Very Skewed Cauchy Distribution: A New Heavy-Tailed Member of Exponential Family

Vahid Nassiri[i] and Adel Mohammadpour[i,ii]

ABSTRACT

In this article we introduce a new class of skewed Cauchy distributions, called very skewed Cauchy, and study properties of this class. Also, we make inference on skewness parameter. We compare this new class with skew Cauchy distribution and the skewed Cauchy distribution introduced by Behboodian et al. in the year 2006. We show that this new class model skew data better than skew Cauchy.

KEYWORDS

Stable family of distributions, skew Cauchy, skew normal, skewness.

1. Introduction

Many real world data that have been collected are skewed. On the other hand, there are some outliers lie in such data. For instance, consider the traffic data of a computer network, obviously in some hours of the day such as 10-12 or 14-16 it is very heavy but in some other hours such as 1-6 there is just little traffic.

Many methods have been developed in order to model data with outliers, or skewed data. A simplest way to achieve symmetry is to exclude the outliers. But a suitable technique to analyze these kinds of data is using a proper model which considers outliers.

Reference [3] extended normal distribution to a general skewed class, called skew-normal. This family of distributions was developed by many authors, such as [4, 5, 6, 10, 15, 18, 21]. This class of distributions is not such heavy tail which may be needed. In order to solve this problem skewed Cauchy is developed by some authors like [1, 7, 13, 16]. But still there are some shortcomings such as difficulties to make inference on skewness parameter or fitting skew data properly.

In this paper we introduce a new class of skewed Cauchy distributions, called *very skewed Cauchy.* This class has two main advantages: 1. It is a member of exponential family (has many inferential advantages), 2. It models skew data better than skew Cauchy (1-stable).

In Section 2 we define a very skewed Cauchy random variable and study some properties of it. In Section 3 we make inference on skewness parameter. Section 4 dedicates to some comparisons among very skewed Cauchy, skew Cauchy and skewed Cauchy distributions. Finally, in Section 5, some concluding remarks are provided.

2. Very skewed Cauchy distribution and some properties

For simplicity computational sake, we introduce a very skewed Cauchy probability density function, pdf, in the following steps. In fact these steps may explain our motivation.

1. Consider Cauchy kernel: $\dfrac{1}{1+x^2}, x \in \mathbb{R}$.

2. Multiply Cauchy kernel by an exponential term, which simplifies our computations, that is: $\dfrac{1}{1+x^2}e^{-\lambda \arctan(x)}$; where λ is a positive parameter. Note that $\dfrac{d}{dx}\arctan(x) = \dfrac{1}{1+x^2}$.

3. Calculate the normalizing constant, which is equal to $\dfrac{\lambda}{2\sinh(\lambda\pi/2)}$.

Based on the previous steps, the pdf of a very Skewed Cauchy random variable can be defined as follows:

[i] Department of Statistics, Faculty of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), 424, Hafez Ave., Tehran 15914, Iran.

[i,ii] Statistical Research and Training Center (SRTC), Tehran, Iran. (e-mail address: adel@aut.ac.ir)

Definition 1 *A random variable X has very skewed Cauchy (VSC) distribution with parameters λ, $X \sim VSC(\lambda)$, if its pdf has the following form:*

$$f_X(x) = \frac{\lambda}{2\sinh(\lambda\pi/2)} \frac{1}{1+x^2} e^{-\lambda \arctan(x)}, \quad x \in \mathbb{R}, \lambda > 0 \tag{1}$$

From (1) an explicit expression for the cumulative distribution function, cdf, of X can be obtained as:

$$F_X(x) = \frac{1}{2\sinh(\frac{\lambda\pi}{2})}(e^{\lambda\pi/2} - e^{-\lambda \arctan(x)}), \quad x \in \mathbb{R}, \lambda > 0 \tag{2}$$

It can be shown that:

$$F_X^{-1}(x) = \tan(-\frac{1}{\lambda}\ln(e^{\lambda\pi/2} - x(2\sinh(\lambda\pi/2)))) \tag{3}$$

We can use (3) by several means, e.g., finding quantiles or generate random numbers. If $P(X \le q) = p$, then we have:

$$q = \tan(-\frac{1}{\lambda}\ln(e^{\lambda\pi/2} - p(2\sinh(\lambda\pi/2)))) \tag{4}$$

Also we can use *probability integral transform theorem* to generate a random sample, $X_1, \ldots, X_n$, from VSC distribution with parameter λ. We have:

$$X_i = \tan(-\frac{1}{\lambda}\ln(e^{\lambda\pi/2} - U_i(2\sinh(\lambda\pi/2)))), \quad i = 1, \ldots, n$$

where U_i is a random number generated from $Uniform[0,1]$ distribution.

