\alpha}{2})}{\gamma\pi}\left(1+\alpha k\Psi(k\alpha)+\alpha kln(|\frac{x-\delta}{\gamma}|)\right)+\frac{\alpha cos(\frac{k\pi\alpha}{2})k}{2\gamma}\right]\cdot\frac{1}{|\frac{x-\delta}{\gamma}|^{k\alpha+1}} \tag{4.23}$$

where Ψ is the digamma function: the logarithmic derivative of the gamma function.

$$\frac{\partial}{\partial\alpha}\left[a_k^*\left(\frac{x-\delta}{\gamma}\right)^k\right]=\begin{cases}\frac{(-1)^{\frac{k+3}{2}}\cdot\Gamma[(k+1)/\alpha][\Psi[(k+1)/\alpha](2k+1)\alpha]\cdot\left(\frac{x-\delta}{\gamma}\right)^k}{\alpha^3\cdot\pi\cdot\gamma\cdot k!}, & k \text{ is even}\\ 0, & k \text{ is odd}\end{cases} \tag{4.24}$$

$$e_k^*=\begin{cases}\frac{(-1)^{\frac{k+3}{2}}\cdot\Gamma[(k+1)/\alpha][\Psi[(k+1)/\alpha](2k+1)\alpha]}{\alpha^3\cdot\pi\cdot\gamma\cdot k!}, & k \text{ is even}\\ 0, & k \text{ is odd}\end{cases} \tag{4.25}$$

Then compute the Padé approximants of $\partial f/\partial\alpha$ based on (4.24):

$$R^\alpha_{[m,n]}(x)=p_\alpha(x)/q_\alpha(x) \tag{4.26}$$

$$p_\alpha(x)=\begin{vmatrix} e^*_{m-n+1} & e^*_{m-n+2} & \cdots & e^*_{m+1}\\ \vdots & \vdots & \ddots & \vdots\\ e^*_m & e^*_{m+1} & \cdots & e^*_{m+n}\\ \sum_{j=n}^m e^*_{j-n}\left(\frac{x-\delta}{\gamma}\right)^j & \sum_{j=n-1}^m e^*_{j-n+1}\left(\frac{x-\delta}{\gamma}\right)^j & \cdots & \sum_{j=0}^m e^*_j\left(\frac{x-\delta}{\gamma}\right)^j\end{vmatrix} \tag{4.27}$$

$$q_\alpha(x)=\begin{vmatrix} e^*_{m-n+1} & e^*_{m-n+2} & \cdots & e^*_{m+1}\\ \vdots & \vdots & \ddots & \vdots\\ e^*_m & e^*_{m+1} & \cdots & e^*_{m+n}\\ \left(\frac{x-\delta}{\gamma}\right)^n & \left(\frac{x-\delta}{\gamma}\right)^{n-1} & \cdots & 1\end{vmatrix} \tag{4.28}$$

where e_k^* is defined as (4.25).

The connecting point x_c^α in this case can be calculated numerically from the following inequalities:

for $0 < \alpha < 1$

$$err \geq \left| \frac{\Gamma(N\alpha) \cdot (1 + \alpha N\Psi(N\alpha) - \alpha N ln(|\frac{x-\delta}{\gamma}|))}{\pi |\frac{x-\delta}{\gamma}|^{N\alpha+1}(N-1)!} + \frac{1}{2} \cdot \frac{\alpha N\Gamma(N\alpha)}{|\frac{x-\delta}{\gamma}|^{N\alpha+1}(N-1)!} \right| \tag{4.29}$$

for $1 < \alpha < 2$

$$|A(N-1, x, \alpha, \gamma, \delta)| \geq |A(N, x, \alpha, \gamma, \delta)| \tag{4.30}$$

where

$$A(N, x, \alpha, \gamma, \delta) = \frac{\Gamma(N\alpha)}{(k-1)! \cdot |\frac{x-\delta}{\gamma}|^{N\alpha+1}} \cdot \left(1 + \frac{1}{2}\alpha N\pi + \alpha N\Psi(N\alpha) - \alpha N ln(|\frac{x-\delta}{\gamma}|)\right) \tag{4.31}$$

Lemma 4.3. *The derivative of density functions with respect to the stable index parameter α can be approximated by the following function:*

$$\frac{\partial f(x)}{\partial \alpha} \approx g_\alpha(x) = \begin{cases} R^\alpha_{[2d,2d+2]}(x), & |\frac{x-\delta}{\gamma}| < x_c^\alpha \\ \sum_{k=1}^N h_k(x, \alpha, \gamma, \delta), & |\frac{x-\delta}{\gamma}| \geq x_c^\alpha \end{cases} \tag{4.32}$$

where d is a positive integer, and

$$h_k(x, \alpha, \gamma, \delta) =$$
$$\frac{(-1)^k\Gamma(k\alpha)}{(k-1)! \cdot |\frac{x-\delta}{\gamma}|^{k\alpha+1}} \cdot \left[\frac{sin(\frac{k\pi\alpha}{2})}{\gamma\pi}\left(1 + \alpha k\Psi(k\alpha) + \alpha k ln(|\frac{x-\delta}{\gamma}|)\right) + \frac{\alpha cos(\frac{k\pi\alpha}{2})k}{2\gamma}\right]$$

The connecting points x_c^α are determined by (4.29)-(4.31).

