Tail Behavior, Modes and Other Characteristics of Stable Distributions

Hippolyte Fofack and John P. Nolan
American University

8 January 1998, revised 13 January 1999

Abstract

Stable distributions have heavy tails that are asymptotically Paretian. Accurate computations of stable densities and distribution functions are used to analyze when the Paretian tail actually appears. Implications for estimation procedures are discussed. In addition to numerically locating the mode of a general stable distribution, analytic and numeric results are given for the mode. Extensive tables of stable percentiles have been computed; aspects of these tables and the appropriateness of infinite variance stable models are discussed.

Abbreviated title: Stable tails and modes

1 Introduction

The Paretian tail behavior of stable distributions is the basis for their use as models when heavy tails are observed. The asymptotic behavior is also used to estimate the index of stability α. However, little is actually known about the point at which this asymptotic behavior becomes accurate. Understanding when the approximation is accurate is also of interest for numerical purposes, where tail approximations are useful for fast density calculations, e.g. in maximum likelihood estimation of stable parameters.

In the Gaussian case, $X \sim N(0,1)$, there is a well known formula to approximate the tail probability: as $x \to \infty$,

$$P(X > x) \sim \frac{\exp(-x^2/2)}{x\sqrt{2\pi}}, \tag{1}$$

see Feller (1968, pg. 175). (Here and below, $h(x) \sim g(x)$ as $x \to \infty$ will mean $\lim_{x\to\infty} h(x)/g(x) = 1$.) When $\alpha < 2$, Lévy (1925) has shown that the tails of non-Gaussian stable distributions are asymptotically equivalent to a Pareto law. Specifically, if X is a standardized stable random variable with characteristic exponent $0 < \alpha < 2$ and skewness parameter β, then as $x \to \infty$,

$$P(X > x) \sim (1+\beta)C_\alpha x^{-\alpha} \tag{2}$$

where

$$C_\alpha = \left(2\int_0^\infty x^{-\alpha}\sin x\,dx\right)^{-1} = \frac{1}{\pi}\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)$$

[0]AMS 1991 subject classifications. Primary 60E07, 62E20. Secondary 65D20, 65U05.
[0]Keywords and phrases. Stable distribution, Pareto distribution, characteristic functions, modes, Hill estimator.

(Property 1.2.15, Samorodnitsky and Taqqu (1994)).

One objective of this paper is to investigate the asymptotic tail behavior of non-Gaussian stable distributions (symmetric and asymmetric). In Section 2, accurate calculations of stable densities and distributions functions are used to determine the point at which the tails of a stable distribution are close to Paretian. Section 3 discusses the implications of those results for tail estimators. Although Yamazato (1978) showed that stable distributions are unimodal, it is not known how mode, median and other distribution characteristics are related with the index and the skewness parameters. Section 4 investigates the behavior of stable modes and stable distributions around the mode. Extensive tabulations of general stable percentiles are described in Section 5 and available on the Web. These values are used to discuss the appropriateness of using infinite variance stable models in applications.

We end this introduction with formal definitions of stable distributions. Since densities and distribution functions are not known in closed form for most stable distributions, they are generally specified by their characteristic functions. Four parameters are needed: an index of stability $\alpha \in (0,2]$, a skewness parameter $\beta \in [-1,1]$, a scale parameter $\sigma > 0$, and a location parameter $\mu \in (-\infty,\infty)$. The most popular parameterization is defined by Samorodnitsky and Taqqu: $X \sim S_\alpha(\sigma,\beta,\mu)$ if the characteristic function of X is given by

$$\mathbf{E}\exp(i\theta X) = \begin{cases} \exp\left(-\sigma^\alpha|\theta|^\alpha\left[1 - i\beta\mathrm{sign}(\theta)\tan(\frac{\alpha\pi}{2})\right] + i\mu\theta\right) & \text{if } \alpha \neq 1 \\ \exp\left(-\sigma|\theta|\left[1 + i\beta\frac{2}{\pi}\mathrm{sign}(\theta)\ln|\theta|\right] + i\mu\theta\right) & \text{if } \alpha = 1. \end{cases}$$

For numerical purpose, it is preferable to use a different parameterization: we will say $X^0 \sim S^0_\alpha(\sigma,\beta,\mu^0)$ if the characteristic function of X^0 is given by

$$\mathbf{E}\exp(i\theta X^0) = \begin{cases} \exp\left(-\sigma^\alpha|\theta|^\alpha\left[1 + i\beta\tan(\frac{\alpha\pi}{2})\mathrm{sign}(\theta)((\sigma|\theta|)^{1-\alpha} - 1)\right] + i\mu^0\theta\right) & \text{if } \alpha \neq 1 \\ \exp\left(-\sigma|\theta|\left[1 + i\beta\frac{2}{\pi}\mathrm{sign}(\theta)\ln(\sigma|\theta|)\right] + i\mu^0\theta\right) & \text{if } \alpha = 1 \end{cases}$$

The S^0 parameterization is a variant of the (M) parameterization of Zolotarev (1986), with the characteristic function and hence the density and the distribution function jointly continuous in all four parameters. In particular, modes, percentiles and convergence to Paretian tail behavior vary in a continuous way as α and β vary. The parameters α, β, and σ have the same meaning in both the S and S^0 parameterization, while the location parameters of the two representations are related by $\mu = \mu^0 - \beta(\tan\frac{\alpha\pi}{2})\sigma$ if $\alpha \neq 1$; $\mu = \mu^0 - \beta\frac{2}{\pi}\sigma\ln\sigma$ when $\alpha = 1$. Unlike the S parameterization, S^0 is a scale and location family of distributions: if $Y \sim S^0_\alpha(\sigma,\beta,\mu^0)$, then for any a, b, $aY + b \sim S^0_\alpha(|a|\sigma, (\mathrm{sign}\, a)\beta, a\mu^0 + b)$. This and related issues are discussed in Nolan (1998).

2 Asymptotic equivalence

This section is concerned with determining when the Paretian tail approximation is accurate for a stable distribution. For a symmetric stable distribution ($\beta = 0$), the mode is at $x = 0$ in either parameterization, and it makes sense to directly compare the tail probabilities in (2). However, when $\beta \neq 0$, stable distributions have shifted modes. In Section 4, it is shown that $m(\alpha,\beta) =$ mode of an $S^0_\alpha(1,\beta,0)$ distribution is uniformly bounded, in fact $|m(\alpha,\beta)| \leq 1$. But the mode of $X \sim S_\alpha(1,\beta,0)$ is at $m(\alpha,\beta) + \beta\tan\frac{\alpha\pi}{2}$, and this latter term is very large when α is near 1. While (2) will eventually hold for any fixed α, what we generally care about is the shape of the tail and not the shift. Hence, we will determine when the Paretian tail occurs in the S^0 parameterization, not the S parameterization.

In what follows, we will restrict ourselves to normalized stable distributions in the S^0 parameterization, i.e. $X \sim S^0_\alpha(1,\beta,0)$. The density and tail probability of such an X will be denoted by

$f(x;\alpha,\beta)$ and $\bar{F}(x;\alpha,\beta) = P(X > x)$ respectively. We note three facts about stable tails. First, (2) tells us about lower tails:

$$P(X < -x) = P(-X > x) \sim (1-\beta)C_\alpha x^{-\alpha} \quad \text{as} \quad x \to \infty,$$

because $-X \sim S^0_\alpha(1,-\beta,0)$. Second, (2) can be formally differentiated to show that

$$f(x;\alpha,\beta) \sim \alpha(1+\beta)C_\alpha x^{-(1+\alpha)} \quad as \quad x \to \infty.$$

This is a consequence of the regularity of the density. Lastly, when $\beta = -1$, (2) tells us that the tail decays faster than a power. In this paper, we restrict ourselves to cases where the upper tail is Paretian, i.e. $\beta > -1$; information about the $\beta = -1$ case can be found in Zolotarev (1986) or Samorodnitsky and Taqqu (1994).

