The Annals of Statistics
1996, Vol. 24, No. 2, 787–808

MÖBIUS TRANSFORMATION AND CAUCHY PARAMETER ESTIMATION

BY PETER MCCULLAGH

University of Chicago

Some properties of the ordinary two-parameter Cauchy family, the circular or wrapped Cauchy family, and their connection via Möbius transformation are discussed. A key simplification is achieved by taking the parameter $\theta = \mu + i\sigma$ to be a point in the complex plane rather than the real plane. Maximum likelihood estimation is studied in some detail. It is shown that the density of any equivariant estimator is harmonic on the upper half-plane. In consequence, the maximum likelihood estimator is unbiased for $n \geq 3$, and every harmonic or analytic function of the maximum likelihood estimator is unbiased if its expectation is finite. The joint density of the maximum likelihood estimator is obtained in exact closed form for samples of size $n \leq 4$, and in approximate form for $n \geq 5$. Various marginal distributions, including that of Student's pivotal ratio, are also obtained. Most results obtained in the context of the real Cauchy family also apply to the wrapped Cauchy family by Möbius transformation.

1. Cauchy distributions and Möbius groups.

1.1. *Parameterization.* Let Y be a Cauchy random variable with the usual parameterization in which μ represents the median or centre of symmetry, and σ is the scale parameter or probable error. The parameter space is then taken to be the upper half-plane with $\sigma > 0$. For almost all purposes, however, it is advantageous to take $\theta = \mu + i\sigma$ as a point in the complex plane rather than the real plane. Rather than restrict the parameter space to the upper half of the complex plane, it is generally more convenient to use the whole plane and to identify conjugate pairs of points. This parameter space, which we denote by U or Θ is isomorphic to the upper half-plane. For $\theta_2 \neq 0$, θ and $\bar{\theta}$ give rise to the same Cauchy density

$$f(y;\theta) = \frac{|\theta_2|}{\pi|y-\theta|^2}. \tag{1}$$

It is helpful here to think of the sample space as the real axis embedded in the parameter space. For $\theta_2 = 0$ the Cauchy distribution is a point mass at θ_1.

Received November 1992; revised March 1995.
AMS 1991 *subject classifications*. Primary 62A05; secondary 62E15.
Key words and phrases. Bartlett adjustment, Brownian motion, circular Cauchy distribution, complex parameter, equivariance, fractional linear transformation, harmonic measure, invariant measure, invariant statistic, likelihood ratio statistic, Möbius group, robustness, wrapped Cauchy distribution.

For notational simplicity we write $Y \sim C(\theta)$, with a single complex argument, meaning that Y has the Cauchy distribution with median $\Re(\theta)$ and probable error $|\Im(\theta)|$.

1.2. *Real Möbius group*. The standard Cauchy distribution ($\theta = \pm i$) is obtained as the ratio of two independent standard normal variables. As a consequence, if X is a bivariate normal random variable with zero mean and covariance matrix Σ, then $Y = X_1/X_2$ has the Cauchy distribution with median $\theta_1 = \sigma_{12}/\sigma_{22}$ and probable error $\theta_2 = |\Sigma|^{1/2}/\sigma_{22}$. If we make a real nonsingular linear transformation from X to X',

$$\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} = \begin{pmatrix} aX_1 + bX_2 \\ cX_1 + dX_2 \end{pmatrix}$$

with matrix A, it is evident that X' is also bivariate normal. Consequently, $Y' = X_1'/X_2'$ must also be Cauchy. The relation between Y and Y' is

$$Y' = \frac{aY + b}{cY + d}, \tag{2}$$

which is called a fractional linear transformation. The set of such transformations with real coefficients and nonzero determinant forms a group known variously as the real fractional linear group, the real Möbius group, the projective group or SL(2, R).

It follows from the preceding construction that the Cauchy family is closed under the real Möbius group. The advantage of using a complex parameter space is evident in the expression for the induced transformation on the parameter space:

$$\frac{aY + b}{cY + d} \sim C\left(\frac{a\theta + b}{c\theta + d}\right). \tag{3}$$

Note that even if $\Im(\theta) > 0$, the transformed parameter may have negative imaginary part. It is only by using the entire complex plane and identifying conjugate pairs that the induced transformation can be presented in the simple form (3).

The real Möbius group is generated by composing three types of transformation, namely, location shifts, scale multiples and reciprocals:

$$Y \mapsto Y + b, \qquad Y \mapsto aY \quad \text{and} \quad Y \mapsto 1/Y.$$

Numerous families of distributions are closed under location–scale transformation, but the Cauchy family is the only univariate location–scale family that is also closed under reciprocals [Knight (1976), Knight and Meyer (1976)]. For that reason any results given here are unlikely to extend beyond the two-parameter Cauchy family.

The standard Cauchy distribution has parameter $\theta = \pm i$. From this one distribution the entire family can be generated from the standard by applying location–scale transformations alone,

$$Y = \theta_1 + \theta_2 \varepsilon,$$

where $\varepsilon \sim C(i)$. Apart from the sign of θ_2, this is the unique transformation in the location–scale group that generates $C(\theta)$ from $C(i)$. Within the real Möbius group, however, there is an entire one-dimensional family all of which transform $C(i)$ to $C(\theta)$. To see this, we observe that if $\varepsilon \sim C(i)$,

$$\varepsilon' = \pm \frac{a\varepsilon + b}{-b\varepsilon + a} \sim C\left(\pm \frac{ai + b}{-bi + a} \right) = C(\pm i) \tag{4}$$

is also standard Cauchy provided that a and b are real and not both zero. The set of transformations (4), under which the parameter $\{i, -i\}$ is invariant, is known as the stabilizer of $\{i, -i\}$ or the isotropy group of $\{i, -i\}$. We denote this group by I. The isotropy group associated with an arbitrary point $\{\theta, \bar{\theta}\}$ is the set of transformations

$$Y \mapsto \theta_1 \pm \theta_2 g\left(\frac{Y - \theta_1}{|\theta_2|} \right),$$

where $g \in I$.

1.3. *Circular Cauchy distribution.* Let Z be a random variable on the unit circle in the complex plane given by

$$Z = \frac{1 + iY}{1 - iY},$$

where $Y \sim C(\theta)$ is real-valued. This is a 1–1 transformation from the extended real line to the unit circle. It is an elementary exercise to show that the density of Z with respect to arc length on the circle is

$$f(z; \psi) = \frac{|1 - |\psi|^2|}{2\pi|z - \psi|^2}, \tag{5}$$

where $\psi = (1 + i\theta)/(1 - i\theta)$. We write $Z \sim C^*(\psi)$ to denote the Cauchy distribution (5) on the unit circle.