Theorem 1 *Let X be a VSC random variable with parameter λ. Then it is unimodal.*

Proof. To prove that $f_X(x;\lambda)$, VSC density, is unimodal we find maximum of $\ln(f_X(x;\lambda))$. $\frac{\partial \ln f_X(x;\lambda)}{\partial x} = 0 \Rightarrow x = -\frac{\lambda}{2}$, and $\frac{\partial^2 \ln f_X(x;\lambda)}{\partial x^2} = 0$ if and only if $x^2 + x\lambda - 1 = 0$. The last equation has two real roots $x_1 = \frac{-\lambda - \sqrt{\lambda^2+4}}{2}, x_2 = \frac{-\lambda + \sqrt{\lambda^2+4}}{2}$. We have $x_1 < -\lambda/2 < x_2$ and $\frac{\partial^2 \ln f_X(x;\lambda)}{\partial x^2}$ is negative for any point between x_1 and x_2. Therefore, $\frac{-\lambda}{2}$ is the unique mode.□

Corollary 1 *If $X \sim VSC(\lambda)$ then $mode(X) = -\lambda/2$.*

Theorem 2 *Let $X \sim VSC(\lambda)$ and $Y \sim Cauchy(0,1)$, if $\lambda \to 0$ then X and Y are identically distributed.*

Proof. By using the Hospital's theorem the proof is straightforward. □

The VSC pdf can be written as $h(x)c(\lambda)e^{w(\lambda)t(x)}$, where $h(x) = \frac{1}{1+x^2} \ge 0$, $c(\lambda) = \frac{\lambda}{2\sinh(\lambda\pi/2)} \ge 0$, $w(\lambda) = \lambda \; (\lambda \in \mathbb{R}^+)$ $\quad t(x) = -\arctan(x) \; (x \in \mathbb{R})$. Thus VSC belongs to exponential family (see [14], p. 23) and $-\sum_{i=1}^n \arctan(X_i)$ is a complete sufficient statistic for λ. The expectation and variance of $T_i = \arctan(X_i)$ are:

$$E(T_i) = \frac{1}{\lambda} - \frac{\pi}{2}\coth(\lambda\pi/2), Var(T_i) = \frac{\pi^2}{4}csch^2(\lambda\pi/2) - \frac{1}{\lambda^2}$$

Theorem 3 *If $X \sim VSC(\lambda)$ then $E|X|^r$ does not exist if $r \ge 1$.*

Proof. By Lyapounov's inequality, it suffices to prove that $E(|X|) = \infty$. We have:

$$E(|X|) = \int_{-\infty}^{+\infty} |x| \frac{\lambda}{2\sinh(\lambda\pi/2)} \frac{1}{1+x^2} e^{-\lambda \arctan(x)} dx$$

First consider the case $X > 0$, we have:

$$E(X) = \frac{-1}{2\sinh(\lambda\pi/2)} \int_0^{+\infty} x \frac{-\lambda}{1+x^2} e^{-\lambda \arctan(x)} dx$$

Now by taking $x = u$, $\frac{-\lambda}{1+x^2} e^{-2\lambda \arctan(x)} dx = dv$ and using integration by parts, we have:

$$\int_0^{+\infty} x \frac{-\lambda}{1+x^2} e^{-\lambda \arctan(x)} dx = x e^{-\lambda \arctan(x)} \Big|_0^{\infty} - \int_0^{+\infty} e^{-\lambda \arctan(x)} dx$$

Now we just need to calculate $\int_0^{\infty} e^{-\lambda \arctan(x)} dx$. We know that:

$$-\frac{\pi}{2} \le \arctan(x) \le \frac{\pi}{2} \Leftrightarrow \infty = \int_0^{+\infty} e^{-\frac{\lambda\pi}{2}} dx \le \int_0^{+\infty} e^{-\lambda \arctan(x)} dx$$

For the case $X \le 0$ a similar method can be applied, and the proof is completed. □

3. Making Inference on Skewness Parameter

In order to draw inference on skewness parameter of VSC distribution, first we find maximum likelihood estimator of λ. For this purpose we should maximize (5). There exist standard methods for this target.

$$l(\lambda) = (\frac{\lambda}{e^{\lambda\pi/2} - e^{-\lambda\pi/2}})^n e^{-\lambda \sum_{i=1}^n \arctan(x_i)} \prod_{i=1}^n \frac{1}{1+x_i^2} \tag{5}$$

The proof of following lemma is straightforward.