We note that equation (1) is in the standard normal parameterization, i.e. $N(0,1) = S_2(\sqrt{1/2},0,0)$. Restating (1) for $X \sim S_2(1,0,0)$ gives

$$\bar{F}(x;2,0) \sim (2/x)f(x;2,0) \quad \text{as} \quad x \to \infty,$$

whereas for $\alpha < 2$ and $\beta > -1$, (2) gives

$$\bar{F}(x;\alpha,\beta) \sim (x/\alpha)f(x;\alpha,\beta) \quad \text{as} \quad x \to \infty.$$

For notational convenience, we describe the Pareto family by,

$$f_{Pareto}(x;\alpha,\beta) = \alpha(1+\beta)C_\alpha x^{-(1+\alpha)} \quad \text{and} \quad \bar{F}_{Pareto}(x;\alpha,\beta) = (1+\beta)C_\alpha x^{-\alpha}.$$

The scale above has been chosen so that $f(x;\alpha,\beta) \sim f_{Pareto}(x;\alpha,\beta)$ and $\bar{F}(x;\alpha,\beta) \sim \bar{F}_{Pareto}(x;\alpha,\beta)$ as $x \to +\infty$. As a density and tail probability, the above formulas hold for $x \geq ((1+\beta)C_\alpha)^{1/\alpha}$; for comparison purposes, we will sometimes use these Pareto formulas for all $x > 0$.

Because there are no explicit formulas for $f(x;\alpha,\beta)$ or $\bar{F}(x;\alpha,\beta)$, we will numerically determine when the approximation is accurate; the next Section will discuss implications for tail estimation. Modifications to the program STABLE described in Nolan (1997) were made to increase accuracy on the tails. The density and d.f. calculations in this paper are highly accurate; double precision arithmetic is used throughout and the relative (not absolute) error is less than 10^{-10}.

We start by focusing on the density. Figure 1 shows graphs of symmetric stable densities with the corresponding Pareto density for $\alpha = 0.7, 1.2, 1.9$. For small α, the Pareto density is always above the symmetric stable one, while for α large, the Pareto crosses the stable one. A convention is needed on when to say f and f_{Pareto} are close. Requiring $|f - f_{Pareto}|$ to be small is not a good measure, because this holds as soon as both are small, which occurs quickly, especially when α is above 1.5. Two measures of closeness are used below.

The first measure is useful for understanding when the tail behavior gives a good estimate of the tail index α. It is simply the ratio

$$L_{pdf}(x) = L_{pdf}(x;\alpha,\beta) = f(x;\alpha,\beta)/f_{Pareto}(x;\alpha,\beta).$$

Equation (2) implies that $L_{pdf}(x) \to 1$ as $x \to \infty$ and the goal is to understand when $L_{pdf}(x)$ gets close to 1. The middle row of Figure 1 shows selected plots of $L_{pdf}(x)$. The second measure is useful for numerical approximation purposes: compute the relative error

$$Relerr_{pdf}(x) = \frac{f(x;\alpha,\beta) - f_{Pareto}(x;\alpha,\beta)}{f(x;\alpha,\beta)}.$$

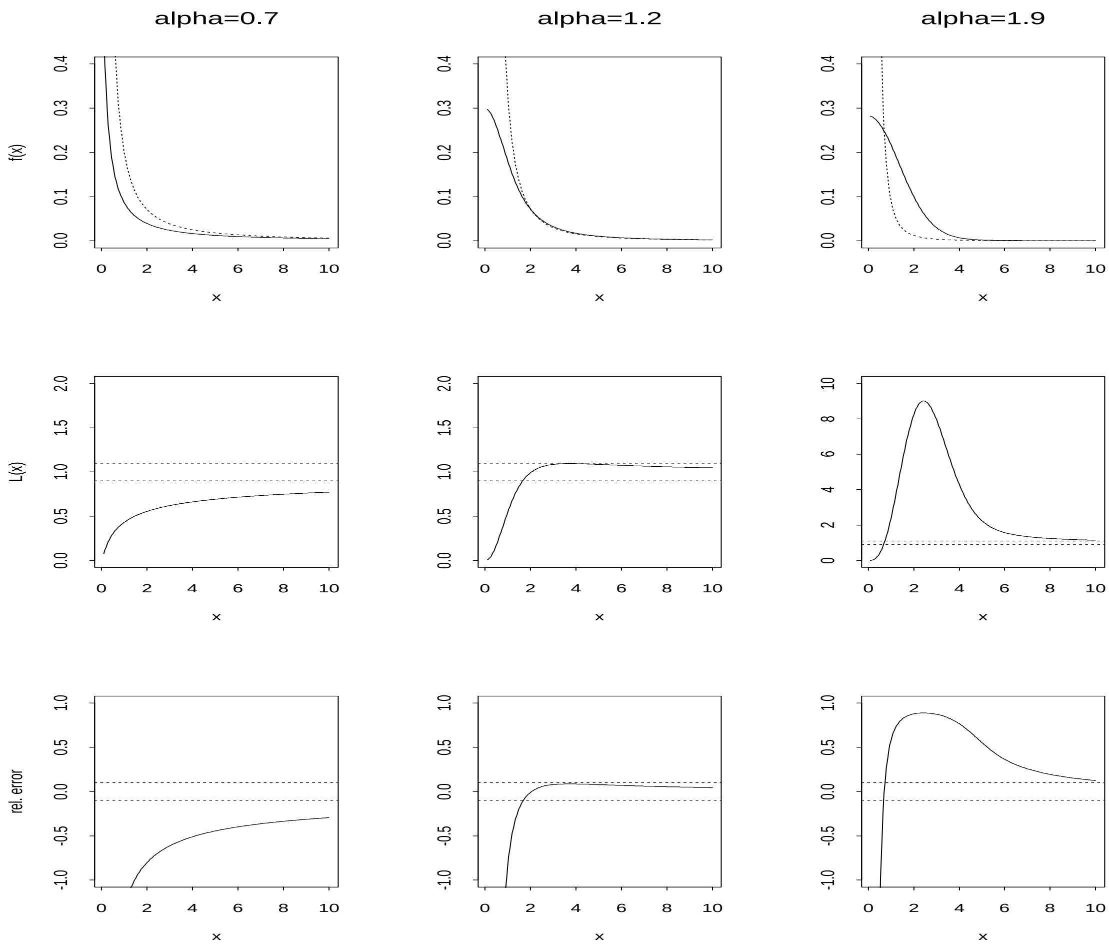

Figure 1: Top row shows symmetric stable densities (solid) versus Pareto density (dotted) for $\alpha = 0.7, 1.2, 1.9$. The middle row shows the ratio $L_{pdf}(x)$ and the bottom row shows the relative error $Relerr_{pdf}(x)$. The middle row plots include a band to indicate $0.9 \leq L_{pdf}(x) \leq 1.1$; the bottom row includes a band to indicate $-0.1 \leq Relerr_{pdf}(x) \leq +0.1$. Note the different vertical scale on the rightmost graph of the middle row.

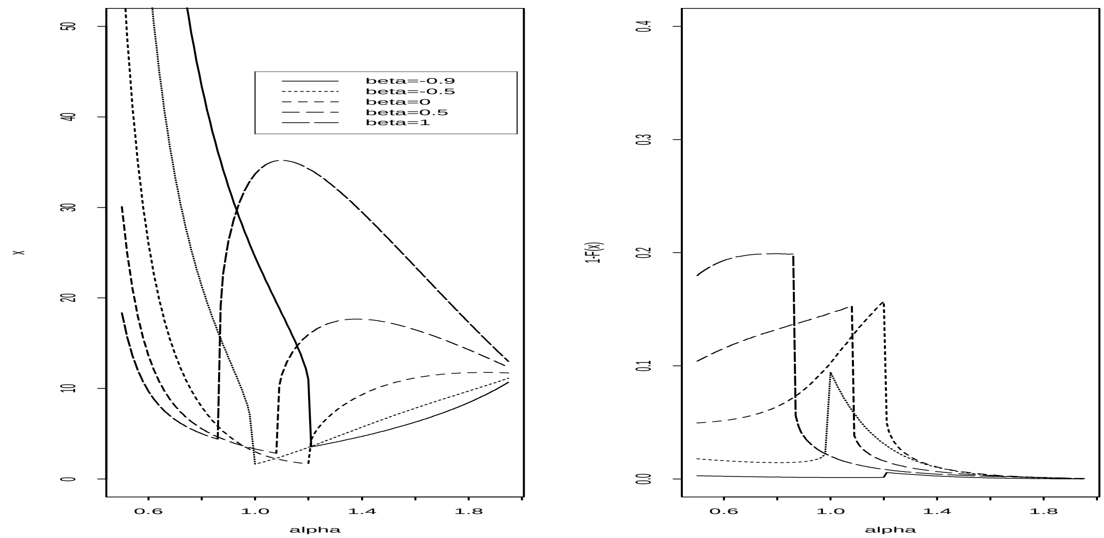

Figure 2: $x_{L_{pdf},0.1}$ (left plot) as a function of α for $\beta = 1, 0.5, 0, -0.5, -0.9$ and $\bar{F}(x_{L_{pdf}})$ (right plot).