If $\Im(\theta) > 0$, then $|\psi| < 1$, in which case the absolute value sign can be omitted from the numerator in (5). However, the image of $\bar{\theta}$ is $1/\bar{\psi}$, a point outside the unit circle. Thus, the parameter space for the circular Cauchy is the extended complex plane in which pairs of points $\{\psi, 1/\bar{\psi}\}$ are identified. Note that, in (5), $f(z; 1/\bar{\psi}) = f(z; \psi)$ as required.

The uniform distribution is a special case given by the parameter point $\{0, \infty\}$.

Although it appears to bear scant resemblance to the real Cauchy distribution, the circular Cauchy does possess many of the properties of the real Cauchy distribution. In particular, the family (5) is closed under the action of the Möbius group on the unit circle. To say the same thing in another way, the transformations

$$Z \mapsto \exp(i\alpha)Z \quad \text{and} \quad Z \mapsto \frac{Z - \gamma}{\bar{\gamma}Z - 1}, \tag{6}$$

with $|\gamma| \neq 1$, map the unit circle onto itself. If $Z \sim C^*(\psi)$, the induced distributions are

$$C^*(\exp(i\alpha)\psi) \quad \text{and} \quad C^*\left(\frac{\psi - \gamma}{\bar{\gamma}\psi - 1}\right),$$

respectively. The first of these transformations is a rotation: for $|\gamma| < 1$ the second is a straight-line projection through the point γ onto the opposite side of the circle.

Note in particular that the uniform distribution is unaffected by rotation: the set of rotations is the isotropy group for $\psi = \{0, \infty\}$. If Z is uniform, projection through ψ with $|\psi| < 1$ produces a Cauchy distribution with parameter ψ: if $Z \sim C^*(\psi)$, projection through ψ gives a uniform distribution. The isotropy group for ψ is obtained by (i) projecting through ψ, (ii) rotating through an arbitrary angle and (iii) projecting through ψ.

1.4. *Harmonicity.* Suppose $u \equiv u(x, y)$ is a continuous real function on the closed unit disc satisfying

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

for all $z = x + iy$ in the open disc. Then u is said to be harmonic on the open unit disc. If Z has the circular Cauchy distribution (4) with $|\psi| < 1$, then $E(u(Z)) = u(\psi)$. The proof follows from Rudin [(1987), Theorem 11.9].

Since the real and imaginary parts of a holomorphic (analytic) function are harmonic, it follows that $E(g(Z)) = g(\psi)$ provided that $g(\cdot)$ is continuous on the closed unit disc and analytic on the open unit disc. In particular, the complex moments of the circular Cauchy distribution (4) are

$$E(Z^k) = \psi^k, \tag{7}$$

for integer $k = 0, 1, \dots$.

For any measure on the unit circle, the complex moments are also the Fourier coefficients. These are sufficient to identify the distribution uniquely. Consequently, property (7) uniquely identifies the circular Cauchy among all measures on the unit circle.

1.5. *Wrapped Cauchy distribution.* Let $W = \exp(iY)$, where Y is real. Then $|W| = 1$, and W is said to be the wrapped version of Y on the unit circle. The terminology is taken from Mardia (1972), although the wrapping there is done entirely in terms of real variables. If Y has the real Cauchy distribution (1), its characteristic function is

$$E(\exp(itY)) = \exp(it\theta),$$

for all $t \geq 0$ provided that $\Im(\theta) \geq 0$. It follows then that $E(W) = \exp(i\theta) = \psi$, say, and

$$E(W^k) = \psi^k,$$

for positive integer k. These are the moments of the circular Cauchy distribution, from which it follows that the wrapped Cauchy family coincides with the circular Cauchy family.

1.6. *Convolutions*. In the context of Cauchy distributions there are two natural convolution operations depending on whether we are dealing with the real-valued Cauchy or the circular Cauchy distribution. For independent real-valued Cauchy variables

$$Y_1 + Y_2 \sim C(\theta_1 + \theta_2), \tag{8}$$

provided that $\Im(\theta_1)$ and $\Im(\theta_2)$ are both nonnegative. This result follows from the characteristic function

$$\begin{aligned} E\big(\exp(it(Y_1 + Y_2))\big) &= E\big(\exp(itY_1)\big)E\big(\exp(itY_2)\big) \\ &= \exp(it\theta_1)\exp(it\theta_2) = \exp\big(it(\theta_1 + \theta_2)\big), \end{aligned}$$

for $t \geq 0$.

For independent circular Cauchy variables we have the alternative convolution operation

$$Z_1 Z_2 \sim C^*(\psi_1 \psi_2), \tag{9}$$

provided that $|\psi_1| \leq 1$ and $|\psi_2| \leq 1$. The latter result follows simply from moment calculations:

$$E\big((Z_1 Z_2)^k\big) = E\big(Z_1^k\big)E\big(Z_2^k\big) = \psi_1^k \psi_2^k = (\psi_1 \psi_2)^k,$$

for nonnegative integer k.

In a certain sense, the proofs and statements of (8) and (9) are identical. The proof of (8) can be interpreted as showing that the product of two wrapped Cauchy variables is also a wrapped Cauchy variable. The proof of (9) says exactly the same for circular Cauchy variables. However, the circular and wrapped Cauchy families are identical.

It is convenient in what follows to assume that the parameters satisfy $\Im(\theta_j) \geq 0$ and $|\psi_j| \leq 1$. Also, $Y_j \sim C(\theta_j)$ are independent. The image of Y_j on the unit circle is $Z_j = (1 + iY_j)/(1 - iY_j)$, which has the circular Cauchy distribution with parameter ψ_j.

On the real line, the image of (9) is given by

$$Y_1 \oplus Y_2 = \frac{Y_1 + Y_2}{1 - Y_1 Y_2} \sim C\left(\frac{\theta_1 + \theta_2}{1 - \theta_1 \theta_2}\right). \tag{10}$$

It can be shown by induction that the "circular sum" of $Y_1, \dots, Y_n$, obtained by repeated application of (10), is

$$Y_1 \oplus \cdots \oplus Y_n = \frac{\Sigma(-1)^r S_{2r+1}(Y)}{\Sigma(-1)^r S_{2r}(Y)} \sim C\left(\frac{\Sigma(-1)^r S_{2r+1}(\theta)}{\Sigma(-1)^r S_{2r}(\theta)}\right),$$

where, for $0 \leq r \leq n$, $S_r(Y)$ is the reduced monomial symmetric function in Y of degree r:

$$S_r(y) = \sum y_{i_1} y_{i_2} \cdots y_{i_r},$$

with summation over $\binom{n}{r}$ terms having unequal subscripts.