Lemma 1 Consider $g(x) = \frac{1}{x} - \frac{\pi}{2}\coth(x\pi/2), x > 0$. Then:

(*i*) $g(x)$ is strictly decreasing.

(ii) $g(x) \in (-\frac{\pi}{2}, 0)$

Theorem 4 *Let $X_1,\ldots,X_n$ be a random sample from a VSC distribution, and $x_1,\ldots,x_n$ be the corresponding observation. Then MLE(λ) exists if and only if* $-n\pi/2 \le \sum_{i=1}^{n} \arctan(x_i) < 0$.

Proof. Let

$$h(\lambda) = \frac{\partial}{\partial\lambda} \ln l(\lambda) = n(\frac{1}{\lambda} - \frac{\pi}{2}\coth(\lambda\pi/2)) - \sum_{i=1}^{n} \arctan(x_i) = ng(\lambda) + r(x)$$

By Lemma 1, $ng(\lambda)$ is strictly decreasing and lies in $(-n\pi/2, 0)$. $h(\lambda)$ has exactly one root (MLE exists) if and only if $-n\pi/2 \le h(x) < 0$ by the *intermediate value theorem* (see [19], p. 93). □

Consider the case $n = 1$, we have:

$$P(-\pi/2 \le \arctan(X) < 0) = \frac{1}{2\sinh(\lambda\pi/2)(\exp(-n\lambda\pi/2) - 1)} > 0, \ \forall\lambda \in (0, \infty)$$

i.e., there exits a set with positive probability which MLE does not exist. Therefore, we introduce an alternative estimator of λ based on quantiles. The following measure of skewness is introduced in [8]:

$$\beta_p = \frac{(Q_{1-p} - Q_{0.5}) - (Q_{0.5} - Q_p)}{Q_{1-p} - Q_p}$$

see also [12]. Reference [9] has shown that β_p satisfies four properties introduced by Van Zwet [22] for a proper skewness index. It is obvious that β_p always exists and robust with respect to outliers. For robustness of β_p we use *breakdown value*. The breakdown value of an estimator is the fraction of that can be given arbitrary values without making the estimator arbitrary bad, the idea attributable to [11]. For instance, the breakdown value of median is equal to 0.5 which means that 50% of sample can be changed arbitrary while the median stays with no change, this quantity is equal to 0 for mean. It can be shown that breakdown point of β_p is equal to p.

We find the quantiles of VSC in (4). Consider $\hat{Q}_p$, the p-th sample quantile computed from $X_{(\lfloor np \rfloor)}$, where n is the sample size, $\lfloor np \rfloor$ is the integer part of np and $X_{(k)}$ is the k-th order statistic of the random sample $X_1,\ldots,X_n$.

Thus, sample β_p computed by $\hat{\beta}_p = \frac{(\hat{Q}_{1-p} - \hat{Q}_{0.5}) - (\hat{Q}_{0.5} - \hat{Q}_p)}{\hat{Q}_{1-p} - \hat{Q}_p}$. An estimator of λ can be computed by solving $\beta_p(\lambda) = \hat{\beta}_p(X_1,\ldots,X_n)$.

A difficulty in using the quantile-based estimator is to choose the appropriate p amount. Now we introduce a similar estimator based on mode of the distribution which does not depend on any unknown parameter.

Reference [2] introduced a measure of skewness with respect to mode. That is:

$$\gamma = 1 - 2F_X(M)$$

where $F_X(.)$ is the cdf of a continuous unimodal random variable and M is the unique mode of this distribution. They also have shown that γ satisfies all properties of an proper skewness index.