The bottom row of Figure 1 shows selected plots of $Relerr_{pdf}(x)$. The relative error tends to 0 as $x \to \infty$ and the goal is understand when it is close to 0.

We have not been able to find a concise way of answering the question of how close the Pareto approximation is to a stable distribution, even for the symmetric case. For $\beta \neq 0$, the behavior is more complex - graphs like Figure 1 change considerably as α and β vary. In an attempt to summarize some part of this behavior, both quantities $L_{pdf}(x)$ and $Relerr_{pdf}(x)$ were computed and a search was done to determine when the functions got close to their limit and that they remain close for all larger x. For a given tolerance $\epsilon > 0$, define the points

$$x_{L_{pdf},\epsilon} = \inf\{x > 0 : \ |L_{pdf}(y) - 1| \leq \epsilon \ \text{ for all } \ y \geq x\} \quad \text{and}$$

$$x_{Relerr_{pdf},\epsilon} = \inf\{x > 0 : \ |Relerr_{pdf}(y)| \leq \epsilon \ \text{ for all } \ y \geq x\}.$$

The above points were numerically located for $\epsilon = 0.1$ and are plotted in Figure 2 and Figure 3 respectively. It is useful to reformulate these raw values of x in terms of the tail probability, so the right plot in these figures show $\bar{F}(x; \alpha, \beta)$.

We were skeptical of the unusual shape of these graphs, but repeated calculations have verified these results. In particular, the $\beta = 0$ curve in Figures 2 and 3 have abrupt changes at around $\alpha = 1.2$ because the stable and Pareto densities cross and depart by more than $\epsilon = 0.1$, forcing $x_{L_{pdf},0.1}$ and $x_{Relerr_{pdf},0.1}$ abruptly to the right. (For example, when $\alpha = 1.2$ and $\beta = 0$, $x_{pdf,0.1} = 1.691$ and $\bar{F}(1.691; 1.2, 0) = 0.154$, but when $\alpha = 1.25$ and $\beta = 0$, $x_{pdf,0.1} = 5.479$ and $\bar{F}(5.479; 1.25, 0) = 0.0337$.) For skewed stable distributions, the location of this abrupt change shifts and its magnitude can be more pronounced.

Using a different measure of closeness instead of relative error would change the exact shape of these curves, but a similar sort of behavior would still result because the Pareto density crosses the stable density and can give a poor approximation for certain values of α and β.

We have repeated this procedure for tail probabilities to determine when $\bar{F}$ and $\bar{F}_{Pareto}$ are close. Figure 4 is similar to Figure 1, comparing tail probabilities $\bar{F}$ and $\bar{F}_{Pareto}$, their ratio $L_{cdf}(x)$ and

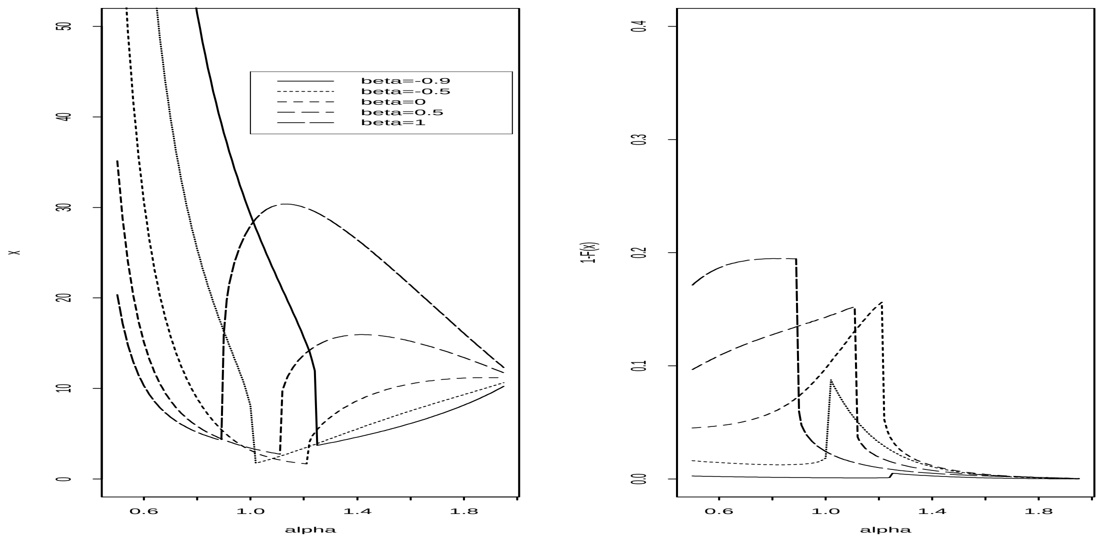

Figure 3: $x_{Relerr_{pdf},0.1}$ (left plot) as a function of α for $\beta = 1, 0.5, 0, -0.5, -0.9$ and $\bar{F}(x_{Relerr_{pdf}})$ (right plot).

the relative error $Relerr_{cdf}(x)$. As above, define

$$x_{L_{cdf},\epsilon} = \inf\{x > 0 : \ |(\bar{F}(y;\alpha,\beta)/\bar{F}_{Pareto}(y;\alpha,\beta)) - 1| \leq \epsilon \text{ for all } y \geq x\} \quad \text{and}$$

$$x_{Relerr_{cdf},\epsilon} = \inf\{x > 0 : \ |\bar{F}(y;\alpha,\beta) - \bar{F}_{Pareto}(y;\alpha,\beta)|/\bar{F}(y;\alpha,\beta) \leq \epsilon \text{ for all } y \geq x\}$$

Figures 5 and 6 shows these points as a function of (α, β) for $\epsilon = 0.1$.

We close this section with some remarks. First, since the plots for x_L and x_{Relerr} are qualitatively similar, it appears that $L(x)$ and $Relerr(x)$ are measuring roughly the same thing. Second, the distribution functions approach the Pareto limit faster than the densities, so any inference about tail behavior is likely to more accurate if the empirical d.f. is used than if the empirical density is used. Third, when a stable distribution is highly skewed, i.e. $|\beta|$ is close to 1, the light tail takes a very long time before the Paretian behavior appears. And finally, when $\alpha \uparrow 2$, the Paretian tail behavior does not occur until the tail probability $\bar{F}(x_{cdf};\alpha,\beta)$ is very small, making any tail estimator unreliable unless a massive data set is available. Intuitively, when α is near 2, stable distributions approach the normal distribution and the Paretian tail is only evident on the extreme tails. This will be illustrated with the Hill estimator next.

3 Implications for tail index estimation

We start with a discussion of the simplest estimate of the index of stability α. It is common to take the top $p\%$ of a sample to estimate α by plotting the upper $p\%$ on a log-log scale and estimating the slope. The accuracy of this or any other tail procedure depends on (i) shifts in the distribution and (ii) on when the Paretian tail behavior starts.

For data coming from stable distribution, some guidelines can be given for the first issue. Shifting the data to center on the mode or median will generally lead to a quicker convergence to Paretian