For the special case in which the Y's are identically distributed, (9) gives $\prod Z_j \sim C^*(\psi^n)$. If $|\psi| < 1$, the limiting distribution is uniform on the circle. The image on the real line is

$$Y_1 \oplus \cdots \oplus Y_n = \frac{\Sigma(-1)^r S_{2r+1}(Y)}{\Sigma(-1)^r S_{2r}(Y)} \sim C\left(\frac{\Sigma(-1)^r \binom{n}{2r+1}\theta^{2r+1}}{\Sigma(-1)^r \binom{n}{2r}\theta^{2r}}\right) \to C(i), \tag{11}$$

in which the limiting distribution is standard Cauchy. In other words, the parameter ratio on the right of (11) tends to $\pm i$ provided only that $\Im(\theta) \neq 0$. In fact (11) can be viewed as a kind of central limit theorem for independent and identically distributed random variables, not necessarily Cauchy. This central limit theorem asserts that $Y_1 \oplus \cdots \oplus Y_n$ has a standard Cauchy limit provided that the distribution of Y_1 has a continuous component. To see why this is so, we observe that the circular sum is the image on the real line of the product $Z_1 \cdots Z_n$ on the unit circle, which is in turn determined by the sum of the arguments modulo 2π. If the distribution of Z_1 has a continuous component, the product is asymptotically uniform on the unit circle [Wold (1934), Kolassa and McCullagh (1990), Section 3]. Hence the circular sum $Y_1 \oplus \cdots \oplus Y_n$ on the real line is asymptotically standard Cauchy.

1.7. *Connection with Brownian motion.* The results given in preceding sections are most easily understood in connection with Brownian motion in the complex plane. A complete description of the connection between Brownian motion and harmonic analysis can be found in Rogers and Williams (1986), but the following brief description will suffice for present purposes.

Let $B(t)$ be the position at time t of a Brownian particle in the complex plane starting at $B(0) = \psi$ with $|\psi| < 1$. Such a particle will eventually leave the unit disc. Since Brownian paths are continuous, the time of first exit T is the smallest t for which $|B(t)| = 1$. The point of first exit is a random variable $Z \equiv B(T)$ whose distribution is $C^*(\psi)$. It is a property of Brownian motion, sometimes known as the martingale property, that $E(u(Z)) = u(\psi)$ provided that $u(\cdot)$ is harmonic in the unit disc.

Lévy's theorem states that Brownian paths are preserved by analytic transformation. In other words, if $w(z)$ transforms the unit disc to the unit disc, then $w(B(t))$ is a Brownian path in which the particle moves with variable instantaneous "velocity" given by $|dw/dz|$ [see Rogers and Williams (1986), Sections 4.33 and 4.34]. This transformed particle starts at $\psi^* = w(\psi)$ and exits at $Z^* = w(Z)$. A consequence of Lévy's theorem is that the transformed exit point is again circular Cauchy, $Z^* \sim C^*(\psi^*)$. A specific example is the function

$$w(z) = \exp\left(\frac{z-1}{z+1}\right),$$

which is analytic on the open unit disc, but not 1–1. Again, the transformed exit distributions is $w(Z) \sim C^*(w(\psi))$.

The real Cauchy distribution is the exit distribution of Brownian motion from the upper half-plane, or lower half-plane if $\theta_2 < 0$. The Cauchy transformation property (3) is a consequence of Lévy's theorem in which $w(\cdot)$ is fractional linear.

2. Maximum likelihood estimation. In this section $y_1, \dots, y_n$ are the observed values of independent and identically distributed random variables $Y_1, \dots, Y_n$, where $Y_i \sim C(\theta)$. It is known from the work of Copas (1975) that the likelihood function is unimodal for $n \geq 3$. In this section, various nonnumerical aspects of the problem of parameter estimation are considered, in particular the implications of equivariance under the real Möbius group. The distribution of the maximum likelihood estimator is derived in closed form for $n \leq 4$ and studied in some detail.

2.1. *Closed-form maximum likelihood estimate.* We consider first the case $n = 2$ with observations $y_1 = -1$ and $y_2 = 1$. Then the likelihood function is

$$L_2(\theta) = \frac{\theta_2^2}{\left(1 + 2\theta_1 + |\theta|^2\right)\left(1 - 2\theta_1 + |\theta|^2\right)}.$$

The likelihood contours at level $L_2(\theta) = 1/4r^2$, with $r \geq 1$, satisfy the equation

$$\left(\theta_1^2 + (\theta_2 - \rho)^2 - r^2\right)\left(\theta_1^2 + (\theta_2 + \rho)^2 - r^2\right) = 0,$$

which identifies a pair of circles of radius r, one with center at $i\rho = i\sqrt{r^2 - 1}$, the other with center at $-i\rho$. All such circles pass through the sample points -1 and 1. The contour of maximum likelihood is the unit circle, or more generally, the circle having (y_1, y_2) as diameter.

For $n = 4$ it is convenient to order the sample points $y_{(1)} < y_{(2)} < y_{(3)} < y_{(4)}$. The likelihood function can be written as the product

$$L_4(\theta; y) = L_2\left(\theta; y_{(1)}, y_{(3)}\right) L_2\left(\theta; y_{(2)}, y_{(4)}\right)$$

in which the first factor is maximized on the circle with diameter $(y_{(1)}, y_{(3)})$, and the second factor is maximized on the circle with diameter $(y_{(2)}, y_{(4)})$. The overall maximum thus occurs at the intersection of these circles and is given by

$$\hat{\theta}_1 = \frac{y_{(2)} y_{(4)} - y_{(1)} y_{(3)}}{y_{(4)} - y_{(3)} + y_{(2)} - y_{(1)}},$$

$$\hat{\theta}_2 = \frac{\sqrt{(y_{(4)} - y_{(3)})(y_{(3)} - y_{(2)})(y_{(2)} - y_{(1)})(y_{(4)} - y_{(1)})}}{y_{(4)} - y_{(3)} + y_{(2)} - y_{(1)}}.$$

This expression is correct for any cyclic permutation of the ordered sample values.