We find the unique mode and cdf of VSC in Theorem 1 and Equation 2. Therefore,

$$\gamma = 1 - 2\left\{\frac{1}{2\sinh(\lambda/2)}\left(e^{\lambda\pi/2} - e^{-\lambda\arctan(-\lambda/2)}\right)\right\}$$

The mode-based estimator is the solution of $\gamma(\lambda) = \hat{\gamma}(x_1,\ldots,x_n)$, where $\hat{\gamma} = 1 - 2F_X(\hat{M})$ and $\hat{M}$ is the sample mode.

The main difficulty of three previous estimators is their numerically computation. The rest of this section dedicates to some estimators which can be calculated analytically.

We find that $-\sum_{i=1}^{n} \arctan(X_i)$ is a complete sufficient statistic for λ. By an informal way, a complete sufficient statistic for the parameter λ is a function of a random sample from this distribution which consists of all of the sample information about λ, i.e., it knows all the sample information about λ and all of its information is about λ. Therefore, it is desirable to find an estimator of λ based on such a statistic.

Remark 1 *Rao-Blackwell theorem justifies that a function of complete sufficient statistic which is unbiased for λ is uniformly minimum variance unbiased estimator (UMVUE) of λ.*

We introduce $T = T(X_1,\ldots,X_n) = -\sum_{i=1}^{n} \arctan(x_i)/n$ as an estimator of λ which is a one-to-one function of the complete sufficient statistic of λ. Some properties of this estimator will be studied.

Remark 2 *When $n \to \infty$, $Var(T) \to 0$.*

The proof of following lemma is straightforward.

Lemma 2 $g(x) = \coth(x) - \frac{1-x^2}{x}, x > 0$ *is a strictly* *increasing function.*

Theorem 5 $bias(T,\lambda)\in(0,\infty)$.

Proof. Using Lemma 2 and Hospital's theorem the proof is straightforward. □

Remark 3 *T always overestimates λ.*

Now we introduce a Bayesian estimator of λ. For the sake of computational simplicity we introduce the following prior distribution:

$$\pi(\lambda)\propto\frac{2\sinh(\lambda\pi/2)}{\lambda};\ \lambda>0$$

the corresponding posterior is attained as follows,

$$\pi(\lambda\mid x_1,\dots,x_n)\propto e^{-\lambda\sum_{i=1}^{n}\arctan(x_i)}$$

therefore,

$$\lambda\mid x_1,\dots,x_n\sim EXP(1/\textstyle\sum_{i=1}^{n}\arctan(x_i))$$

where $EXP(\varsigma)$ indicates the exponential distribution with paremeter ς. The Bayes estimator with respect to quadratic error loss is

$$E(\lambda\mid y)=\frac{1}{\sum_{i=1}^{n}\arctan(x_i)}.$$

The proof of following proposition is straightforward.

Proposition 1 *If $X\sim EXP(\lambda)$ and $E(X)=\lambda$, then the $p-th$ quantile of X, q_p, is as follows:*

$$q_p=-\lambda\ln(1-p)$$

By Proposition 1, the Bayes estimator with respect to absolute error loss is

$$median(\lambda\mid x)=-ln0.5\frac{1}{\sum_{i=1}^{n}\arctan(x_i)}.$$

$$\approx 0.69\frac{1}{\sum_{i=1}^{n}\arctan(x_i)}$$

In Section 2, we show that VSC pdf belongs to exponential family and $T=-\arctan(X)$ is a sufficient statistic for λ. It can be shown that this family has monotone likelihood ratio, MLR, property with respect to T. By using Karlin-Rubin's theorem one can find the uniformly most powerful, UMP, level (size) α test for testing $H_0:\lambda<\lambda_0$ vs $H_1:\lambda\ge\lambda_0$. That is,

$$\varphi(X)=\begin{cases}1 & T\le t_0\\ 0 & T>t_0\end{cases}$$

is an UMP size α test, where $P(T\ge t_0)=\alpha$. It can be shown that:

$$F_T(t)=\frac{1}{2\sinh(\lambda\pi/2)}(e^{\lambda\pi/2}-e^{-\lambda t}),$$

$$t_0=\frac{1}{\lambda_0}\ln\left(e^{\lambda_0\pi/2}-2(1-\alpha)\sinh(\lambda_0\pi/2)\right)$$

For testing $H_0:F=F_0$, where F_0 is a known VSC cdf, against $H_1:F\ne F_0$, one can use *nonparametric methods* such as tests based on chi-square test or empirical distribution, e.g., Kolmogrov-Smirnov, KS, test. Since we have the closed form of the cdf, performing these tests is straightforward. Also we can use a quantile-quantile plot for this purpose.