With the above functions defined in Lemma 4.1-4.3, we can approximate the score function as following:

Lemma 4.4. *The score function $s(\alpha, \gamma, \delta)$ can be approximated by the following functions:*

$$s(\alpha, \gamma, \delta) = \left(\frac{\partial l}{\partial \alpha}, \frac{\partial l}{\partial \gamma}, \frac{\partial l}{\partial \delta}\right)$$

$$\frac{\partial l}{\partial \gamma} \approx \sum_{i=1}^n g_\gamma(x_i) \cdot \frac{1}{f(x_i|\alpha, \gamma, \delta)} \tag{4.33}$$

$$\frac{\partial l}{\partial \delta} \approx \sum_{i=1}^n g_\delta(x_i) \cdot \frac{1}{f(x_i|\alpha, \gamma, \delta)} \tag{4.34}$$

$$\frac{\partial l}{\partial \alpha} \approx \sum_{i=1}^n g_\alpha(x_i) \cdot \frac{1}{f(x_i|\alpha, \gamma, \delta)} \tag{4.35}$$

where $g_\gamma(x_i)$, $g_\delta(x_i)$, and $g_\alpha(x_i)$ are defined respectively in (4.15), (4.22), and (4.32).

Applying the same idea of taking the derivative of the power and asymptotic series twice theoretically allows one to calculate the second order derivative of the density functions, and approximate the observed information matrix, which makes it possible to utilize the Newton's method in optimization. However, approximating the observed information matrix requires defining six extra functions and makes the algorithm complicated and slow in calculation. Besides, non-derivative optimization methods such like Nelder-Mead method was also tried, but it is slower in convergence than quasi-Newton method, requiring more time in computation. Thus, considering the computational convenience, we will go with the quasi-Newton method in this thesis. The algorithm for calculating the maximum likelihood estimates using a quasi-Newton method is introduced in the next part.

4.2.2 Computational algorithm

As stated above, the quasi-Newton method (BFGS) will be used to find the maximum value of the log-likelihood functions, and quantile estimator will be used as the initial point. We denote the parameters by $\vec{\theta} = (\alpha, \gamma, \delta)$ and the density functions by $f(x|\vec{\theta})$. The parameter space is $\Theta = (0, 2] \times (0, \infty) \times (-\infty, \infty)$.

For a given sample $(x_1, x_2, \dots x_n)$, the maximum likelihood estimator $\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\delta})$ can be calculated through the following steps:

- Calculate McCulloch's quantile estimator $\tilde{\theta} = (\tilde{\alpha}, \tilde{\gamma}, \tilde{\delta})$ from the formula (4.1)-(4.4), and set $\tilde{\theta}$ as the initial value $\theta^{(0)}$. If α exceeds the available range, i.e. $\alpha < 0.6$, and cannot be obtained in the table then randomly choose $\alpha^{(0)} \in (0, 0.6]$

- For each point $\theta^{(k)}$, choose the proper degree for Padé approximants and connecting points, and calculate the density function $f(x_i|\theta^{(k)})$ from lemma 3.2 and lemma 3.3, and the first order derivative $g_\gamma(x_i)$, $g_\delta(x_i)$ and $g_\alpha(x_i)$ from lemma 4.1-4.3.

- Compute the score function $s(\theta^{(k)})$ using Lemma 4.4, and quasi-Newton direction $\Delta_k = -H_k^{-1} s(\theta^{(k)})$, where H_k is an approximation to the Hessian matrix $(-I(\theta^{(k)}))$.

- Set step size $t = 1$, and compute the updated point $\theta^{(k+1)}$ as $\theta^{(k+1)} = \theta^{(k)} + \Delta_k$. If α exceeds the boundary of the parameter space, set it as the extreme value, i.e. set $\alpha = 2$ if $\alpha > 2$, and $\alpha = 0.01$ if $\alpha < 0$.

- Check and see whether the iteration satisfies the stopping criteria: the number of iteration meets the limit, or $\|\Delta_k\|$ is less than the preset error bound. Stop the process if any of the above conditions is satisfied.

- Compute the updated H_{k+1}:

$$H_{k+1} = H_k + \frac{y_{k+1}y_{k+1}^T}{y_{k+1}^T\Delta_k} - \frac{H_k\Delta_k\left(H_k\Delta_k\right)^T}{\Delta_k^T H_k\Delta_k} \tag{4.36}$$

  where $y_{k+1} = s(\theta^{(k+1)}) - s(\theta^{(k)})$, and H_0 is chosen as nI, where n is the sample size and I is the identity matrix.

  or directly its inverse H_{k+1}^{-1} using the Sherman-Morrison formula [29]:

$$H_{k+1}^{-1} = \left(I - \frac{\Delta_k y_{k+1}^T}{y_{k+1}^T\Delta_k}\right)H_k^{-1}\left(I - \frac{y_{k+1}\Delta_k^T}{y_{k+1}^T\Delta_k}\right) + \frac{\Delta_k\Delta_k^T}{y_{k+1}^T\Delta_k} \tag{4.37}$$

4.2.3 Observed information and confidence intervals

Let $X = (x_1, x_2, \ldots x_n)$ denote an i.i.d. stable sample with sample size n, and $\hat{\theta}_n$ denote the maximum likelihood estimator calculated from X. DuMouchel [8] verified the needed regularity conditions, and stated that the asymptotic behavior of $\hat{\theta}_n$ satisfies:

$$\left[I(\hat{\theta}_n; X)\right]^{1/2}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} Z \sim MVN(0_3, I_3)$$

where $I(\hat{\theta}_n; X)$ is the observed information matrix and θ_0 is the true value of the parameter.