For $n = 3$ a different argument is required. First, if the observed values are $(-1, 0, 1)$, the maximum likelihood estimate is $i/\sqrt{3}$. For any configuration of three distinct values (y_1, y_2, y_3) there exists a real Möbius transformation matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

that transforms the points $(-1, 0, 1)$ to (y_1, y_2, y_3). The required coefficients are given by

$$a = y_1 y_2 + y_2 y_3 - 2 y_1 y_3, \qquad b = y_2(y_3 - y_1),$$
$$c = 2y_2 - y_1 - y_3, \qquad d = y_3 - y_1.$$

Since the maximum likelihood estimator is equivariant, it follows that

$$\hat{\theta}(y) = A \circ \frac{i}{\sqrt{3}} = \frac{ai/\sqrt{3} + b}{ci/\sqrt{3} + d}.$$

Simplification of the real and imaginary parts gives

$$\hat{\theta}_1 = \frac{y_3(y_1 - y_2)^2 + y_1(y_2 - y_3)^2 + y_2(y_3 - y_1)^2}{(y_1 - y_2)^2 + (y_2 - y_3)^2 + (y_3 - y_1)^2},$$
$$\hat{\theta}_2 = \sqrt{3}\,\frac{(y_1 - y_2)(y_2 - y_3)(y_3 - y_1)}{(y_1 - y_2)^2 + (y_2 - y_3)^2 + (y_3 - y_1)^2}.$$

Ferguson (1978) gives the same formulae without derivation and shows that they satisfy the likelihood equations.

2.2. *Robustness in small samples*? Given that closed-form estimates are available for $n = 3$ and $n = 4$, one can examine in some detail how the maximum likelihood estimate is affected by perturbations of arbitrary size in one or more observations. We consider here the case $n = 4$ with y_1, y_2, y_3 held fixed. It follows from the geometrical construction described in the preceding section that $\hat{\theta}$ must lie on one of the semicircles with diameters (y_1, y_2), (y_1, y_3) and (y_2, y_3). Further, $\hat{\theta}$ is a continuous function of y. Consequently, as y_4 varies from $-\infty$ to ∞, $\hat{\theta}$ moves continuously on these semicircular arcs. This locus is shown in detail in Figure 1. It can be seen that the scale component $\hat{\theta}_2$ is particularly sensitive to near-coincidences in the observations. Also, as y_4 increases beyond $\max(y_1, y_2, y_3)$, $\hat{\theta}_1$ decreases to a limiting value, namely, the median of (y_1, y_2, y_3). This limiting value is not the same as the estimate with y_4 omitted.

Similar calculations for $n = 5$ show that the locus is a smooth closed curve, but the qualitative phenomenon of sensitivity to clustering in the data points remains apparent.

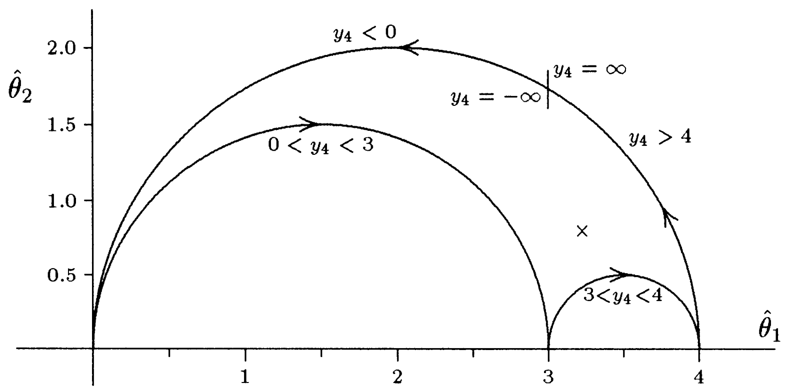

FIG. 1. *Locus of the maximum likelihood estimator for* $-\infty < y_4 < \infty$ *with* $(y_1, y_2, y_3) = (0, 3, 4)$. *The maximum likelihood estimator with* y_4 *omitted is marked by* $\times$ *at* $(3.23, 0.80)$.

2.3. *Equivariance.* The term equivariance refers to the behaviour of an estimator T of θ under the action of a group $\mathscr{G}$ on the sample space. It is assumed that the family of distributions is closed under the action of the group, so that, for each $g \in \mathscr{G}$, Y and gY belong to the same family with parameters θ and $g\theta$, respectively. In the present context, the group acts componentwise on $y \in R^n$ as in (2): its action on the parameter space is given by (3). To estimate $\theta' = g\theta$, one might first transform the data to $y' = gy$ and then compute $T(y')$. Alternatively, one could compute $T(y)$ and then transform to $gT(y)$. There is no guarantee, even for consistent estimators, that these two procedures yield the same numerical value. If, however,

$$T(gy) = gT(y), \tag{12}$$

for all $g \in \mathscr{G}$, the estimator $T(\cdot)$ is said to be equivariant under $\mathscr{G}$. Typical estimators based on the sample median and sample probable error, or semiinterquartile range, are equivariant under location–scale transformation, but not under the larger Möbius group. The maximum likelihood estimator, however, is automatically equivariant because the likelihood function based on gy is the same as the likelihood function based on y [Eaton (1989), Section 3.3].

From the definition, an equivariant estimator T of θ satisfies the property

$$\mathrm{pr}(T \in S; \theta) = \mathrm{pr}(T \in gS; g\theta),$$

for all $g \in \mathscr{G}$ and for all measurable $S \subset U$. Consequently, the density, if it exists, must satisfy

$$\mathrm{pr}(T \in dt; \theta) = p(t; \theta)\mu(dt),$$

where $\mu(\cdot)$ is invariant measure under the action of $\mathscr{G}$ on U, and $p(gt; g\theta) = p(t; \theta)$ is invariant under the action of $\mathscr{G}$ on $U \times \Theta$. In the case of the Möbius group acting on $U \times \Theta$ via

$$gt = \frac{at + b}{ct + d}, \qquad g\theta = \frac{a\theta + b}{c\theta + d}, \tag{13}$$

$\chi(t; \theta) = |t - \theta|^2/4t_2\theta_2$ is a maximal invariant, and $\mu(dt) = dt_1\, dt_2/t_2^2$ is invariant measure, unique up to a multiplicative constant. In other symbols, $\chi(gt; g\theta) = \chi(t; \theta)$ and

$$\int_S \frac{dt_1\, dt_2}{t_2^2} = \int_{gS} \frac{dt_1\, dt_2}{t_2^2}.$$

Consequently, any equivariant estimator of θ in the Cauchy problem must have a density of the form

$$\operatorname{pr}(T \in dt) = \frac{p_n(\chi)}{4\pi t_2^2}\, dt_1\, dt_2, \tag{14}$$

for some function $p_n(\cdot)$ on the positive real line. The reason for inserting the constant 4π is that $p_n(\cdot)$ is then the marginal density of the pivotal statistic $|T - \theta|^2/4T_2\theta_2$.