4. COMPARISONS

In this section we compare VSC (very skewed Cauchy), skew Cauchy and skewed Cauchy distributions. We note that VSC is a negative skewed distribution (despite its positive valued skewness parameter).

Now we define a skew Cauchy random variable, which is well known as stable random variable with index 1, or 1-*stable* random variable, e.g., [20].

Definition 2 *A random variable X is skew Cauchy (in standard form) if and only if X has the following characteristic function:*

$$\phi_X(t)=\exp\left\{-|t|(1+i\beta\frac{2}{\pi}\operatorname{sgn}(t)\ln|t|)\right\},$$

where β is the skewness parameter, $-1\le\beta\le1$, and $\operatorname{sgn}(u)=-1$, 0, or 1 if u $<$, $=$, or $>$ 0, respectively.

By [7], skewed Cauchy random variable defined as follows:

Definition 3 *A random variable X has skewed Cauchy distribution with parameter λ, if its pdf has the following form:*

$$f_X(x)=\frac{1}{\pi(1+x^2)}(1+\frac{\lambda x}{\sqrt{1+(1+\lambda^2)x^2}})$$

$$=\frac{1}{\pi(1+x^2)}(1+sgn(\beta')\frac{x\sqrt{\frac{\beta'^2}{1-\beta'^2}}}{\sqrt{1+(1+\frac{\beta'^2}{1-\beta'^2})x^2}})$$

where the last equation is skewed Cauchy pdf with $\beta'=\frac{\lambda}{\sqrt{1+\lambda^2}}$ reparametrization.

As we mentioned before, in the skew Cauchy distribution the skeweness parameter, β, belongs to $[-1,1]$. Therefore, we need a proper reparametrization for

VSC to enable us comparing it by skew Cauchy distribution. We introduce $\beta^* = \frac{-\lambda}{\sqrt{1+\lambda^2}}$ reparametrization. It can be shown that, $\beta^* \in [-1,0]$, $\beta^* \to -1 \Leftrightarrow \lambda \to \infty$, and $\beta^* \to 0 \Leftrightarrow \lambda \to 0$.

The pdf in (1) with respect to the parameter β^* is as follows.

$$f_X(x) = \frac{\sqrt{\frac{\beta^{\Sigma^2}}{1-\beta^{\Sigma^2}}}}{2\sinh\left(\frac{\pi}{2}\sqrt{\frac{\beta^{\Sigma^2}}{1-\beta^{\Sigma^2}}}\right)} \frac{1}{1+x^2} e^{-\sqrt{\frac{\beta^{\Sigma^2}}{1-\beta^{\Sigma^2}}}\arctan(x)}$$

Now, we are able to compare very skewed Cauchy, skewed Cauchy and skew Cauchy via their pdfs and cdfs in Figures 1 and 2.

By Figures 1 and 2 one can find out VSC has heavier tail than two other distributions.

In order to compare goodness of fit of VSC and skew Cauchy distribution on skew heavy tail data, the following procedure is introduced. We use abbreviations SC for skewed Cauchy and SSC for skew Cauchy (first "S" refers to stable).

1. Generate 50 random numbers from VSC, $\lambda = 1,10,20$; SC, $\lambda = -1,-10,-20$, and SSC, $\lambda = -0.7,-0.995,-0.999$.

2. Calculate MLE of skewness parameters of VSC and SSC for each 3 data sets in step 1.

3. Calculate Kolmogrov-Smirnov (KS) goodness of fit test statistic for 2×3 different permutations of models and data sets for each parameter.

4. Repeat steps $1-3$, 10000 times.

5. Calculate means and standard deviations (S.D.) of KS statistics.

Table 1 shows simulation results based on previous 5 steps. As one can find out by Table 1, fitting the VSC model generally leads to smaller KS statistics. On the other hand, the results show that VSC dominates SSC in 3 different types of data sets for each parameter. That is, the means of KS statistics for fitted VSC distribution is less than the means of KS statistics for fitted SSC distribution. Figure 3 may show this matter better in a visual way. Obviously, just where the sample generated from SSC the performance of VSC and SSC are close to each other and in any other cases VSC is clearly better. In particular, when the skewness becomes greater; i.e., λ becomes greater, the performance of VSC model is much better than the SSC which shows that VSC is a much suitable model for heavy-tailed data.