One should notice the fact that when $\vec{\theta}$ is near the boundary of the parameter space, the asymptotic behavior of the estimators from finite sample is not well known. When $\alpha = 2$, the asymptotic normal distribution for the estimators becomes a degenerate distribution [7]. When $\vec{\theta}$ is away from the boundary, an approximate $100p\%$ confidence interval for θ_i is given by

$$\left(\hat{\theta}_i - z_p\sqrt{\hat{v}_{ii}}, \hat{\theta}_i + z_p\sqrt{\hat{v}_{ii}}\right) \tag{4.38}$$

where $\hat{v}_{ii}$ is the (i,i) entry of $\left[I(\hat{\theta}_n; X)\right]^{-1}$.

To calculate the observed information matrix $I(\hat{\theta}_n; x)$, one can use $-H_k$, which is introduced in the algorithm in section 4.2.2, to approximate it. In other words, if the iteration stops after s times, then we would use h_{ii}, the (i,i) entry of matrix $-H_s^{-1}$, instead of $\hat{v}_{ii}$ in the formula (4.38) to compute the confidence interval. The advantage of doing this

is the computational convenience. Once the point estimates are obtained, the confidence interval can be determined without extra calculation. However, one should note that h_{ii} is an approximation of $\hat{v}_{ii}$, and the confidence interval calculated from h_{ii} will be different from the true value.

Another approach is to calculate the observed information matrix directly by applying the idea of taking the second order derivative of the power and asymptotic series, and defining an approximate function to the $\sqrt{\hat{v}_{ii}}$ as we did before. It is more accurate, but also more time consuming and complicated in computation. In this thesis, we will use $-H_k$ to obtain the confidence interval, and as can be seen in the next section, the MLE with this confidence interval performs well in the simulation.

4.3 Simulation

A stable data set can be generated using the algorithm of Chambers, Mallows and Stuck [6], and the result of simulations in this section are all based on the data generated in this method.

Generate a stable data set with the parameter $\alpha = 1.3$, $\gamma = 5$, $\delta = 10$, and sample size $n = 1000$. The quantile estimators of the parameters are: $\tilde{\alpha} = 1.2427$, $\tilde{\gamma} = 5.2807$, and $\tilde{\delta} = 10.4882$. The maximum likelihood estimators with 95% confidence intervals are: $\hat{\alpha} = 1.3037 \pm 0.0850$, $\hat{\gamma} = 5.1110 \pm 0.3437$, and $\hat{\gamma} = 10.1719 \pm 0.3618$.

For different values of (α, γ, δ), a single sample with size $n = 1000$ is generated, and the algorithm is run to get the parameter estimates. Table 4.1 below lists the point estimation calculated from quantile estimator and the maximum likelihood estimator. The true value of the parameters is shown in the first column, the quantile estimator $\tilde{\theta}$ in the second, and the MLE $\hat{\theta}$ in the last. The quantile estimator in the first row cannot be obtained because this method is only available for $\alpha > 0.6$ as discussed in section 4.1.

The 95% confidence interval of (α, γ, δ) for the samples above are given in the Table 4.2 below.

As shown in Table 4.1, the quantile estimates as initial values are usually not far from the true value of the parameters. Thus, the iterative procedure converges to a satisfactory point within 4 iterates when $0.6 \leq \alpha \leq 2$. In contrast, for $0 < \alpha < 0.6$, it usually takes more time to get an acceptable estimates by integration. As discussed in section 3.2.3, evaluating the stable density by Padé approximants and asymptotic series is much faster than by Nolan's integral form. Thus, it is not surprising that our method is faster in parameter estimation compared to Nolan's algorithm.

The algorithm is also run on the sample with small size, and the result shows that this method still works. Generating a stable data set with the parameter $\alpha = 1.3$, $\gamma = 5$,

(α, γ, δ)	Quantile estimator	MLE
$(0.25, 1, 0)$		$(0.2477, 1.1713, -0.0515)$
$(0.6, 2.5, 0)$	$(0.5764, 2.5529, 0.0273)$	$(0.5872, 2.4789, 0.0360)$
$(0.8, 5, 0)$	$(0.7654, 5.1516, 0.0444)$	$(0.7843, 4.9387, 0.0214)$
$(0.95, 1, 5)$	$(0.9174, 0.9742, 5.0403)$	$(0.9667, 1.0246, 4.9632)$
$(1.05, 2.5, 5)$	$(1.0354, 2.4331, 4.8359)$	$(1.0700, 2.4794, 5.0541)$
$(1.3, 5, 5)$	$(1.3291, 5.0807, 5.4223)$	$(1.3064, 5.1110, 5.2719)$
$(1.5, 1, 10)$	$(1.3839, 0.9933, 9.9977)$	$(1.4737, 1.0371, 10.0728)$
$(1.7, 2.5, 10)$	$(1.8275, 2.5542, 10.0593)$	$(1.7325, 2.5372, 10.0208)$
$(1.9, 5, 10)$	$(1.9578, 5.2114, 9.9146)$	$(1.8794, 5.0734, 10.1059)$