In many problems involving group action, the action of the group on the parameter space is isomorphic to the action of the group on itself, either left action or right action. The Cauchy problem is not an instance of this because the parameter space is not isomorphic to the group. For that reason, the distinction between left- and right-invariant measures does not arise. The invariant measure given above is a measure on the parameter space, not Haar measure on the group.

2.4. *The orbit of a nonequivariant estimator*. Apart from the maximum likelihood estimator, most other estimators of θ are not equivariant under $\mathscr{G}$, although they are typically equivariant under location–scale transformation. It is of interest, therefore, to examine the set of possible estimators that could be obtained by first transforming y componentwise to gy, and then correcting by applying g^{-1} to the estimate. In symbols, if $T(\cdot)$ is not equivariant, its orbit is defined as the set

$$\{g^{-1}T(gy) \colon g \in \mathscr{G}\}.$$

If $T(y)$ is consistent for θ, so also is $g^{-1}T(gy)$.

The estimator used here is the sample median and the semiinterquartile range defined as follows. With the observations indexed from 0 to $n - 1$ and arranged in increasing order, we define

$$i_1 = \left[\frac{n-2}{4}\right]_-, \qquad i_4 = \left[\frac{n-2}{4}\right]_+, \qquad i_2 = i_1 + 1, \qquad i_3 = i_4 - 1;$$

$$w = \frac{n-2}{4} - i_1; \qquad T_2 = \frac{(1-w)(y_{i_4} - y_{i_1}) + w(y_{i_3} - y_{i_2})}{2},$$

where $[x]_-$ is the integer part of x rounding down, and $[x]_+$ is the smallest integer not less than x, that is, rounding up.

Figure 2 exhibits the orbits for two data configurations with $n = 5$ and $n = 6$. The remarkable observation is that the orbit is a smooth curve in Θ for $n \leq 5$, but for $n \geq 6$ the orbit is the union of n arc segments. This discontinuity at $n = 6$ seems to be characteristic of estimators based on order statistics. Estimators that are not equivariant under sign reversal have more complicated orbits.

If we regard the real part of $g^{-1}T(gy)$ as a generalized median, Figure 2a exhibits a paradox, in that this generalized median can occur slightly beyond the range of the observations. Also, the generalized probable error can be arbitrarily close to zero for $n \leq 5$.

2.5. *Distribution of the maximum likelihood estimator.* Equation (14) specifies the distribution of an equivariant estimator of θ up to an undetermined function $p_n(\chi)$ on the positive real line. In this section we focus on the

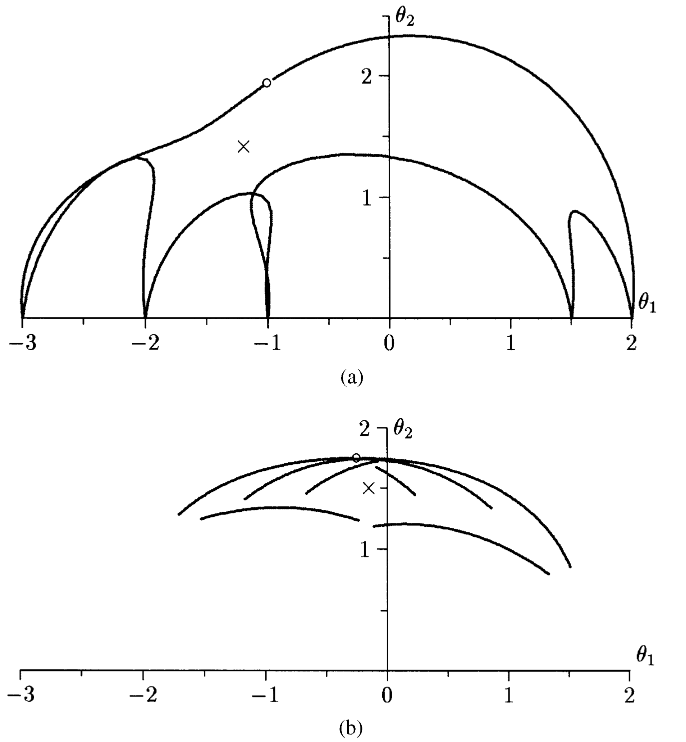

FIG. 2. (a) *Orbit of a nonequivariant estimator* $g^{-1}T(gy)$ *for* $n = 5$; $y = (-3.0, -2.0, -1.0, 1.5, 2.0)$: *the value of* $T(y)$ *is marked by* $\circ$; *the mle is marked by* $\times$. (b) *Orbit of a nonequivariant estimator* $g^{-1}T(gy)$ *for* $n = 6$; $y = (-3.0, -2.0, -1.0, 0.5, 1.5, 2.0)$: *the value of* $T(y)$ *is marked by* $\circ$; *the mle is marked by* $\times$.

maximum likelihood estimator and obtain expressions for $p_3(\chi)$, $p_4(\chi)$ and the asymptotic form as $n \to \infty$ of $p_n(\chi)$.

The strategy is to make a nonsingular transformation from $(y_1, \dots, y_n)$ to (t, a), in which $t = (t_1, t_2)$ and $a_j = (y_j - t_1)/t_2$ are the components of the configuration statistic. Maximum likelihood estimation imposes constraints on the configuration such that (a_{n-1}, a_n) can be expressed in terms of $(a_1, \dots, a_{n-2})$. The marginal distribution of T can then be obtained by integrating out the auxiliary $(n-2)$-dimensional variable a. The transformation $y_j = t_1 + t_2 a_j$ has a Jacobian of the form

$$\frac{\partial(y)}{\partial(t, a)} = t_2^{n-2} J_n(a),$$

in which $J_n(a)$ depends on the expression for (a_{n-1}, a_n) in terms of $(a_1, \dots, a_{n-2})$. Specifically, $J_n(a)$ is the absolute value of the determinant of

$$\begin{pmatrix} 1 - \sum \dfrac{\partial a_{n-1}}{\partial a_j} & a_{n-1} - \sum a_j \dfrac{\partial a_{n-1}}{\partial a_j} \\ 1 - \sum \dfrac{\partial a_n}{\partial a_j} & a_n - \sum a_j \dfrac{\partial a_n}{\partial a_j} \end{pmatrix}, \tag{15}$$

in which the sums run from $j = 1$ to $n - 2$.

