Making the Cauchy work

M. Shobana
Assistant Professor
M.sc, M.phil (maths)
Malla Reddy Engineering College for Women(autonomous)

Abstract- *A truncated version of the Cauchy distribution is introduced. Unlike the Cauchy distribution, this possesses finite moments of all orders and could therefore be a better model for certain practical situations. More than 10 practical situations where the truncated distribution could be applied are discussed. Explicit expressions are derived for the moments, L moments, mean deviations, moment generating function, characteristic function, convolution properties, Bonferroni curve, Lorenz curve, entropies, order statistics and the asymptotic distribution of the extreme order statistics. Estimation procedures are detailed by the method of moments and the method of maximum likelihood and expressions derived for the associated Fisher information matrix. Simulation issues are discussed. Finally, an application is illustrated for consumer price indices from the six major economics.*

I. INTRODUCTION

The Cauchy distribution given by the probability density function (pdf):

$$f(x) = \frac{1}{\pi\theta}\left\{1+\left(\frac{x-\mu}{\theta}\right)^2\right\}^{-1} \qquad (1.1)$$

(for $-\infty < x < \infty$, $\theta > 0$ and $-\infty < \mu < \infty$)

has been studied in the mathematical world for over three centuries. An excellent historical account of the distribution has been prepared by Stigler (1974). As he points out, (1.1) seems to have appeared first in the works of Pierre de Fermat in the mid-17th century and was subsequently studied by many including Sir Issac Newton, Gottfried Leibniz, Christian Huygens, Guido Grandi, and Maria Agnesi. The parameters μ and θ are the location and scale parameters, respectively. The distribution is symmetrical about $x = \mu$. The median is μ; the upper and lower quartiles are $\mu \pm \theta$; and, the points of inflexion are at $\mu \pm \theta/\sqrt{3}$. The main weakness of (1.1) is that it has no moments. In this paper, we overcome this weakness by introducing a truncated version. It has the pdf and cumulative distribution function (cdf) specified by

$$f(x; A, B) = \frac{1}{\theta D}\left\{1+\left(\frac{x-\mu}{\theta}\right)^2\right\}^{-1} \qquad (1.2)$$

and

$$F(x; A, B) = \frac{1}{D}\left\{\arctan\left(\frac{x-\mu}{\theta}\right) - \arctan\left(\frac{A-\mu}{\theta}\right)\right\}, \qquad (1.3)$$

respectively, for $-\infty \le A \le x \le B \le \infty$, $-\infty < \mu < \infty$ and $\theta > 0$, where

$$D(A, B) = \arctan(\beta) - \arctan(\alpha), \qquad (1.4)$$

where $\alpha = (A - \mu)/\theta$ and $\beta = (B - \mu)/\theta$. This distribution originally appeared in Johnson and Kotz (1970) and Rohatgi (1976) in simpler forms. Johnson and Kotz (1970) derived the variance and discussed estimation issues for the symmetric standard case given by $A = -B$, $\mu = 0$ and $\theta = 1$. Rohatgi (1976) derived expressions for the first two moments for the standard case given by $\mu = 0$ and $\theta = 1$. Note that (1.2) is unimodal. The mode is at $x = \mu$ if $A \le \mu \le B$. If $B < \mu$ then the mode is at $x = B$. If $\mu < A$ then the mode is at $x = A$.

Because (1.2) is defined over a finite interval, the truncated Cauchy distribution has all its moments. So, (1.2) may prove to be a better model for certain practical situations than one based on just the Cauchy distribution. Below, we discuss more than 10 such situations.

The Cauchy distribution given by (1.1) has been applied in the past as models for depth map data, prices of speculative assets such as stock returns and the phase derivative (random frequency of a narrow-band mobile channel) of air components in an urban environment. For data of this kind, there is no reason to believe that empirical moments of any order should be infinite. So, the choice of the Cauchy distribution as a model is unrealistic since none of its moments are finite. The alternative truncated Cauchy distribution given by (1.2) will be a more appropriate model for the kind of data mentioned. The choice of the limits, A and B, could be easily based on historical records.

A main problem with characterizing employment productivity distributions is to find a reasonable measure of the minimal and maximal productivity. In both ends of the distribution one is likely to find accumulations of measurement errors due either to downright faulty data or time aggregation problems associated with, for example, plant closures and new plants. With respect to Swedish employment data, Forslund and Lindh (2004) took average wage costs as the measure of minimal sustainable productivity and 95th percentile productivity as a fairly reliable indicator of maximal sustainable productivity. Forslund and Lindh (2004) foundthat the empirical employment distribution between these two productivity values was well described by a truncated Cauchy distribution.