Table 4.1: Parameter estimates for sample size $n = 1000$

(α, γ, δ)	α	γ	δ
$(0.25, 1, 0)$	0.2477 ± 0.0178	1.1713 ± 0.0725	-0.0515 ± 0.0214
$(0.6, 2.5, 0)$	0.5872 ± 0.0563	2.4789 ± 0.0942	0.0360 ± 0.0612
$(0.8, 5, 0)$	0.7843 ± 0.0703	4.9387 ± 0.2617	0.0214 ± 0.0787
$(0.95, 1, 5)$	0.9667 ± 0.0758	1.0246 ± 0.0963	4.9632 ± 0.0808
$(1.05, 2.5, 5)$	1.0700 ± 0.0816	2.4794 ± 0.2236	5.0541 ± 0.1067
$(1.3, 5, 5)$	1.3064 ± 0.0865	5.1110 ± 0.3845	5.2719 ± 0.9681
$(1.5, 1, 10)$	1.4737 ± 0.1037	1.0371 ± 0.0714	10.0728 ± 0.1062
$(1.7, 2.5, 10)$	1.7325 ± 0.0838	2.5372 ± 0.1365	10.0208 ± 0.9936
$(1.9, 5, 10)$	1.8794 ± 0.0706	5.0734 ± 0.2434	10.1059 ± 0.1096

Table 4.2: Confidence interval for sample size $n = 1000$

$\delta = 10$ and sample size $n = 50$. The quantile estimators of the parameters are: $\tilde{\alpha} = 1.0612$, $\tilde{\gamma} = 5.2510$, and $\tilde{\delta} = 10.6814$. The maximum likelihood estimators with 95% confidence intervals are: $\hat{\alpha} = 1.2436 \pm 0.3667$, $\hat{\gamma} = 5.2032 \pm 1.6585$, and $\hat{\gamma} = 10.2009 \pm 0.4629$.

According to the asymptotic theory, the $100p\%$ confidence interval can be expressed as $(\hat{\theta} \pm z_p SE(\hat{\theta}))$, where $\hat{\theta}$ is the point estimate and $SE(\hat{\theta})$ is the standard error of $\hat{\theta}$. A stable sample with the parameter $\alpha = 1.3$, $\gamma = 1$, $\delta = 0$, and sample size $n = 1000$ is generated, and the parameter estimation is performed based on such data (all three parameters are estimated). The simulation is run repeatedly 5000 times to evaluate the mean and standard error of the maximum likelihood estimator $\hat{\alpha}$: $mean(\hat{\alpha}) = 1.2997$, and $SE(\hat{\alpha}) = 0.0436$. The result is close to what we got in table 4.3, by using approximate observed information matrix to calculate the confidence interval.

(α, γ, δ)	α	γ	δ
$(1.3, 1, 0)$	1.2927 ± 0.0883	1.0257 ± 0.0762	0.0365 ± 0.0845
$(1.3, 1, 5)$	1.3142 ± 0.0844	1.0161 ± 0.0775	5.0433 ± 0.0872
$(1.3, 1, 10)$	1.3158 ± 0.0811	1.0684 ± 0.0743	9.9787 ± 0.0817
$(1.3, 2.5, 0)$	1.3268 ± 0.0903	2.5484 ± 0.1947	-0.0307 ± 0.1038
$(1.3, 2.5, 5)$	1.2895 ± 0.0915	2.5899 ± 0.2071	5.0341 ± 0.1164
$(1.3, 2.5, 10)$	1.3188 ± 0.0850	2.4864 ± 0.2137	10.0394 ± 0.0918
$(1.3, 5, 0)$	1.2888 ± 0.0877	5.1864 ± 0.3849	-0.0377 ± 0.0987
$(1.3, 5, 5)$	1.3064 ± 0.0865	5.1110 ± 0.3845	5.2719 ± 0.9681
$(1.3, 5, 10)$	1.3177 ± 0.0909	5.1622 ± 0.3719	10.1281 ± 0.0975

Table 4.3: Performance of $\hat{\gamma}$ and $\hat{\delta}$ for fixed $\alpha = 1.3$

The performance of $\hat{\alpha}$ varies with different values of α, which is shown in Figure 4.1 respectively when sample size $n = 1000$ and $n = 10000$. For each point in the graph, the algorithm is run 5000 times, and the standard error is calculated as a function of α. In this simulation and the following two, the data are generated with the parameter $\theta = (\alpha, 1, 0)$ (all three parameters are estimated).

As can be seen in the graph, $SE(\hat{\alpha})$ tends to be smaller when α gets closer to the boundary. Our method works quite well for almost all $\alpha \in (0, 2]$, except when α is very close to the boundary ($0 < \alpha < 0.2$ or $1.95 < \alpha < 2$), where the confidence interval will be too narrow to cover the true value. Besides, when $0 < \alpha < 0.2$, one will sometimes observe a very slow convergence of $\hat{\alpha}$ if the true value is far from the initial value. For example, in a single experiment, we generated a stable data set with parameter $(0.15, 1, 0)$ and sample size $n = 1000$. The iteration starts at $(0.5874, 1.3412, 0.0243)$ and the estimates we had after 11 times iteration is: $(\hat{\alpha}, \hat{\gamma}, \hat{\delta}) = (0.1078, 1.3712, -0.2852)$. The example above does not suggest that the estimates will always be disappointing in these cases, but suggest that the result is not as good as those when $0.2 \leq \alpha \leq 2$.