In the case of the Cauchy distribution, the configuration takes on a particularly simple form when expressed in terms of $z_j = (1 + ia_j)/(1 - ia_j)$ on the unit circle. The likelihood equation is $\sum z_j = 0$ [McCullagh (1992)]. Given values $z_1, \dots, z_{n-2}$ satisfying $|z_.| \le 1$, where

$$z_. = \frac{z_1 + \cdots + z_{n-2}}{2},$$

the remaining two points are $-z_.(1 \pm i(|z_.|^{-2} - 1)^{1/2})$ provided that $z_. \ne 0$.

For $n = 3$ the configuration is necessarily of the form

$$(z, \omega z, \omega^2 z) \quad \text{or} \quad (z, \omega^2 z, \omega z),$$

with $|z| = 1$ and $\omega = \exp(2\pi i/3)$. On the original scale this translates to

$$\left(a, \frac{a + \sqrt{3}}{1 - a\sqrt{3}}, \frac{a - \sqrt{3}}{1 + a\sqrt{3}}\right) \quad \text{or} \quad \left(a, \frac{a - \sqrt{3}}{1 + a\sqrt{3}}, \frac{a + \sqrt{3}}{1 - a\sqrt{3}}\right).$$

Equation (15) gives $J_3(a) = 6\sqrt{3}(1 + a^2)^2/|1 - 3a^2|^2 \times 2$, the doubling factor coming from the two distinct configurations. Likewise, for $n = 4$, the configuration is necessarily a permutation of $(z_1, z_2, -z_1, -z_2)$ or $(a_1, a_2, -1/a_1, -1/a_2)$. This gives

$$J_4(a) = \frac{2|(a_1 - a_2)(1 + a_1 a_2)|}{a_1^2 a_2^2} \times 3.$$

For $n = 3$ and $\theta = i$ the joint density of T is

$$\int_{-\infty}^{\infty} \frac{t_2 J_3(a)\, da}{\pi^3\{1 + (t_1 + t_2 a_1)^2\}\{1 + (t_1 + t_2 a_2)^2\}\{1 + (t_1 + t_2 a_3)^2\}},$$

which simplifies to $p_3(\chi)/(4\pi t_2^2)$, with

$$p_3(\chi) = \frac{3\sqrt{3}}{\pi(1 + 3\chi + 3\chi^2)}. \tag{16}$$

Likewise, for $n = 4$ and $\theta = i$, the density of T is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Big[\pi^4\{1 + (t_1 + t_2 a_1)^2\}\{1 + (t_1 + t_2 a_2)^2\}$$
$$\times \{1 + (t_1 - t_2/a_1)^2\}\{1 + (t_1 - t_2/a_2)^2\}\Big]^{-1} t_2^2 J_4(a)\, da_1\, da_2.$$

After much simplification this reduces to $p_4(\chi)/4\pi t_2^2$, with

$$p_4(\chi) = \frac{12\log(1 + 2\chi)}{\pi^2\chi(\chi + 1)(2\chi + 1)}. \tag{17}$$

The process of simplifying these integrals is exceedingly tedious and error-prone, but the preceding formulae can easily be verified numerically. This direct method seems not to be feasible for $n > 4$.

The large-sample behavior of $p_n(\chi)$ is difficult to compute, but the following argument gives an upper bound for the large-deviation rate, $-\lim_{n\to\infty} n^{-1} \log p_n(\chi)$. First compute the conditional density of T given the value of the configuration statistic a. From the Jacobian calculations given above, this conditional density is proportional to $t_2^{n-2}\prod f(t_1 + a_j t_2; \theta)$. Thus the conditional density with respect to invariant measure $(dt_1\, dt_2/t_2^2)$ is proportional to

$$\prod_j t_2 f(t_1 + a_j t_2; \theta).$$

The maximum of this density occurs at $t = \theta$. The determinant of the logarithmic second-derivative matrix at $t = \theta$ is $(n^2 - |\sum z_j^2|^2)/4$, which can be approximated by $n^2/4$. Laplace approximation then gives

$$f_{T|A}(t|a;\theta) \cong \frac{n}{4\pi t_2^2}\prod_j \frac{f(t_1 + a_j t_2;\theta)t_2}{f(\theta_1 + a_j\theta_2;\theta)\theta_2}$$
$$= \frac{n}{4\pi t_2^2}\prod_j \frac{t_2\theta_2(1 + a_j^2)}{\theta_2^2 + (t_1 - \theta_1 + a_j t_2)^2},$$

with relative error $O(n^{-1})$. Since $\hat{\theta}$ is $\sqrt{n}$-consistent, the components of a satisfy

$$a_j = \frac{(Y_j - \hat{\theta}_1)}{\hat{\theta}_2} = \varepsilon_j - \frac{\delta_j}{\sqrt{n}}$$

where $\varepsilon_j = (Y_j - \theta_1)/\theta_2$ are independent standard Cauchy and δ_j are $O_p(1)$ and exchangeable. In fact,

$$\frac{\delta_j}{\sqrt{n}} = y_j\left(\frac{1}{\theta_2} - \frac{1}{\hat{\theta}_2}\right) + \frac{\hat{\theta}_1}{\hat{\theta}_2} - \frac{\theta_1}{\theta_2},$$

so the average of the δ's is $O_p(1)$. For any differentiable function $h(\cdot)$, the mean-value theorem gives $h(a_j) = h(\varepsilon_j) + \delta_j h'(\xi_j)/n^{1/2}$, where ξ_j lies in the interval (a_j, ε_j). If $h(\varepsilon)$ has finite mean and h' is bounded, the law of large numbers gives

$$\sum \frac{h(a_j)}{n} = \sum \frac{h(\varepsilon_j)}{n} + O_p(n^{-1/2}) \to E(h(\varepsilon)).$$

From the appendix to McCullagh (1993), if ε is standard Cauchy,

$$E\log(1+\varepsilon^2) = \log 4,$$
$$E\log\left(\theta_2^2 + (t_1 - \theta_1 + \varepsilon t_2)^2\right) = \log(|t - \bar{\theta}|^2),$$

where $\Im(t)$ and $\Im(\theta)$ are both positive. Hence, taking

$$h(a) = \log(1+a^2) - \log\left(\theta_2^2 + (t_1 - \theta_1 + a t_2)^2\right),$$

we have

$$E\log\left(\frac{t_2\theta_2(1+a_j^2)}{\theta_2^2 + (t_1 - \theta_1 + a_j t_2)^2}\right) = \log\left(\frac{4t_2\theta_2}{|t - \bar{\theta}|^2}\right) + O(n^{-1/2}).$$