The method of finding MLE for skewness parameter of skew Cauchy is described in [17, p. 379-400]. Note that since the MLE(λ) of skewed Cauchy distribution does not exist in many cases, and there is no estimator as well as MLE, we ignore comparison its goodness of fit with VSC and SSC. However, we keep the samples generated from SC in order to compare goodness of fit of fitting VSC and SSC models on a third different heavy tail sample.

At last, the similarities of VSC, SC, and SSC are classified in Table 2 ("●" shows property of the distributions).

5. Conclusions and Remarks

In this paper, a new class of heavy tailed distribution was introduced which form an exponential family. Also, its properties were studied and made inference on its skewness parameter. We showed that this distribution dominates $1-$stable distribution for fitting on skew heavy tail data. The similarities of VSC, SC, and SSC are classified.

Finally, we should note that, VSC pdf can be extended to location and scale distribution family, to reach more flexibility for modeling different data sets. VSC pdf with location, scale and skewness parameters is given by:

$$f_X(x) = \frac{\lambda}{2\sinh(\frac{\lambda\pi}{2})}\frac{1}{\sigma}$$

$$\frac{1}{1+(\frac{x-\mu}{\sigma})^2} e^{-\lambda\arctan(\frac{x-\mu}{\sigma})}, \quad x,\mu \in \mathbb{R}; \lambda,\sigma > 0$$

For future research we will investigate to inference on parameters of the location-scale VSC family and its applications.

6. Acknowledgment

The authors gratefully acknowledge the anonymous referees for their useful and technical comments and suggestions.

7. References

[1] B. C. Arnold and R. J. Beaver, "The skew-Cauchy distribution", Statist. Probab. Lett. 49 (3), 285-290. 2000.

[2] B. C. Arnold and R. A. Groeneveld, "Measuring skewness with respect to the mode", Amer. Statist. 49 (1), 34-38. 1995.

[3] A. Azzalini, "A class of distributions include the normal ones", Scand. J. Statist. 12, 171-178. 1985.

[4] A. Azzalini, "Further results on a class of distributions which includes the normal ones", Statistica XLVI, 199-208. 1986.

[5] A. Azzalini and A. Capitanio, "Statistical applications of the multivariate skew normal distribution", J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (3), 579--602. 1999.

[6] A. Azzalini and A. Dalla Valle, "The multivariate skew-normal distribution", Biometrika 83 (4), 715-726. 1996.

[7] J. Behboodian, A. Jamalizadeh and N. Balakrishnan, "A new class of skew-Cauchy distributions", Statist. Probab. Lett. 76 (14), 1488-1493. 2006.
[8] F. N. David and N. L. Johnson, "Some tests of significance with ordered variables", J. Roy. Statist. Soc. Ser. B. 18, 1-20; discussion, 20--31. 1956.
[9] R. A. Groeneveld and G. Meeden, "Some tests of significance with ordered variables", The Statistician 33, 391-399. 1984.
[10] R. C. Gupta and R. D. Gupta, "Generalized skew normal model", Test 13 (2), 501-524. 2004.
[11] F. R. Hampel, "The influence curve and its role in robust estimation", J. Amer. Statist. Assoc. 69, 383-393. 1974.
[12] D. V. Hinkley, "On power transformations to symmetry", Biometrika 62, 101-111. 1975.
[13] W. J. Huang and Y. H. Chen, "Generalized skew Cauchy distribution", Statist. Probab. Lett. 77, 1137-1147. 2007.
[14] E. L. Lehmann and G. Casella, Theory of point estimation, 2nd Edition. Springer-Verlag, New York. 1998.
[15] N. Loperfido, "Quadratic forms of skew-normal random vectors", Statist. Probab. Lett. 54 (3), 381-387. 2001.
[16] S. Nadarajah and S. Kotz, "Skew distributions generated from different families", Acta Appl. Math. 91 (1), 1-37. 2006.
[17] J. P. Nolan, "Maximum likelihood estimation and diagnostics for stable distributions", In: Lévy processes. Birkhäuser Boston, Boston, MA, pp. 379-400. 2001.
[18] A. Pewsey, "Problems of inference for Azzalini's skew-normal distribution", J. Appl. Stat. 27, 351-360. 2000.
[19] W. Rudin, Principles of mathematical analysis, 3rd Edition. McGraw-Hill Book Co., New York. 1976.
[20] G. Samorodnitsky and M. S. Taqqu, Stable non-Gaussian random processes. Chapman & Hall, New York. 1994.
[21] M. Sharafi and J. Behboodian, "The Balakrishnan skew–normal density", Statistical Papers. Online: http://www.springerlink.com/content/kn381u131gvpw337/. 2007.
[22] W. R. van Zwet, "Convex transformations of random variables", Vol. 7 of Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam. 1964.