The behaviors of $\hat{\gamma}$ and $\hat{\delta}$ are also related to the value of α. Figure 4.2 and 4.3 indicate the standard error of $\hat{\gamma}$ and $\hat{\delta}$ respectively as a function of α.

Comparing to Nolan's algorithm [3], the maximum likelihood estimators shown above have smaller standard errors. One should note that Nolan's result is from the general stable data (including skewness), and the conclusion of a superior method cannot be made based on such simulation result. However, the idea of using Padé approximants and asymptotic series in approximation of density functions, and parameters estimations, is proved to be reliable in practice.

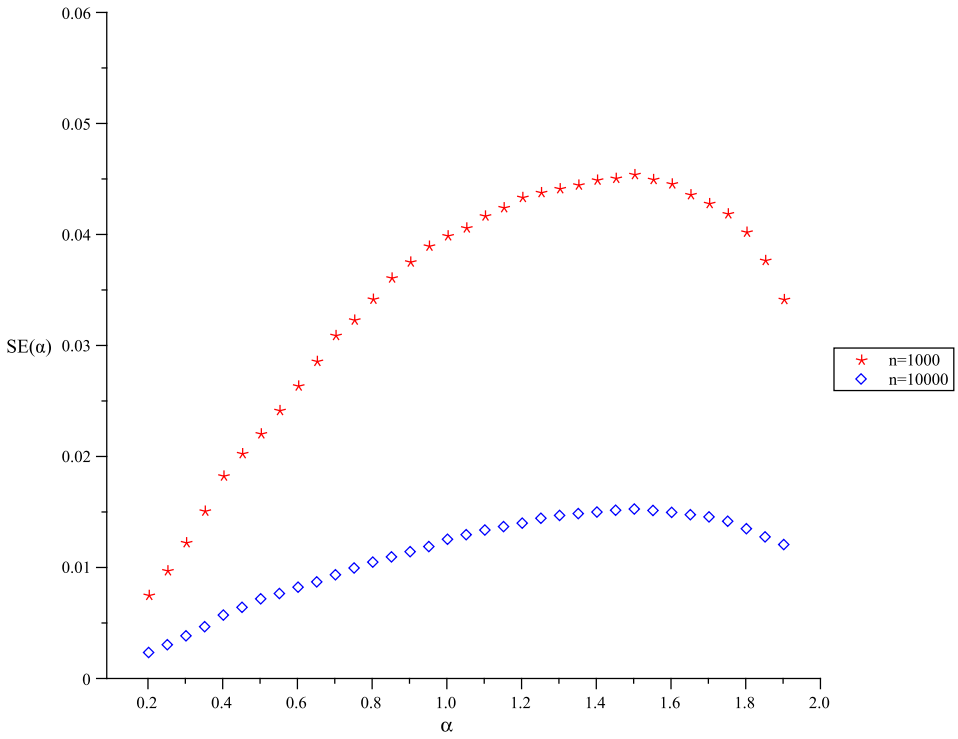

Figure 4.1: The standard error of $\hat{\alpha}$ for different values of α for sample size $n = 1000$ and $n = 10000$

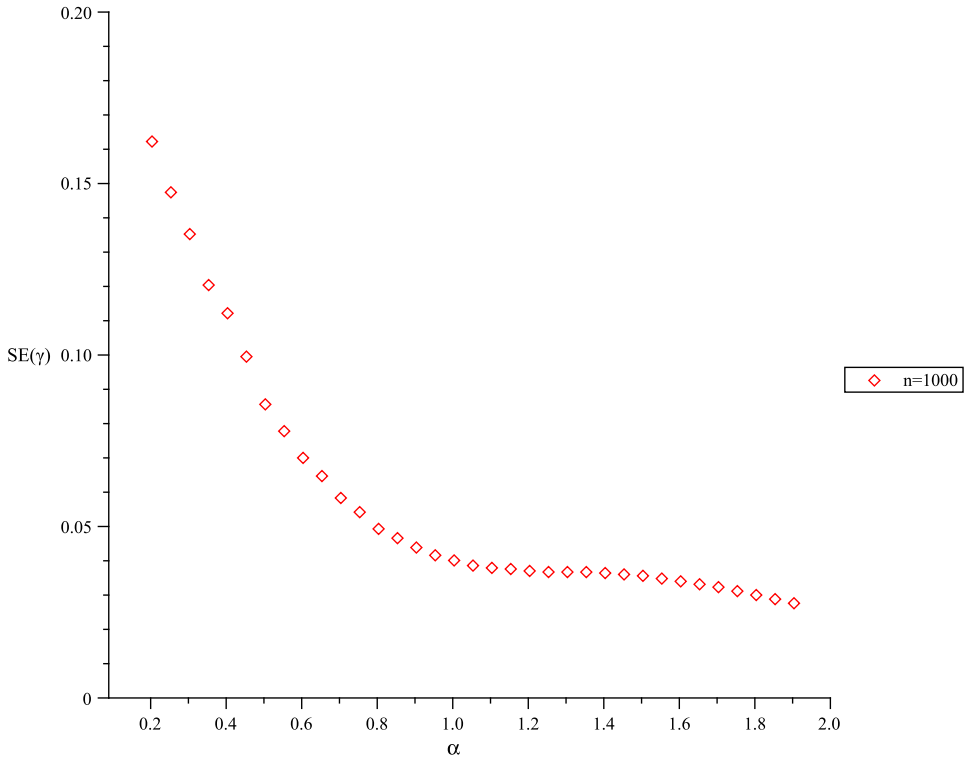

Figure 4.2: The standard error of $\hat{\gamma}$ for different values of α for sample size $n = 1000$