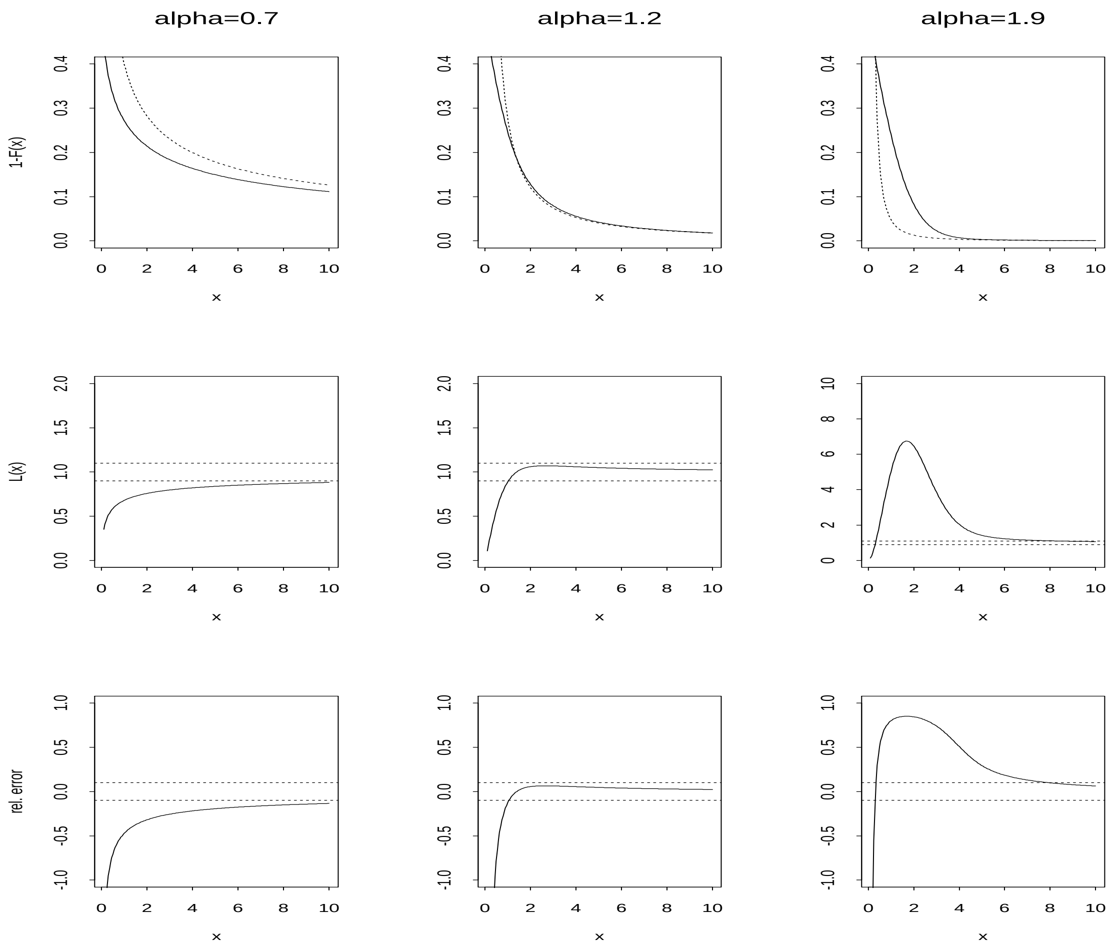

Figure 4: Top row shows symmetric stable d.f. (solid) versus Pareto d.f. (dotted) for $\alpha = 0.7, 1.2, 1.9$. The middle row shows the ratio $L_{cdf}(x)$ and the bottom row shows the relative error $Relerr_{cdf}(x)$. The middle row plots include a band to indicate $0.9 \leq L_{cdf}(x) \leq 1.1$; the bottom row includes a band to indicate $-0.1 \leq Relerr_{cdf}(x) \leq +0.1$. Note the different vertical scale on the rightmost graph of the middle row.

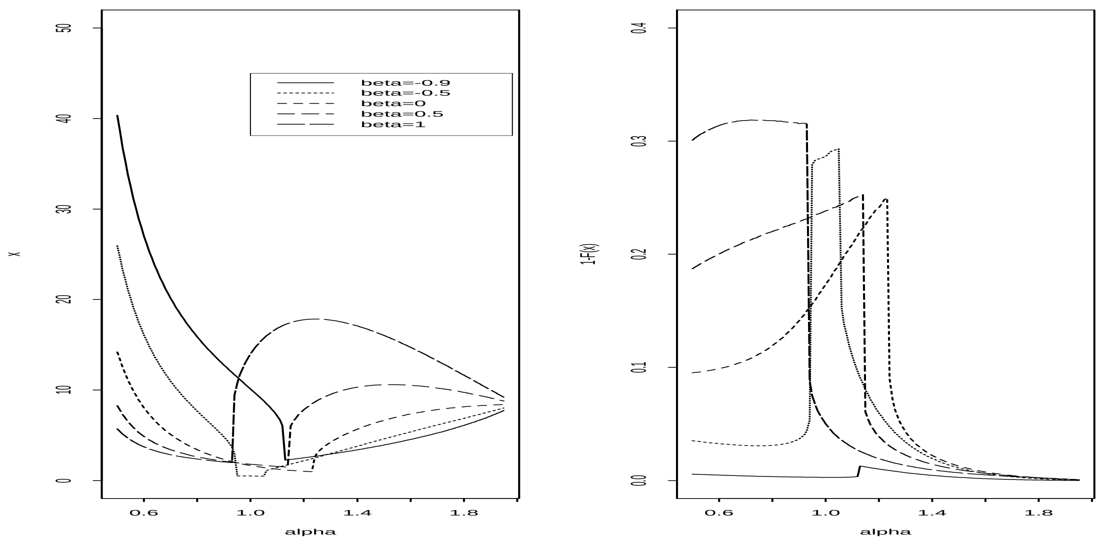

Figure 5: $x_{L_{cdf},0.1}$ (left plot) as a function of α for $\beta = 1, 0.5, 0, -0.5, -0.9$ and $\bar{F}(x_{L_{cdf}})$ (right plot).

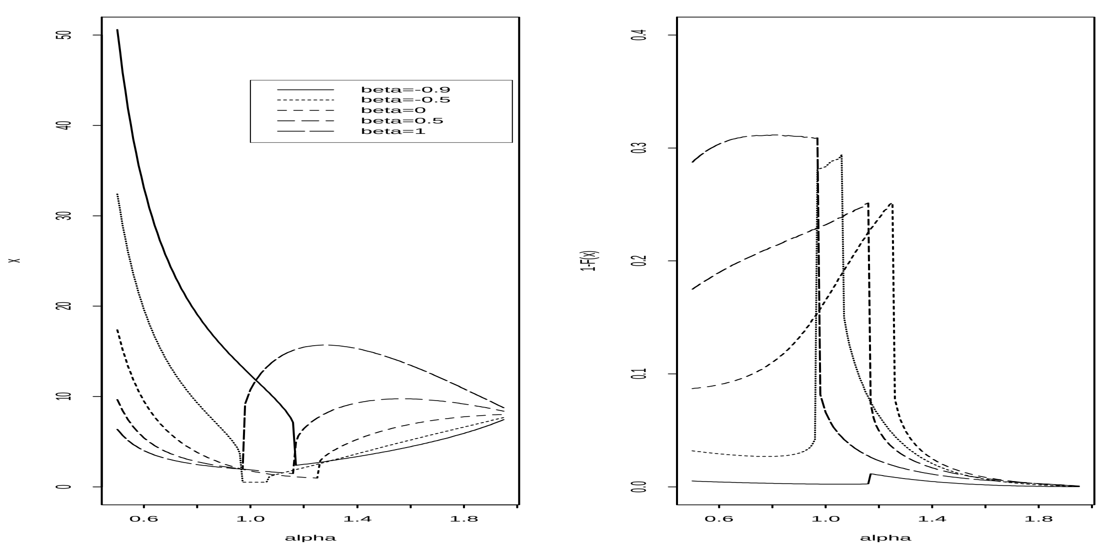

Figure 6: $x_{Relerr_{cdf},0.1}$ (left plot) as a function of α for $\beta = 1, 0.5, 0, -0.5, -0.9$ and $\bar{F}(x_{Relerr_{cdf}})$ (right plot).

tails. In contrast, little advice can be given on point (ii): the Figures in the previous Section imply that the value of p one should take is very dependent on α and β. If one knows what α and β are, then these figures can be used to give a conservative value of p to use. Of course, this is not helpful in a general estimation procedure where the purpose is to estimate α and β.

Hill (1975) proposed a method for estimating the tail index that does not assume that a parametric form holds for the entire distribution function, but focuses only on the tail behavior. The Hill estimator is used to estimate the Pareto index α, when the (upper) tail of the distribution has form $\bar{G}(x) \approx Cx^{-\alpha}$. Here the goal is to study the behavior of Hill's estimator when it is applied to non-Gaussian stable distributions.

Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be the order statistics drawn from a population with distribution G, ordered so that $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$. The Hill estimate of α based on the k largest values is:

$$\hat{\alpha}_{Hill\ ,k} = \left(\frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}} \right)^{-1}$$

In practice, the ordered pairs $(k, \hat{\alpha}_{Hill,k})$ are plotted and one looks for a region where the plot levels off to identify the correct α. This procedure is widely used, see Resnick (1997), Resnick and Stărică (1995, 1996), and Embrechts, Klüppelberg and Mikosch (1997), Reiss and Thomas (1997). In order to determine how the Hill estimate is affected by α for a stable sample, we generated 10000 symmetric standardized ($\beta = 0, \sigma = 1, \mu^0 = 0$) stable random variables for different values of α=.5, 1.2, 1.9 using the algorithm of Chambers, Mallows and Stuck (1975), then used the positive values in the sample as basis for deriving the Hill estimate. In order to remove the random effects associated with simulations, the Hill estimator of α was derived from the same number of (evenly spaced) stable quantiles as well. Hence, convergence of the Hill estimator to the true parameter estimate is assessed from two different data sets: a sample of 10000 simulated standardized symmetric stable random variables with $\alpha = .5, 1.2, 1.9$ and 10000 exact quantiles from the same stable distribution.