Consider the truncated Cauchy distribution in (1.2) for A = 0, B =∞, μ = 0 and θ = 1. So, we have $f(x) = (2/\pi)(1 + x2)$
−1 for $x > 0$ with the asymptotic tail $f(x) \sim (2/\pi)x$
−2 as $x \to \infty$. Yablonsky (1985) established that $(2/\pi)x$ −2 coincides with the classical *Lotka distribution of scientific productivity* (describing the frequency of publication by authors in any given field) up to a normalizing constant. So, Lotka's law is an approximate expression of the asymptotic form of the truncated Cauchy distribution.

Truncated Cauchy distributions are also popular priors for Bayesian models especially with respect to economic data; see, for example, Bauwens et al. (1999). The aim of this paper is to provide a comprehensive account of the mathematical

$$\psi(x) = \frac{d \log \Gamma(x)}{dx},$$

the exponential integral defined by

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{\exp(t)}{t} dt$$

and the Gauss hypergeometric function defined by

$$ {}_2F_1(a, b; c; x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!},$$

II. MOMENTS

Here, we discuss moments of a random variable X having the pdf (1.2). Some of the results given have been reported earlier by Nadarajah and Kotz (2006). They are repr

$$E(X^n; A, B) = \frac{1}{\theta D} \int_A^B x^n \left\{ 1 + \left(\frac{x-\mu}{\theta} \right)^2 \right\}^{-1} dx. \quad (2.1)$$

Setting $y = (x-\mu)/\theta$ and using the binomial expansion

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k,$$

one can rewrite (2.1) as

$$\begin{aligned} E(X^n; A, B) &= \frac{1}{D} \int_\alpha^\beta \frac{(\mu + \theta y)^n}{1+y^2} dy \\ &= \frac{1}{D} \sum_{k=0}^{n} \binom{n}{k} \mu^{n-k} \theta^k \int_\alpha^\beta \frac{y^k}{1+y^2} dy \quad (2.2) \\ &= \frac{1}{D} \sum_{k=0}^{n} \binom{n}{k} \mu^{n-k} \theta^k \{ I_k(\beta) - I_k(\alpha) \}, \end{aligned}$$

where

$$I_k(c) = \int_0^c \frac{y^k}{1+y^2} dy.$$

By equation (3.194.5) in Gradshteyn and Ryzhik (2000), one can calculate $I_k(c)$ as

$$I_k(c) = \frac{c^{k+1}}{k+1} {}_2F_1\left(1, \frac{k+1}{2}; \frac{k+3}{2}; -c^2\right). \quad (2.3)$$

By combining (2.2) and (2.3) it follows that the nth moment of X is given by

$$\begin{aligned} E(X^n; A, B) &= \frac{\mu^n}{D} \sum_{k=0}^{n} \frac{1}{k+1} \binom{n}{k} \left(\frac{\theta}{\mu}\right)^k \left\{ \beta^{k+1} {}_2F_1\left(1, \frac{k+1}{2}; 1 + \frac{k+1}{2}; -\beta^2\right) \right. \quad (2.4) \\ &\quad \left. - \alpha^{k+1} {}_2F_1\left(1, \frac{k+1}{2}; 1 + \frac{k+1}{2}; -\alpha^2\right) \right\} \end{aligned}$$

for $n \geq 1$. In the standard case $\mu = 0$ and $\theta = 1$, using standard properties of the Gauss hypergeometric function, one can obtain the first four moments of X from (2.4) as:

$$E(X; A, B) = \{\log(1+B^2) - \log(1+A^2)\}/(2D), \quad (2.5)$$

$$E(X^2; A, B) = \{\arctan(A) - \arctan(B) - A + B\}/D, \quad (2.6)$$

$$E(X^3; A, B) = \{\log(1+A^2) - \log(1+B^2) - A^2 + B^2\}/(2D) \quad (2.7)$$

and

$$E(X^4; A, B) = \{3\arctan(B) - 3\arctan(A) - A^3 + B^3 + 3A - 3B\}/(3D). \quad (2.8)$$

III. L MOMENTS

L-moments are summary statistics for probability distributions and data samples (Hoskings, 1990). They are analogous to ordinary moments but are computed from linear combinations of the ordered data values. The *n*th *L* moment is defined by

$$\lambda_n = \sum_{j=0}^{n-1} (-1)^{n-1-j} \binom{n-1}{j} \binom{n-1+j}{j} \beta_j, \qquad (3.1)$$

where

$$\beta_r = \int x \{F(x)\}^r f(x)\, dx.$$

The *L* moments have several advantages over ordinary moments: for example, they apply for any distribution having finite mean; no higher-order moments need be finite.