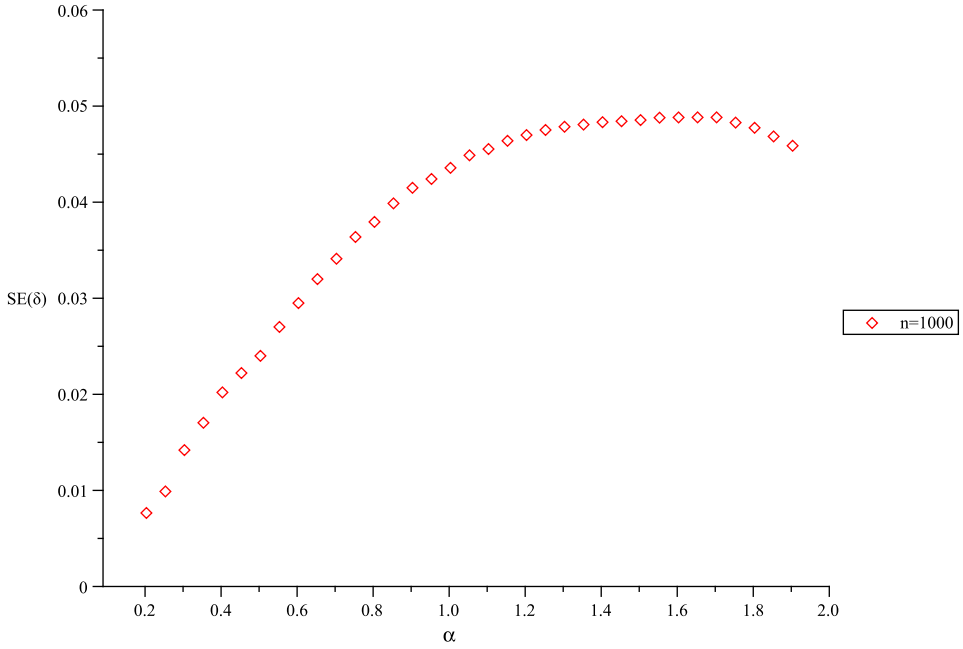

Figure 4.3: The standard error of $\hat{\delta}$ for different values of α for sample size $n = 1000$

Chapter 5

Conclusion and future work

We have introduced in this thesis a new computational method of using Padé approximants and asymptotic series to approximate the density functions of the symmetric stable distributions, which are commonly useful in practice but facing the problem of having no closed-form. Specifically, we derive the method of choosing the proper degree and central point of the Padé approximants, and explore the connecting point to combine the Padé approximants and asymptotic series in order to obtain the optimal approximation. Based on such approximation, an algorithm using quasi-Newton method is developed to calculate the maximum likelihood estimator and confidence interval of the parameters. The simulation results indicate that the new method is fast and accurate in density approximation. In parameter estimation, the algorithm is fast and reliable for $0.2 \leq \alpha \leq 1.95$. Moreover, by comparing to results under Nolan's integral method, which is the most commonly used in evaluating stable densities in practice, the maximum likelihood estimators calculated from this method have a smaller standard error.

The main drawback of this method is that it is restricted for the symmetric stable distributions. Further possible investigation could be carried out to analyse general cases if the power series and asymptotic series are provided. The approximation of general stable density functions and estimation of skewness parameter are available by applying a similar idea.

Appendix A

A.1 Proof of Lemma 3.1

Let $R_{[m,n]}(x)$ denote the Padé approximant of the stable density function with degree (m, n), which is calculated from equations (3.4) to (3.6). For a positive integer d, it follows that

$$R_{[2d,2d+2]}(x) = R_{[2d+1,2d+2]}(x) = R_{[2d,2d+3]}(x) = R_{[2d+1,2d+3]}(x)$$

Proof. We will prove the proposition $R_{[2d,2d+2]}(x) = R_{[2d+1,2d+2]}(x)$ here. By definition 2.2 and power series (3.3), we have:

$$R_{[2d,2d+2]}(x) = P_1(x)/Q_1(x)$$

$$P_1(x) = |P_1| = \begin{vmatrix} 0 & a_0 & 0 & \cdots & 0 \\ a_0 & 0 & a_2 & \cdots & a_{2d+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{2d} & 0 & a_{2d+2} & \cdots & a_{4d+2} \\ 0 & 0 & a_0 x^{2d} & \cdots & \sum_{j=0}^{d} a_{2j} x^{2j} \end{vmatrix}$$

$$Q_1(x) = |Q_1| = \begin{vmatrix} 0 & a_0 & 0 & \cdots & 0 \\ a_0 & 0 & a_2 & \cdots & a_{2d+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{2d} & 0 & a_{2d+2} & \cdots & a_{4d+2} \\ x^{2d+2} & x^{2d+1} & x^{2d} & \cdots & 1 \end{vmatrix}$$

and

$$R_{[2d+1,2d+2]}(x) = P_2(x)/Q_2(x)$$

$$P_2(x) = |P_2| = \begin{vmatrix} a_0 & 0 & a_2 & \cdots & a_{2d+2} \\ 0 & a_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{2d+2} & 0 & \cdots & 0 \\ 0 & a_0 x^{2d+1} & a_0 x^{2d} & \cdots & \sum_{j=0}^{d} a_{2j} x^{2j} \end{vmatrix}$$

$$Q_2(x) = |Q_2| = \begin{vmatrix} a_0 & 0 & a_2 & \cdots & a_{2d+2} \\ 0 & a_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{2d+2} & 0 & \cdots & 0 \\ x^{2d+2} & x^{2d+1} & x^{2d} & \cdots & 1 \end{vmatrix}$$

Finding the LU decomposition of P_1 and P_2: $P_1 = L_1^p U_1^p$, $P_2 = L_2^p U_2^p$, where $U_1^p = \{up_{ij}\}$ and $U_2^p = \left\{up_{ij}^*\right\}$ are upper triangular matrixes.