Table 1: Mean and S.D. of KS statistic of fitted VSC and SSC distribution to the simulated data from VSC, SC, and SSC distribution for different parameters.

Distribution		KS Statistic		
		VSC(1)	SC(-1)	SSC(-0.7)
VSC	Mean (S.D.)	0.086 (0.024)	0.091 (0.025)	0.085 (0.022)
SSC	Mean (S.D.)	0.273 (0.054)	0.221 (0.055)	0.187 (0.058)
Distribution		VSC(10)	SC(-10)	SSC(-0.995)
VSC	Mean (S.D.)	0.090 (0.027)	0.159 (0.030)	0.105 (0.0234)
SSC	Mean (S.D.)	0.770 (0.035)	0.357 (0.022)	0.120 (0.038)
Distribution		VSC(20)	SC(-20)	SSC(-0.999)
VSC	Mean (S.D.)	0.094 (0.029)	0.163 (0.030)	0.105 (0.023)
SSC	Mean (S.D.)	0.865 (0.026)	0.367 (0.017)	0.120 (0.038)

Table 2: Similarities between VSC, SC, and SSC

Property	VSC	SC	SSC
Form an exponential family	●	○	○
Has more heavier tail than two other distributions	●	○	○
MLE does not always exist	●	●	○
Has close form of pdf, cdf and quantiles	●	●	○
When $\lambda \to 0$, converges to symmetric Cauchy distribution	●	●	●
$E\|X\|^r$ does not exist for $r \geq 1$	●	●	●
Is unimodal	●	●	●

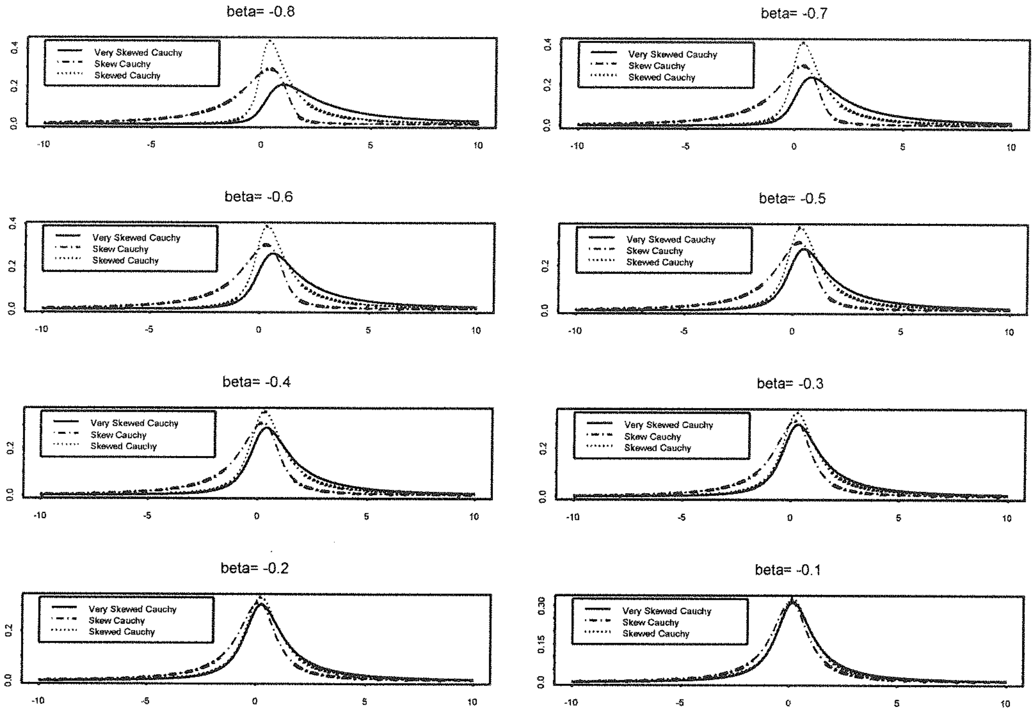

Figure 1: Comparing VSC (very skewed Cauchy), skew Cauchy, and skewed Cauchy pdfs