Suppose *X* is a truncated Cauchy random variable with its pdf specified by (1.2).

Assume without loss of generality that $\mu = 0$ and $\theta = 1$. Then the *n*th *L* moment of *X* is given by (3.1), where

$$\beta_r = D^{-r-1} \sum_{i=0}^{r} \binom{r}{i} (-\arctan(A))^{r-i} \int_A^B x (\arctan(x))^i (1+x^2)^{-1} dx. \quad (3.2)$$

Using the series expansion

$$\arctan(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2k+1}, \qquad (3.3)$$

one can calculate (3.2) as

$$\beta_r = D^{-r} \sum_{i=0}^{r} \binom{r}{i} (-\arctan(A))^{r-i} \times \sum_{k_1=0}^{\infty} \cdots \sum_{k_i=0}^{\infty} \frac{(-1)^{k_1+\cdots+k_i}}{(2k_1+1)\cdots(2k_i+1)} E\left(X^{2(k_1+\cdots+k_i)+i+1}; A, B\right).$$

IV. CONVOLUTION

If *X*1 and *X*2 are independent Cauchy random variables then it is well known that their convolution, *X*1 + *X*2, is also a Cauchy random variable. It is natural to ask whether this property holds for truncated Cauchy random variables. Suppose *X*1 and *X*2 are independent random variables with the pdf (1.2) for $(\mu, \theta, A, B) = (\mu 1, \theta 1, A1, B1)$ and $(\mu, \theta, A, B) = (\mu 2, \theta 2, A2, B2)$, respectively. Let *D*1 and *D*2 denote that corresponding normalizing constants given by (1.4). Let $S = X1 + X2$.

Then the pdf of *S* can be written as

$$\begin{aligned} f_S(s) &= \frac{\theta_1\theta_2}{D_1 D_2} \int \frac{dx}{\{(x-\mu_1)^2+\theta_1^2\}\{(s-x-\mu_2)^2+\theta_2^2\}} \\ &= \frac{1}{4D_1D_2} \int \left\{ \frac{\alpha_1-\alpha_3}{x-\mu_1+i\theta_1} + \frac{\alpha_4-\alpha_2}{x-\mu_1-i\theta_1} \right. \\ &\quad \left. + \frac{\alpha_2-\alpha_1}{x+\mu_2-s-i\theta_2} \frac{\alpha_3-\alpha_4}{x+\mu_2-s+i\theta_2} \right\} dx \end{aligned} \qquad (6.1)$$

by partial fractions, where $i = \sqrt{-1}$, $\alpha_1 = (\mu_2 - s - i\theta_2 + \mu_1 - i\theta_1)^{-1}$, $\alpha_2 = (\mu_2 - s - i\theta_2 + \mu_1 + i\theta_1)^{-1}$, $\alpha_3 = (\mu_2 - s + i\theta_2 + \mu_1 - i\theta_1)^{-1}$ and $\alpha_4 = (\mu_2 - s + i\theta_2 + \mu_1 + i\theta_1)^{-1}$. If $B_1 + A_2 \le A_1 + B_2$ then (6.1) can be reduced to