$P_1(x) = |P_1| = |U_1| = \prod_{i=1}^{2d+3} up_{ii}$ and $P_2(x) = \prod_{i=1}^{2d+3} up_{ii}^*$.

Note that the x^i terms are only contained in the numerator of $up_{2d+3,2d+3}$ and $up_{2d+3,2d+3}^*$ so that $P_1(x) = C_1 p(x)$, where $p(x) = numer(up_{2d+3,2d+3})$ is a polynomial and $C_1 = \prod_{i=1}^{2d+3} up_{ii} \cdot denom(up_{2d+3,2d+3})$ is a constant coefficient. Similarly, $P_2(x) = C_2 p^*(x)$. Note that by elementary column operations, the last column of P_1 and P_2 can be transformed to be the same, which means $p(x) = p^*(x)$.

At the same time, finding the LU decomposition of Q_1 and Q_2: $Q_1 = L_1^q U_1^q$, $P_2 = L_2^q U_2^q$. Note that P_1 and Q_1 are the same except for the last row. Thus, by the same idea before, $Q_1(x) = C_1 q(x)$, and $Q_2(x) = C_2 q(x)$.

Combine the above argument, we have the fact that

$$R_{[2d,2d+2]}(x) = P_1(x)/Q_1(x) = p(x)/q(x) = P_2(x)/Q_2(x) = R_{[2d+1,2d+2]}(x)$$

Similarly, we can prove the rest of Lemma 3.1 by the same argument. □

A.2 Generating stable data

The algorithm for constructing a standard stable random variable in representation (1.4) is given by Weron, R. in 1996 [31]

- generate a random variable V uniformly distributed on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and an independent exponential random variable W with mean 1;

- for $\alpha \neq 1$ compute:

$$X = S_{\alpha,\beta} \cdot \frac{sin\left\{\alpha(V + B_{\alpha,\beta})\right\}}{\left\{cos(V)\right\}^{1/\alpha}} \cdot \left[\frac{cos\left\{V - \alpha(V + B_{\alpha,\beta})\right\}}{W}\right]^{(1-\alpha)/\alpha}$$

where

$$B_{\alpha,\beta} = \frac{arctan(\beta tan(\frac{\pi\alpha}{2}))}{\alpha}$$
$$S_{\alpha,\beta} = \left\{1 + \beta^2 tan^2\left(\frac{\pi\alpha}{2}\right)\right\}^{1/(2\alpha)}$$

- for $\alpha = 1$ compute:

$$X = \frac{2}{\pi}\left\{\left(\frac{\pi}{2} + \beta V\right) tan(V) - \beta ln\left(\frac{\frac{\pi}{2} W cos(V)}{\frac{\pi}{2} + \beta V}\right)\right\}$$

- A stable random variable for all admissible values of the scale and location parameter γ and δ can be easily simulated by:

$$Y = \begin{cases} \gamma X + \delta, & \alpha \neq 1 \\ \gamma X + \frac{2}{\pi}\beta\gamma ln(\gamma) + \delta, & \alpha = 1 \end{cases}$$

To generate the symmetric stable random variable used in the thesis, the formula is simplified to:

$$Y = \gamma\left\{\frac{sin(\alpha V)}{\left\{cos(V)\right\}^{1/\alpha}} \cdot \left[\frac{cos\left\{V(1-\alpha)\right\}}{W}\right]^{(1-\alpha)/\alpha}\right\} + \delta$$

A stable data sample of size $n = 4000$, with parameter $\alpha = 1.3$, $\gamma = 5$ and $\delta = 10$ is generated by the above method, and the histogram is plotted as below:

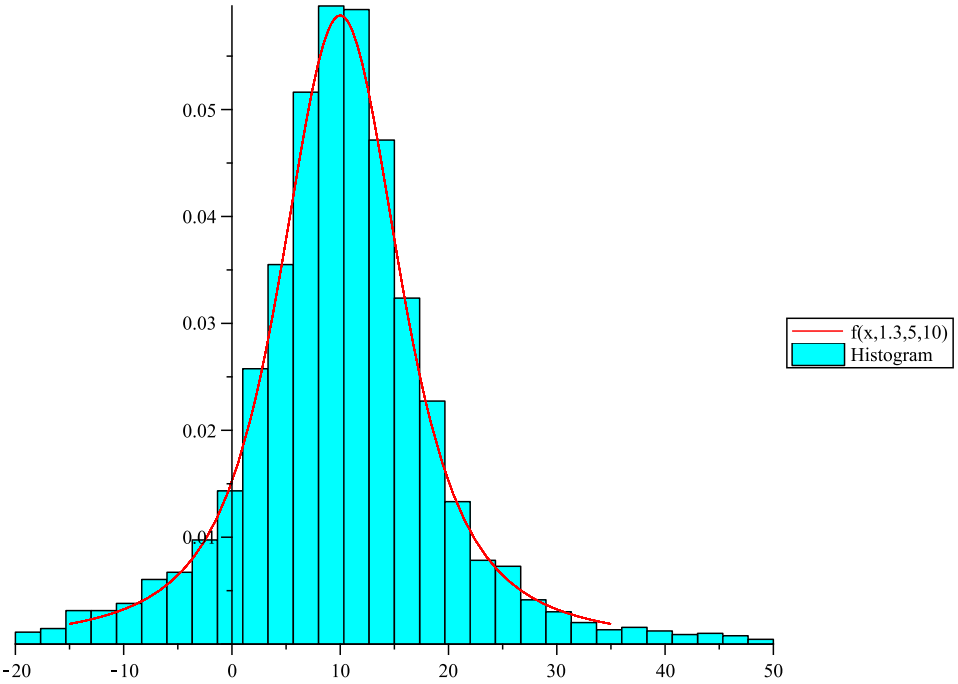

Figure A.1: Histogram of i.i.d. sample with parameter $(1.3, 5, 10)$ and size $n = 4000$

Bibliography

[1] V. Akgiray and C. G. Lamoureux. Estimation of stable law parameters: a comparative study. *J. Bus. Econ. Stat*, 7(85-93), 1989. 29

[2] George A. Baker, JR, and Peter Graves-Morries. *Padé approximants(Encyclopedia of mathematics and its applications)*. Cambridge University Press, Cambridge, second edition, 2010. 7

[3] O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick. *Lévy Processes - Theory and Application*. Birkh 2001. 40

[4] H. Bergström. On some expansions of stable distributions. *Arkiv for Mathematik*, 2(463-474), 1952. 12

[5] D. J. Buckle. Bayesian inference for stable distributions. *Journal of the American Statistical Association*, 90(605-613), 1995. 28

[6] J. M. Chambers, C. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(340-344), 1976. 38

[7] W. H. DuMouchel. Stable distribution in statistical inference. 1971. Unpublished Ph.D. Thesis, Department of statistics, Yale university. 11, 37

[8] W. H. DuMouchel. On the asymptotic normality of the maximum likelihood estimate when sampling from a stable distribution. *Annals of Statistics*, 1(948-957), 1973. 37

[9] W. H. DuMouchel. Estimating the stable index α in order to measure tail thickness: A critique. *Annals of Statistics*, 11(1019-1031), 1983. 28

[10] Paul. Embrechts, C. Klppelberg, and T. Mikosc. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin, 1997. 1

[11] E. F. Fama and R. Roll. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66(331-338), 1971. 29

[12] W Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, New York, ii edition, 1966. 12

[13] H. Fofack and J. P. Nolan. Tail behavior, modes and other characteristics of stable distributions. *Extremes*, 2(39-58), 1999. 28

[14] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(1163-1174), 1975. 28

[15] D. R. Holt and E. L. Crow. Tables and graphs of the stable probability density functions. *Journal of research of the National Bureau of Standards. B. Mathematical Sciences*, 77B(3-4)(143-198), 1973. 11

[16] I. A. Koutrouvelis. Regression type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75(918-928), 1980. 28

[17] Paul. P. Lévy. *Calcul des Probabilites*. Gauthier-Villars, Paris, 1925. 1

[18] J. H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics-Simulations*, 15(1109-1136), 1986. 29, 30

[19] J. H. McCulloch. Financial applications of stable distributions. *Handbook of Statistics*, Vol. 14(Ch. 13), 1996. 1

[20] J. H. McCulloch. Measuring tail thickness to estimate the stable index alpha: A critique. *J. Bus. Econ. Stat*, 15(74-81), 1997. 28

[21] J. H. McCulloch. Characteristic function based estimation of stable parameters. in adler, r., feldman, r. and taqqu, m.(eds). *A Practical Guide to Heavy Tailed Data*, (311-335), 1998. 29

[22] C. L. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley, New York, 1995. 28

[23] J. P. Nolan. Numerical calculation of stable densities and distribution functions. *Communications is Statistics - Stochastic Models*, 13(759-774), 1997. 4, 11

[24] J. P. Nolan. An algorithm for evaluating stable densities in zolotarev's (m) parametrization. *Mathematical and Computer Modelling*, Vol.29(229-233), 1999. 11

[25] J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhäuser, Boston, 2010. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan. 2

[26] S. J. Press. Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, 67(842-846), 1972. 28

[27] S. T. Rachev, Stefan Mittnik, and M. S. Paolella. Diagnosing and treating the fat tails in financial returns data. *Journal of Empirical Finance*, vol. 7(issue 3-4, pages 389-416), 2000. 1

[28] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York, 1994. 28

[29] Jack Sherman and Winifred J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics*, 21(124-127), 1950. 37

[30] Christopher G. Small. *Expansions and Asymptotics for Statistics*. Chapman and Hall/CRC, 2010. 12

[31] R Weron. On the chambers-mallows-stuck method for simulating skewed stable random variables. *Statistics and Probability Letters*, 28(165-171), 1996. 46

[32] V. M. Zolotarev. One-dimensional stable distributions. *Amer. Math. Soc. Transl. of Math. Monographs*, Vol.65, 1986. 11