From Figures 5 and 6, it is seen that in the symmetric case ($\beta = 0$) convergence of a stable distribution to a Pareto is quickest when $\alpha = 1.2$, where the corresponding tail probability is maximal. On the other hand, when $\alpha \to 2$, the tail probability $\bar{F}(x_{pdf}; \alpha, \beta)$ converges to 0. This implies that an accuracate value from the Hill estimator will require a very large sample size when α is large, because only the extreme tail has Paretian behavior.

The Hill estimator for these data are shown in Figure 7, for both quantile and simulated data. The true value of the parameter α is known in the three cases, and the issue is to estimate what percent p of the (upper) distribution is required to ensure convergence of the Hill estimate $\hat{\alpha}$ to the true parameter value. This indirectly determines the size n of a sample needed to estimate α: if p is very small, then n must be very large before np, the approximate number of points in the upper tail required to get a good estimate of α, is appreciable. In the first case considered here, $\alpha = 0.7$, Figures 5 and 6 both show that it takes a while before either $L_{cdf}(x)$ or $Relerr_{cdf}(x)$ get close to their limiting values. However, the Hill estimator actually does quite well using a large portion of the sample. An explanation for this can be found in Figure 4: when $\alpha = 0.7$, $L_{cdf}(x)$ never gets very far from 1, in contrast to the $\alpha = 1.9$ case discussed below. As expected, when $\alpha = 1.2$, the Hill estimator is close to the true index α for a large proportion of the sample. However, when $\alpha = 1.9$, the Hill estimator never gives a reasonable estimate of α. This can be understood by studying the plot of $L_{cdf}(x)$ from the previous section. Formally, the stable tail is exactly $\bar{F}(x; \alpha, \beta) = L_{cdf}(x)\bar{F}_{Pareto}(x; \alpha, \beta)$. Resnick (1997), considers the Hill estimator for distributions of this type when $L_{cdf}(x)$ is slowly varying. In the stable case, we have more than slowly varying because $L_{cdf}(x) \to 1$ as $x \to \infty$. Yet even with this stronger condition, the Hill estimator performs poorly because $L_{cdf}(x)$ gets quite large and takes a long time before it converges to 1. For example, when $\alpha = 1.9$ and $\beta = 0$, $x_{L_{cdf},0.1} = 8.33$

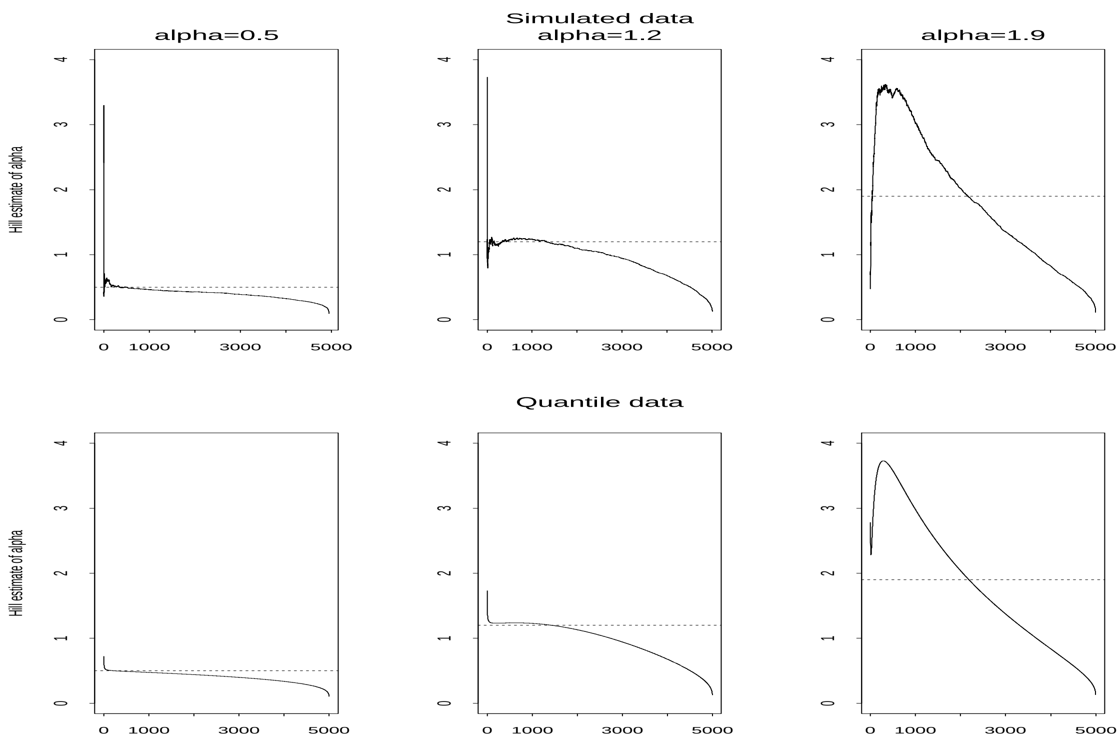

Figure 7: Hill plots of tail index estimate for symmetric stable distributions when $\alpha = 0.5, 1.2, 1.9$ using simulated data (top row) and exact quantiles (bottom row).

and more importantly $\bar{F}(8.33; 1.9, 0) = 0.00093864$. This last value shows that a massive sample is needed to show the Paretian tail behavior when α is close to 2. The situation gets more involved if we consider the nonsymmetric case, e.g. values of $L_{cdf}(x)$ in the hundreds occur for highly skewed data.

For a stable distribution with $\alpha > 1.5$, the Hill estimator generally does a poor job of estimating α. Resnick (1997) includes one such stable distribution in his "Hill Horror Plot" for this reason. This issue has been examined by others. McCulloch (1997) shows that using the Hill estimator (or the generalized Pareto model) on stable data when $1 < \alpha < 2$ leads to overestimates of α. Reiss and Thomas (1997) also discuss why the tail estimator performs poorly for stable data with $\alpha > 1$. Even with the smoothed Hill plots and Alternate Hill plots (change of horizontal scale) described by Resnick (1997), it is not clear how useful the Hill plot is for stable data.

In fact, we see little reason to use the Hill estimator for a stable distribution, since there are now reliable estimation procedures for stable data. Kogon and Williams (1998) improve the Koutrouvelis method, which uses regression on the sample characteristic function. Nolan (1999) describes efficient maximum likelihood estimation for all four stable parameters and shows that the maximum likelihood estmate of α is highly efficient for α near 2, in contrast with the poor performance of the Hill estimator in that region. Of course, if the data is not stably distributed, these procedures may mislead more than the Hill estimator. While there are no formal tests for assessing the stablity of a data set, there are diagnostics for assessing the goodness of fit with a stable model in Nolan (1999).

For an arbitrary (not necessarily stable) distribution having a Paretian tail, it is impossible to give general guidelines on what percentage of the tail is needed by the Hill estimator to get a good approximation of the tail index. One can easily construct examples which will mislead any estimation procedure. For example, a mixture of a normal and a non-Gaussian stable with density $p_1 f(x; 2, 0) + p_2 f(x; \alpha, 0)$ has tail index α from the stable term. However, if p_2 is small, then one is unlikely to get $\hat{\alpha}$ near α, simply because most of the data set is from the normal component. (This is completely separate from the question considered above of how well one can estimate α from a pure stable sample.) Goldie and Smith (1987) discuss the remainder term when the distribution is asymptotically Pareto.

4 Modes of stable distributions

Yamazato (1978) and Sato and Yamazato (1978) proved that all stable distributions are unimodal. But it is not known how the mode of a stable distribution is related to the parameters. In this section, we investigate how the mode varies with tail index and skewness parameters. Also, values of the density and percentiles of the distribution at the mode are computed. In the S^0 parameterization, modes are well behaved because of the joint continuity, and it suffices to calculate the mode of the standardized density because the mode of a general distribution can be obtained by simple scaling and shifting.