$$f_S(s) = \begin{cases} \frac{1}{4D_1D_2}\Big\{(\alpha_1-\alpha_3)\log\frac{s-A_2-\mu_1+i\theta_1}{A_1-\mu_1+i\theta_1} + (\alpha_4-\alpha_2)\log\frac{s-A_2-\mu_1-i\theta_1}{A_1-\mu_1-i\theta_1} \\ \quad +(\alpha_2-\alpha_1)\log\frac{s-A_2+\mu_2-s-i\theta_2}{A_1+\mu_2-s-i\theta_2} \\ \quad +(\alpha_3-\alpha_4)\log\frac{s-A_2+\mu_2-s+i\theta_2}{A_1+\mu_2-s+i\theta_2}\Big\}, \\ \quad \text{if } A_1+A_2 \le s \le B_1+A_2, \\ \frac{1}{4D_1D_2}\Big\{(\alpha_1-\alpha_3)\log\frac{B_1-\mu_1+i\theta_1}{A_1-\mu_1+i\theta_1} + (\alpha_4-\alpha_2)\log\frac{B_1-\mu_1-i\theta_1}{A_1-\mu_1-i\theta_1} \\ \quad +(\alpha_2-\alpha_1)\log\frac{B_1+\mu_2-s-i\theta_2}{A_1+\mu_2-s-i\theta_2} + (\alpha_3-\alpha_4)\log\frac{B_1+\mu_2-s+i\theta_2}{A_1+\mu_2-s+i\theta_2}\Big\}, \\ \quad \text{if } B_1+A_2 \le s \le A_1+B_2, \\ \frac{1}{4D_1D_2}\Big\{(\alpha_1-\alpha_3)\log\frac{B_1-\mu_1+i\theta_1}{s-B_2-\mu_1+i\theta_1} + (\alpha_4-\alpha_2)\log\frac{B_1-\mu_1-i\theta_1}{s-B_2-\mu_1-i\theta_1} \\ \quad +(\alpha_2-\alpha_1)\log\frac{B_1+\mu_2-s-i\theta_2}{s-B_2+\mu_2-s-i\theta_2} \\ \quad +(\alpha_3-\alpha_4)\log\frac{B_1+\mu_2-s+i\theta_2}{s-B_2+\mu_2-s+i\theta_2}\Big\}, \\ \quad \text{if } A_1+B_2 \le s \le B_1+B_2. \end{cases}$$

V. ESTIMATION

Here, we consider estimation by the method of moments and the method of maximum likelihood and provide expressions for the associated Fisher information matrix. Suppose *X*1, . . . , *Xn* is a random sample from (1.2). By equating the first four moments of (2.4) with the corresponding sample estimates, one can obtain the method of moments estimators as the simultaneous solutions of the four equations

$$\frac{\mu^m}{D}\sum_{k=0}^{m}\frac{1}{k+1}\binom{m}{k}\left(\frac{\theta}{\mu}\right)^k\left\{\beta^{k+1}{}_2F_1\left(1,\frac{k+1}{2};1+\frac{k+1}{2};-\beta^2\right)\right.$$

$$\left.-\alpha^{k+1}{}_2F_1\left(1,\frac{k+1}{2};1+\frac{k+1}{2};-\alpha^2\right)\right\}=\frac{1}{n}\sum_{i=1}^{n}X_i^m$$

for $m = 1, 2, 3, 4$, where $D = \arctan(\beta) - \arctan(\alpha)$, $\alpha = (A - \mu)/\theta$ and $\beta = (B - \mu)/\theta$.

Now, consider the method of maximum likelihood. The log-likelihood for the random sample is

$$\log L(\mu,\theta,A,B) = -n\log C - \sum_{i=1}^{n}\log\left\{1+\left(\frac{X_i-\mu}{\theta}\right)^2\right\}, \qquad (11.1)$$

where $C = \theta D$. The first-order derivatives of (11.1) with respect to the four parameters are:

$$\frac{\partial \log L}{\partial A} = -\frac{n}{C}\frac{\partial C}{\partial A},$$

$$\frac{\partial \log L}{\partial B} = -\frac{n}{C}\frac{\partial C}{\partial B},$$

$$\frac{\partial \log L}{\partial \mu} = -\frac{n}{C}\frac{\partial C}{\partial \mu} + \frac{2}{\theta^2}\sum_{i=1}^{n}(X_i-\mu)\left\{1+\left(\frac{X_i-\mu}{\theta}\right)^2\right\}^{-1}$$

and

$$\frac{\partial \log L}{\partial \theta} = -\frac{n}{C}\frac{\partial C}{\partial \theta} + \frac{2}{\theta^3}\sum_{i=1}^{n}(X_i-\mu)^2\left\{1+\left(\frac{X_i-\mu}{\theta}\right)^2\right\}^{-1}.$$

REFERENCES

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.

[2] Bauwens, L., Lubrano, M. and Richard, J. F. (1999). *Bayesian Inference in Dynamic Econometric Models*. New York: Oxford Univ. Press. MR1858176

[3] Bhowmick, K., Mukhopadhyay, A. and Mitra, G. B. (2000). Edgeworth series expansion of the truncated Cauchy function and its effectiveness in the study of atomic heterogeneity. *Zeitschrift Fur Kristallographie* 215, 718–726.

[4] Bonferroni, C. E. (1930). *Elementi di Statistica Generale*. Firenze: Seeber.

[5] Borgia, G. C., Brown, R. J. S. and Fantazzini, P. (1996). The effect of diffusion and susceptibility differences on T-2 measurements for fluids in porous media and biological tissues. *Magnetic Resonance Imaging* 14, 731–736.