By the law of large numbers, therefore,

$$-\frac{1}{n}\log f_{T|A}(t|a;\theta) \to \log\left(\frac{4t_2\theta_2}{|t-\bar{\theta}|^2}\right) = \log(1+\chi).$$

This is the large-deviation rate for the conditional distributions: it is independent of the value of the conditioning statistic. For the unconditional large-deviation rate, Jensen's inequality gives

$$\begin{aligned}-\lim_{n\to\infty} n^{-1}\log f_T(t;\theta) &= -\lim_{n\to\infty} n^{-1}\log E f_{T|A}(T|a;\theta)\\ &\le -E\left\{\lim_{n\to\infty} n^{-1}\log f_{T|A}(t|a,\theta)\right\} = \log(1+\chi).\end{aligned}$$

If this inequality is strict, as it appears to be for positive χ, we have here a conundrum. On the one hand the conditional large-deviation rate does not depend on the value of the conditioning variable. On the other hand the unconditional large-deviation rate is smaller than the conditional rate. Even though it is agreed that the conditioning variable is not relevant for the computation of large-deviation probabilities, there is no agreement on which limit to use.

Application to (14) gives $\lim n^{-1} \log p_n(\chi) \geq -\log(1+\chi)$. Simulation results indicate that

$$p_n(\chi) \cong (n-2)(1+\chi)^{-n+1} \tag{18}$$

is reasonably accurate for samples as small as $n = 5$. This approximation achieves the conditional large-deviation rate. It is also consistent with the familiar moderate-deviation limit that $\sqrt{n/2}(\hat{\theta} - \theta)/\theta_2$ is bivariate standard normal and $n\chi$ is unit exponential. Exact results for $n \leq 4$ suggest that χ has finite moments up to but not including order $n - 2$. Approximation (18) achieves this property, but the claim has not been proved.

2.6. *Moments and other expectations*. Let $u(t)$ be any function that is harmonic on U. If T has a density of the form (14), the expected value of $u(T)$ is given by

$$E(u(T)) = \iint u(t) \frac{1}{4\pi t_2^2} p(\chi)\, dt_1\, dt_2.$$

The curve for χ equal to a constant is a complete circle in U with center $\omega = (\theta_1, \theta_2(1 + 2\chi))$ and radius $\rho = 2\theta_2\sqrt{\chi(1+\chi)}$. As a consequence, it is convenient to make a change of variables from (t_1, t_2) to (χ, ϕ) as follows:

$$t_1 = \theta_1 + 2\theta_2\sqrt{\chi(1+\chi)}\cos\phi,$$

$$t_2 = \theta_2(1+2\chi) + 2\theta_2\sqrt{\chi(1+\chi)}\sin\phi,$$

with Jacobian

$$\left|\frac{\partial(t_1, t_2)}{\partial(\chi, \phi)}\right| = 2t_2\theta_2.$$

This change of variables gives

$$E(u(T)) = \int_0^\infty p(\chi)\, d\chi \int_{-\pi}^{\pi} \frac{\theta_2 u(\omega + \rho\exp(i\phi))\, d\phi}{2\pi(\omega_2 + \rho\sin(\phi))}. \tag{19}$$

The standard form of the Poisson integral formula [Rudin (1987), Section 5.22],

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{(1-r^2)h(\exp(i\phi))\, d\phi}{1 + r^2 - 2r\cos(\phi - \alpha)} = h(r\exp(i\alpha)),$$

with $\alpha = -\pi/2$ and $r = \chi/\sqrt{\chi(\chi+1)}$, gives

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{h(\exp(i\phi))\, d\phi}{1 + 2\chi + 2\sqrt{\chi(1+\chi)}\sin\phi} = h\left(\frac{\chi\exp(i\alpha)}{\sqrt{\chi(1+\chi)}}\right)$$

if $h(\cdot)$ is harmonic on the open unit disc and continuous on the closed disc. Application of the formula in this form to (19) gives

$$E(u(T)) = \int_0^\infty p(\chi) u(\theta)\, d\chi = u(\theta), \tag{20}$$

provided that the integral exists and $u(\cdot)$ is harmonic on the upper half-plane.

Note in particular that the real and imaginary parts of an analytic function are harmonic. Consequently, if $g(\cdot)$ is analytic on U with $E|g(T)| < \infty$, it follows that $E(g(T)) = g(\theta)$. In particular, if $g(t) = t^k$, the integral (6) converges provided that $\int p(\chi)\chi^{k-1}\,d\chi < \infty$. In the case of the Cauchy maximum likelihood estimator, $E(|T|^k) < \infty$ for $k < n-1$ if our conjecture regarding the moments of $p_n(\chi)$ is correct.

2.7. *Marginal distributions.* It is of some interest to obtain the marginal distributions of T_1 and T_2, and of the pivotal statistic $(T_1 - \theta_1)/T_2$, analogous to the Student t-ratio used in normal-theory inference. For $n = 3$ and $\theta = (0, 1)$, integration using residues gives exact expressions as follows:

$$\text{(21a)} \quad \operatorname{pr}(T_1 \in dt_1) = \frac{3\,dt_1}{\pi(1 + t_1^2)\sqrt{4 + 3t_1^2}},$$

$$\text{(21b)} \quad \operatorname{pr}(T_2 \in dt_2) = \frac{2\sqrt{3}\,dt_2}{\pi(1 + t_2^2)^{3/2}} \frac{\sqrt{2}}{(1+\varepsilon)^{1/2}\left(1 + (1+\varepsilon)^{1/2}\right)^{1/2}},$$

$$\text{(21c)} \quad \operatorname{pr}\left(\frac{T_1}{T_2} \in dt\right) = \frac{3\sqrt{3}\,dt}{\pi^2\sqrt{4 + 3t^2}} \log\left|\frac{1 + \sqrt{4 + 3t^2}}{1 + \sqrt{4 + 3t^2}}\right|,$$

where $\varepsilon = 4t_2^2/(3(1 + t_2^2)^2)$ in (21b). Since $0 \le \varepsilon \le \frac{1}{3}$, the second factor in (21b) is bounded between 0.83 and 1.0. Thus, the density is approximately the positive half of a Student t on two degrees of freedom. Expression (21c) is a mixture of Student t-distributions on odd degrees of freedom in which the first few weights are $3/\pi$, $1/(8\pi)$, $9/(640\pi)$ and $15/(7168\pi)$. In other words, the density is essentially Cauchy with probable error 1.1, which is a little less than $2/\sqrt{3}$, the probable error of the leading term. These densities are not members of any of the standard univariate families, nor are they simple transformations of standard distributions.