Let $m(\alpha, \beta)$ be the mode of $X \sim S^0_\alpha(1, \beta, 0)$. For a symmetric stable distribution, it is known that $m(\alpha, 0) = 0$, see Wintner (1936). There is also an exact value for the mode when $X \sim$ Lévy distribution. In the S^0 parameterization, the Lévy density is $f(x; 1/2, 1) = (x+1)^{-3/2} \exp(-1/(2(x+1)))/\sqrt{2\pi}$, for $x > -1$, and $f'(x; 1/2, 1) = 0$ precisely when $x = -2/3$. Hence, $m(1/2, 1) = -2/3$ and $m(1/2, -1) = 2/3$. Modal values are not known for any other stable distributions. By symmetry, $m(\alpha, -\beta) = -m(\alpha, \beta)$, so it suffices to consider $0 \leq \beta \leq 1$. Figure 8 provides accurate calculations of the mode for general α and β, obtained by numerically maximizing $f(x; \alpha, \beta)$ as computed by the program STABLE. Estimated accuracy in the values of $m(\alpha, \beta)$ is ± 0.0001.

Next we give an analytic result on modes that is a modification of Theorem (2.1) of Hall (1984)

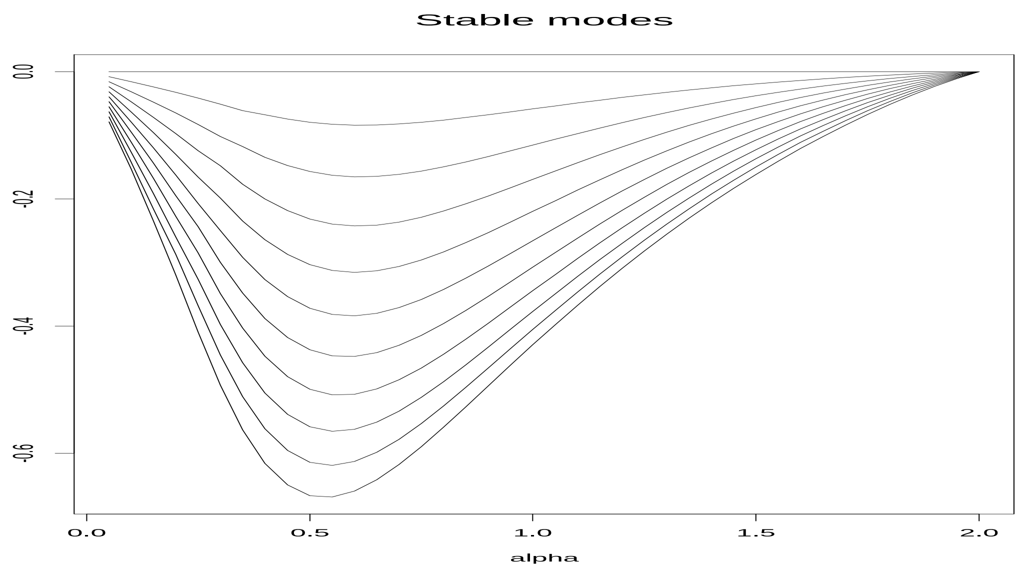

Figure 8: Modes of standardized stable distribution $m(\alpha, \beta)$ versus α with β increasing from $0, 0.1, \ldots, 1$ from top to bottom.

and Equation (87) of Gawronski and Wiessner (1992). Since $m(\alpha, 0) = 0$ for all α, the rate at which $m(\alpha, \beta)$ goes to zero as $\beta \to 0$ is given by $g(\alpha) = \lim_{\beta \to 0} m(\alpha, \beta)/\beta = \partial m(\alpha, 0)/\partial \beta$.

Lemma 1: If $g(\alpha)$ is the rate at which the mode $m(\alpha, \beta)$ goes to zero as $\beta \to 0$, then

$$g(\alpha) = \begin{cases} (\tan \frac{\alpha\pi}{2})[\frac{\Gamma(1+(2/\alpha))}{\Gamma(3/\alpha)} - 1] & \text{if } \alpha \neq 1 \\ \frac{2\gamma - 3}{\pi} & \text{if } \alpha = 1 \end{cases}$$

where $\gamma \approx 0.57721$ is Euler's constant.

Proof: Let $\alpha \neq 1$ and $c = c(\alpha) = \tan \frac{\pi\alpha}{2}$. Using the inversion theorem, e.g. Theorem 1 of Nolan (1997), the pdf of a standardized S^0 stable random variable is:

$$f(x; \alpha, \beta) = \frac{1}{\pi} \int_0^\infty \cos[xt + \beta c(t - t^\alpha)] e^{-t^\alpha} dt$$

so

$$\frac{\partial f(x; \alpha, \beta)}{\partial x} = -\frac{1}{\pi} \int_0^\infty \sin[xt + \beta c(t - t^\alpha)] t e^{-t^\alpha} dt$$

We want $m(\alpha, \beta)$ such that $\partial f/\partial x(m(\alpha, \beta); \alpha, \beta) = 0$. Fix α and define the following series for $\mu(\cdot) = m(\alpha, \cdot)$:

$$\mu(\beta) = \sum_{j=0}^\infty \mu_j(\alpha)\beta^j = \sum_{j=0}^\infty \mu_{2j+1}(\alpha)\beta^{2j+1} = \frac{\mu_1}{c} c\beta + \sum_{j=1}^\infty \mu_{2j+1}\beta^{2j+1},$$

where the even terms drop out because $m(\alpha,\cdot)$ is odd. At the mode,

$$\begin{aligned} 0 &= \int_0^\infty \sin\left[\mu(\beta)t + \beta c(t - t^\alpha)\right] te^{-t^\alpha} dt \\ &= \int_0^\infty \sin\left[\beta c[(\frac{\mu_1}{c}+1)t - t^\alpha] + t\sum_{j=1}^\infty \mu_{2j+1}\beta^{2j+1}\right] te^{-t^\alpha} dt. \end{aligned}$$

As $\beta \downarrow 0$, the above integral is

$$\begin{aligned} &\int_0^\infty \sin\left[\beta c[(\frac{\mu_1}{c}+1)t - t^\alpha] + t\sum_{j=1}^\infty \mu_{2j+1}\beta^{2j+1}\right] te^{-t^\alpha} dt \\ &\simeq \int_0^\infty \left[\beta c[(\frac{\mu_1}{c}+1)t - t^\alpha] + tO(\beta^3)\right] te^{-t^\alpha} dt \\ &\simeq \beta c\int_0^\infty \left[(\frac{\mu_1}{c}+1)t^2 - t^{\alpha+1}\right] e^{-t^\alpha} dt = 0 \\ &\Rightarrow (\frac{\mu_1}{c}+1)\int_0^\infty t^2 e^{-t^\alpha} dt = \int_0^\infty t^{\alpha+1}e^{-t^\alpha} dt \end{aligned}$$

so

$$(\frac{\mu_1}{c}+1)\frac{\Gamma(3/\alpha)}{\alpha} = \frac{\Gamma(1+2/\alpha)}{\alpha}.$$

Solving for $g(\alpha) = \mu_1(\alpha)$ gives the result when $\alpha \neq 1$. The $\alpha = 1$ case is proved in Gawronski and Wiessner (1992). □

In contrast to other parameterizations $g(\alpha)$ is well behaved around $\alpha = 1$: $\lim_{\alpha\to 1} g(\alpha) = g(1) = (2\gamma - 3)/\pi$. This is shown by using properties of the gamma function: $\Gamma(p+1) = p\Gamma(p)$ and $\Gamma(1+x) \sim 1 - \gamma x$ for x small. We conjecture that $m(\alpha,\beta)$ is decreasing and convex in $\beta > 0$. If this conjecture is true, it yields uniform bounds on the mode: for $\beta \geq 0$, $\beta g(\alpha) \leq m(\alpha,\beta) \leq \beta m(\alpha,1) \leq 0$.

For a standardized stable random variable, values of the density and the d. f. at the mode are known only for symmetric stable and Lévy cases. Otherwise, it is not known how the density and d.f. at $m(\alpha,\beta)$ vary with α and β. In the symmetric case, $m(\alpha,\beta) = 0$, and it is known, see Zolotarev (1986), that $f(0;\alpha,0) = \Gamma(1+1/\alpha)/\pi$, which tends to ∞ as $\alpha \downarrow 0$. For general β, $f(m(\alpha,\beta);\alpha,\beta)$ also tends to infinity as $\alpha \downarrow 0$, so it is difficult to display modal values of densities. The following shows that $f(m(\alpha,\beta);\alpha,\beta)$ is bounded by $f(0;\alpha,0)$. Because of this result, values of $f(m(\alpha,\beta);\alpha,\beta)/f(0;\alpha,0)$ are plotted against α in Figure 9.

Lemma 2: If $f(x;\alpha,\beta)$ is the pdf of a standardized stable random variable, then for all $-\infty < x < \infty$ and $-1 \leq \beta \leq 1$,

$$f(x;\alpha,\beta) \leq f(0;\alpha,0) = \Gamma(1+1/\alpha)/\pi$$

where $f(0;\alpha,0)$ is the height of a symmetric α-stable density at its mode.