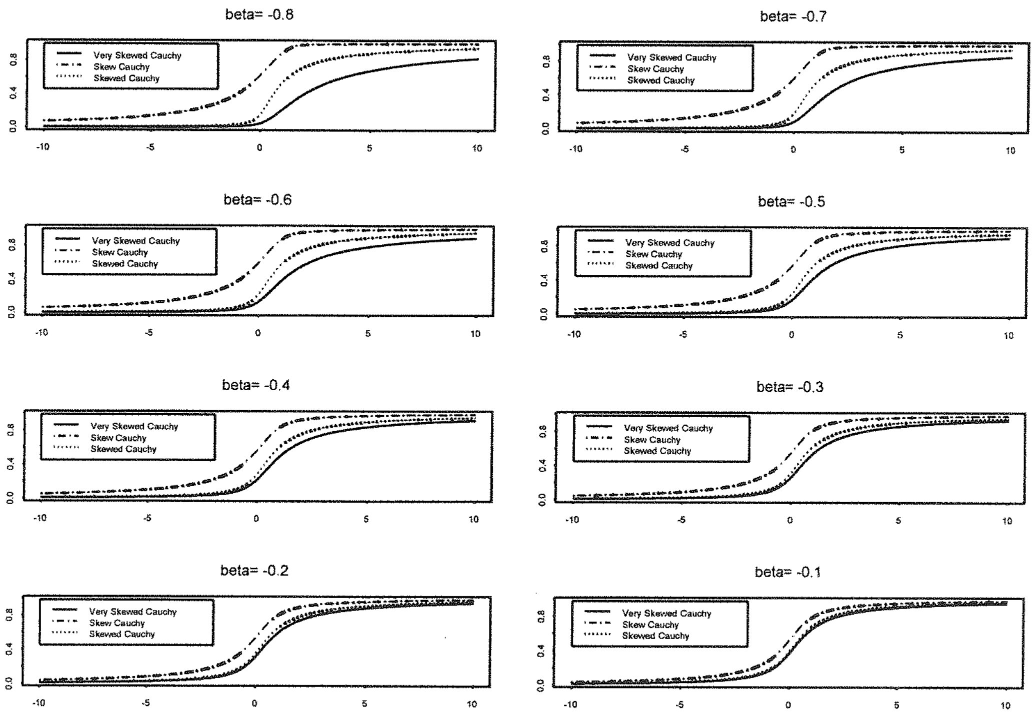

Figure 2: Comparing VSC (very skewed Cauchy), skew Cauchy, and skewed Cauchy cdfs

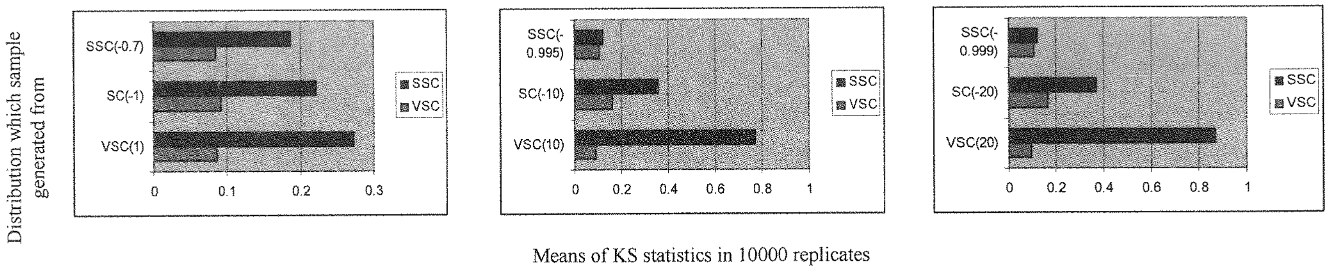

Figure 3: Comparing goodness of fit of VSC and SSC models