Proof: As above, a stable pdf is expressed by:

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\pi}\int_0^\infty e^{-t^\alpha}\cos[(x-\zeta)t + \zeta t^\alpha]dt & \text{if } \alpha \neq 1 \\ \frac{1}{\pi}\int_0^\infty e^{-t}\cos[xt + \beta\frac{2}{\pi}t\ln|t|]dt & \text{if } \alpha = 1 \end{cases}$$

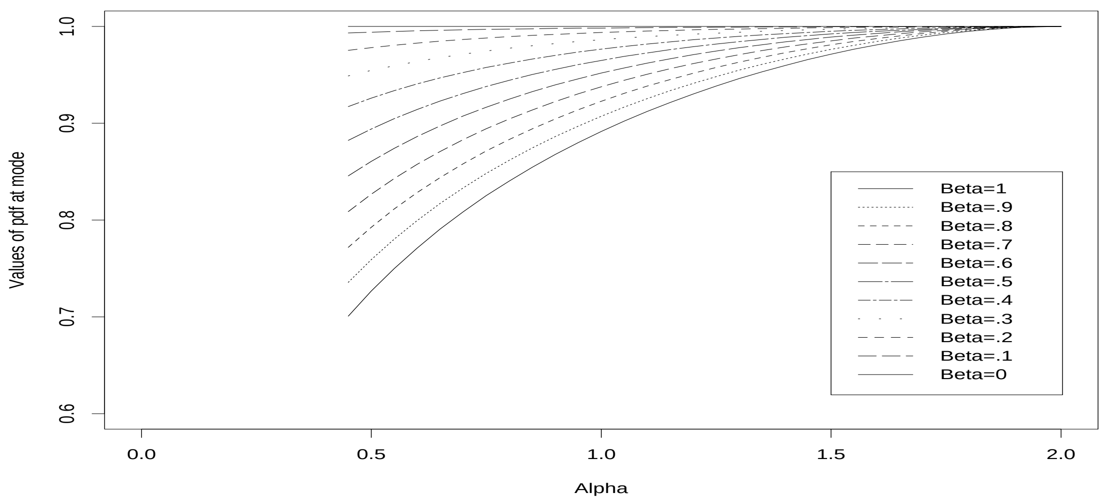

Figure 9: Scaled value of density at the mode of standardized stable random variables $f(m(\alpha,\beta);\alpha,\beta)/f(0;\alpha,0)$ as a function of α for $\beta = 0, 0.1, \ldots, 1$.

When $\alpha \neq 1$, for all x and β, $f(x;\alpha,\beta) \leq \frac{1}{\pi}\int_0^\infty e^{-t^\alpha}dt = f(0;\alpha,0)$, since $\cos[(x-\zeta)t + \zeta t^\alpha] \leq \cos(0) = 1$. Similarly, when $\alpha = 1$, for all x and β, $f(x;\alpha,\beta) = \frac{1}{\pi}\int_0^\infty e^{-t}\cos[xt + \beta\frac{2}{\pi}t\ln|t|]dt \leq \frac{1}{\pi}\int_0^\infty e^{-t}dt = f(0;1,0)$ $\square$

One measure of asymmetry of a distribution is how much of the distribution is to the left of the mode, $F(m(\alpha,\beta);\alpha,\beta)$ = percentile at the mode. Figure 10 shows this quantity for varying α and β.

5 Percentiles and Appropriateness of Stable Models

Extensive tables of percentiles have been computed for general stable distributions in the S^0 parameterization for α=0.1,0.2,. . .,1.9,1.95,1.99,2.0 and β=0,0.1, 0.2,. . ., 0.9,1. (Symmetry can be used for negative beta). The quantiles are tabulated for p-values of 0.00001 to 0.00010 (step 0.00001), 0.0001 to 0.0010 (step 0.0001), 0.001 to 0.010 (step 0.001), 0.01 to 0.10 (step 0.01), 0.10 to 0.90 (step 0.1), and upper tail similar to lower tail. These tables can be downloaded from the Web at `http://www.cas.american.edu/~jpnolan`. Here we discuss a few aspects of stable distributions in terms of those percentiles.

First, there is interest among practitioners in knowing the median of a general stable distribution. Figure 11 shows the median as a function of α and β.

The exact support of stable distributions is known: if $f(x;\alpha,\beta,\sigma,\mu^0)$ is the density of an $S^0_\alpha(\sigma,\beta,\mu^0)$ distribution, adapting results from Samorodnitsky and Taqqu (1994) in the standard

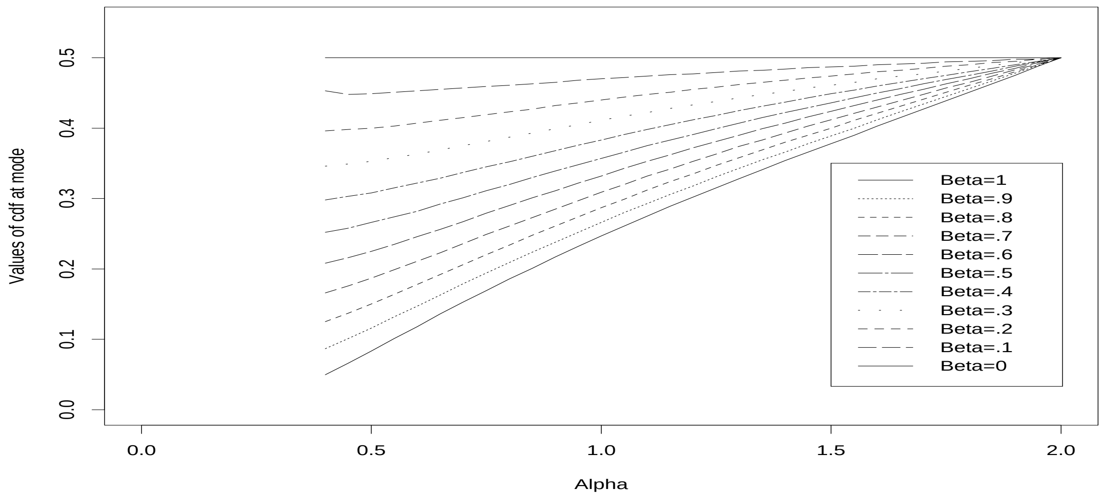

Figure 10: Percentile at mode $F(m(\alpha,\beta);\alpha,\beta)$ as a function of α for $\beta = 0, 0.1, \ldots, 1$.

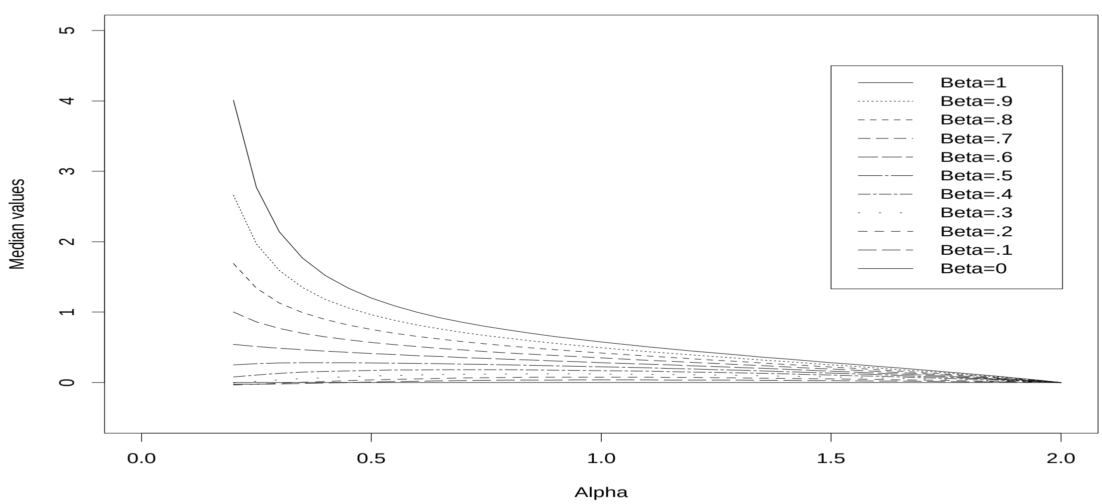

Figure 11: Median of standardized stable random variables as a function of α for $\beta = 0, 0.1, \ldots, 1$.

parameterization,

$$\text{support } f(x;\alpha,\beta,\sigma,\mu^0) = \begin{cases} (\mu^0 - \sigma\tan\frac{\alpha\pi}{2}, \infty) & \text{if } \alpha < 1 \text{ and } \beta = 1 \\ (-\infty, \mu^0 + \sigma\tan\frac{\alpha\pi}{2}) & \text{if } \alpha < 1 \text{ and } \beta = -1 \\ (-\infty, +\infty) & \text{otherwise.} \end{cases}$$

Note that for totally skewed distributions ($\beta = \pm 1$) when $\alpha < 1$, the endpoint of the support goes to $(\text{sign}\beta)\infty$ as $\alpha \uparrow 1$. It can be shown that the mode and most of the distribution stays concentrated near μ^0, so that only a very small probability is near the endpoint of the support, because the light tail in the totally skewed cases decays faster than Pareto by (2). In particular, for $1 - \epsilon < \alpha < 1$, a positive $S_\alpha(1, 1, 0)$ distribution has very little probability in any interval $[0, a]$.

Distributions with heavy tails are regularly seen in applications in economics, finance, insurance, telecommunication and physics. Examples can be found in Resnick (1997), Embrechts, Klüppelberg, and Mikosch (1997), Adler, Feldman and Taqqu (1998), and Rachev, Kim and Mittnik (1997). Even if we acknowledge that large data sets have heavy tails, is it ever reasonable to use a stable model? One of the arguments against using stable models is that they have infinite variance, which is inappropriate for real data that have bounded range. However, bounded data are routinely modeled by normal distributions which have infinite support. The only justification for this is that the normal distribution gives a usable description of the shape of the distribution, even though it is clearly inappropriate on the tails of any bounded data set. The same justification can be used for stable models: does a stable fit gives an accurate description of the shape of the distribution? The variance is one measure of spread; the scale σ in a stable model is another. Perhaps practioners are so used to using the variance as *the* measure of spread, that they automatically retreat from models without a variance. The scale parameter σ can play a similar role for stable models. Of course, all four parameters $(\alpha, \beta, \sigma, \mu^0)$ are necessary to completely describe the distribution.

What we expect from a fit to a data set may depend on the particular application and on the size of the data set. For discussion sake, suppose we are interested in the middle 98% of the distribution and are willing to tolerate an inappropriate model on the upper and lower 1%. We will call the interval $(x_{0.01}, x_{0.99})$ the "approximate range" of the distribution in what follows. Table 1 shows the approximate range for a standardized stable distribution $S^0_\alpha(1, \beta, 0)$ for $\beta = 0$ and $\beta = 1$. When $\alpha < 2$, the distribution is naturally more spread out than the normal ($\alpha = 2$) case, but the range is not outrageous, even for α as small as 1. With a large data set, one should demand a close match between data and mode for a larger percentage of the data, but there will always be a point at which the bounded and discrete nature of the data will make it difficult to justify any model. In Nolan (1999), diagnostics are suggested for assessing the stability of a data set. Here we just note that heavy tailed data sets inherently have more variability on the tails, so standard pp-plots or qq-plots can be misleading.

The authors would like to thank the anonymous reviewers who corrected several typographical mistakes and provided references to related work.

α	$\beta = 0$	$\beta = 1$
0.1	$(-5.1424 \times 10^{16}, 5.1424 \times 10^{16})$	$(-0.16, 5.548 \times 10^{19})$
0.2	$(-1.7837 \times 10^{8}, 1.7837 \times 10^{8})$	$(-0.32, 5.8728 \times 10^{9})$
0.3	(-273949.61,273949.61)	(-0.49,2822032.3)
0.4	(-10812.94,10812.94)	(-0.67,62385.20)
0.5	(-1559.73,1559.73)	(-0.84,6364.87)
0.6	(-429.22,429.22)	(-1.01,1392.84)
0.7	(-170.56,170.56)	(-1.17,470.72)
0.8	(-85.14,85.14)	(-1.33,208.42)
0.9	(-49.41,49.41)	(-1.48,110.30)
1.0	(-31.82,31.82)	(-1.62,66.02)
1.1	(-22.07,22.07)	(-1.77,43.12)
1.2	(-16.16,16.16)	(-1.91,30.01)
1.3	(-12.31,12.31)	(-2.06,21.81)
1.4	(-9.66,9.66)	(-2.21,16.45)
1.5	(-7.74,7.74)	(-2.37,12.65)
1.6	(-6.28,6.28)	(-2.53,9.84)
1.7	(-5.15,5.15)	(-2.70,7.66)
1.8	(-4.28,4.28)	(-2.88,5.86)
1.9	(-3.67,3.67)	(-3.07,4.36)
2.0	(-3.28,3.28)	(-3.28,3.28)

Table 1: Approximate range $(x_{0.01}, x_{0.99})$ of $S^0_\alpha(0,1,0)$ stable distributions for $\beta = 0, 1$.

References

[1] R. J. Adler, R. E. Feldman, and M. S. Taqqu (1998) *A Practical Guide to Heavy Tails.* Birkhäuser, Boston.

[2] J. M. Chambers, C. Mallows and B. W. Stuck (1976) A Method for Simulating Stable Random Variables. *JASA* **71**, 340-344

[3] P. Embrechts, C. Klüppelberg, and T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, NY.

[4] W. Feller (1968) *An Introduction to Probability Theory and its Applications, Vol. I, 2nd Edition*, John Wiley and Sons, NY, NY.

[5] W. Gawronski and M. Wiessner (1992) Asymptotic and inequalities for the mode of stable laws. *Statistics and Decisions* **10**, 183-197.

[6] C. M. Goldie and R. L. Smith (1987) Slow variation and the remainder term: A survey of the theory and its applications. *Quart. J. Math. Oxford Ser.* **38**, 45-71.

[7] B. Hill (1975) A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**, No. 5, 1163-1174.

[8] P. Hall (1984) On unimodality and rates of convergence for stable laws. *J. London Math. Society* **2**, No. 30, 371-384.

[9] Kogon, S. M. and Williams, D. B. (1998), Characteristic function based estimation of stable parameters. In Adler, R., Feldman, R. and Taqqu, M. (eds.) *A Practical Guide to Heavy Tailed Data*, Birkhäuser, Boston, MA.

[10] P. Lévy (1925) *Calcul des Probabilités.* Gauthier Villars, Paris.

[11] McCulloch, J. H. (1997) Measuring Tail Thickness to Estimate the Stable Index Alpha: A Critique *J. Bus. Econ. Stat.*, **15**: 74-81.

[12] T. Mikosch (1997) Heavy tailed modelling in insurance *Comm. in Stat.-Stochastic Models* **13**, 799-896.

[13] J. P. Nolan (1997) Numerical computation of stable densities and distribution functions. *Comm. in Stat.- Stochastic Models.* **13**, 759-774.

[14] J. P. Nolan (1998) Parameterizations and modes of stable distributions. *Statistics and Probability Letters.* **38**, 187-195.

[15] J. P. Nolan (1999) Maximum likelihood estimation of stable parameters. *Submitted.*

[16] S. J. Rachev, J. R. Kim, and S. Mittnik (1997) Econometric modelling in the presence of heavy-tailed innovations. *Comm. in Stat.- Stochastic Models.* **13**, 841-886.

[17] R.-D. Reiss and M. Thomas (1997) *Statistical analysis of extreme values.* Birkhäuser, Boston.

[18] S. Resnick (1997) Heavy tailed modeling and teletraffic data. *Ann. Stat* **25**, 1805-1869.

[19] S. Resnick and C. Stărică (1995) Consistency of Hill's estimator for dependent data. *Journal of Applied Probability*, **32**, 139-167.

[20] S. Resnick and C. Stărică (1996) Tail index for dependent data. *Submitted.*

[21] G. Samorodnitsky and M. S. Taqqu (1994) *Stable Non-Gaussian Random Processes.* Chapman and Hall, NY.

[22] K. Sato and M. Yamazato (1978) On distribution functions of Class L. *Wahr. verw. Gebiete* **43**, 273-308.

[23] A. Wintner (1936) On the class of Fourier Transforms. *Am. J. of Math.*, **58**, 45-90.

[24] M. Yamazato (1978) Unimodality of infinitely divisible distributions of class L. *Annals of Probability.* **6**, 523-531.

[25] V. M. Zolotarev (1986) *One-dimensional Stable distributions. Volume 65 of Translations of Mathematical Monographs*, American Mathematical Society.

Department of Mathematics and Statistics
American University
4400 Massachusetts Avenue, NW
Washington, DC 20016

LaTeX'ed June 16, 2000