For $n = 4$ the available formulae are a little more complicated:

$$\text{(22a)} \quad \operatorname{pr}(T_2 \in dt_2) = \frac{12\,dt_2}{\pi^2 t_2^{3/2}} \left(\frac{\log\left(v + 1 + \sqrt{v(v+2)}\right)}{\sqrt{v+2}} + \frac{\log\left(v - 1 + \sqrt{v(v-2)}\right)}{\sqrt{v-2}} - \frac{2\log 2v}{\sqrt{v}} \right),$$

$$\text{(22b)} \quad \operatorname{pr}\left(\frac{T_1}{T_2} \in dt\right) = \frac{12\sqrt{1+t^2}\,\log\left(|t| + \sqrt{1+t^2}\right) - 6|t|\log(4 + 4t^2)}{\pi^2|t|\sqrt{1+t^2}}\,dt,$$

where $v = t_2 + 1/t_2$. The mean of T_2 is 1, which is consistent with (20): the variance, determined numerically, is 2. The tail behaviour of the density is $36\log t_2/\pi^2 t_2^4$. The tail behaviour of the density of T_1/T_2 is $6\log t/\pi^2 t^3$: the mean absolute deviation is approximately 1.69. I have not succeeded in finding a closed-form expression for the marginal density of T_1, although its variance must be 2.

Resorting to approximation (18) for large n, and keeping $\theta = (0, 1)$ for simplicity, the marginal distribution of T_2 is given by

$$\frac{(n-1)T_2}{n-2} \sim F_{2(n-2),\,2(n-1)},$$

with unit expectation and variance $2/(n-3)$ for $n > 3$. Further, the conditional distribution of T_1 given T_2 is proportional to Student's t on $2n-3$ degrees of freedom:

$$T_1 | T_2 = t_2 \sim \frac{1 + t_2}{\sqrt{2n-3}} t_{2n-3}.$$

This has mean zero, conditional variance $(1 + t_2)^2/(2n-5)$ and unconditional variance $2/(n-3)$ for $n > 3$. Note that harmonicity requires $\mathrm{var}(T_1) = \mathrm{var}(T_2)$ and $\mathrm{cov}(T_1, T_2) = 0$ if these moments exist. The approximate marginal density of the pivot T_1/T_2, obtained from (18) by Laplace approximation, is

$$\text{(23)} \qquad \mathrm{pr}\left(\frac{T_1}{T_2} \in dt\right) \cong \frac{(\nu - 1)2^{3\nu}\Gamma^2(\nu)\, dt}{4\pi\sqrt{2}\,\Gamma(2\nu)(1 + t^2)^{1/4}\left(1 + \sqrt{1 + t^2}\right)^{\nu - 1/2}}.$$

where $\nu = n - 1$. This is a correction of, and an improvement over, a similar formula in McCullagh (1993). The tails of this distribution are $O(1/|t|^{\nu})$, that is, like those of t_{n-2}, in agreement with the exact distribution for $n = 3$. For large n, the centre of the distribution is such that the standardized ratio $\sqrt{n/2 - 1}\, T_1/T_2$ is approximately distributed as Student's t on $2(n-1)/3$ degrees of freedom.

2.8. *Information in the marginal and full likelihoods*. In general, if $T(\cdot)$, as an estimator of θ, is equivariant under location–scale transformation, the data may be partitioned into $y = (t, a)$ in which t is the value of the estimator and a is ancillary. The marginal likelihood based on T alone is thus

$$\text{(24)} \qquad L_M(\theta; t) = E_A L(\theta; t, a),$$

where $L(\theta; y) = \exp(l(\theta; y))$ is the full likelihood based on data y. The Cauchy log-likelihoods are

$$\text{(25)} \qquad l(\theta; y) = n \log \theta_2 - \sum \log|y_j - \theta|^2,$$

$$\text{(26)} \qquad l_M(\theta; t) = \log p_n(\chi).$$

The latter expression applies only if $T(\cdot)$ is equivariant under the real Möbius group, which is assumed in the remainder of this section.

The second term on the right-hand side of (25) is harmonic in θ. Consequently, the Laplacian of the log-likelihood satisfies

$$\nabla^2 l(\theta; y) = n\nabla^2 \log \theta_2 = -\frac{n}{\theta_2^2},$$

which is a constant independent of the data. Similar calculations show that the Laplacian of the marginal log-likelihood is given by

$$\begin{aligned}\nabla^2 l_M(\theta; t) &= \nabla^2 \log p_n(\chi) \\ &= g_n'(\chi)\left[\left(\frac{\partial\chi}{\partial\theta_1}\right)^2 + \left(\frac{\partial\chi}{\partial\theta_2}\right)^2\right] + g_n(\chi)\nabla^2\chi \\ &= \frac{\{\chi(1+\chi)g_n'(\chi) + (1+2\chi)g(\chi)\}}{\theta_2^2},\end{aligned}$$

where $g_n = p_n'/p_n$. Note that $-\nabla^2 l(\hat\theta; y) = n/\hat\theta_2^2$ is the trace of the observed information matrix and has a natural statistical interpretation in terms of precision of estimation. Likewise $-\nabla^2 l_M(t; t)$ is the trace of the marginal observed information matrix based on the statistic $T(y) = t$.

The Laplacian of the likelihood is given in terms of logarithmic derivatives by

$$\nabla^2 L = L\left[\left(\frac{\partial l}{\partial\theta_1}\right)^2 + \left(\frac{\partial l}{\partial\theta_2}\right)^2\right] + L\nabla^2 l.$$

At stationary points we have $\nabla^2 L = L\nabla^2 l$. Using (24), the Laplacian of the marginal likelihood is

$$\begin{aligned}\nabla^2 L_M(\theta; t) &= E_A \nabla^2 L(\theta; t, a) \\ &= E_A\left[L(\theta; t, a)\left\{\left(\frac{\partial l}{\partial\theta_1}\right)^2 + \left(\frac{\partial l}{\partial\theta_2}\right)^2\right\}\right] + E_A L(\theta; t, a)\left(-\frac{n}{\theta_2^2}\right) \\ &= E_A\left[L(\theta; t, a)\left\{\left(\frac{\partial l}{\partial\theta_1}\right)^2 + \left(\frac{\partial l}{\partial\theta_2}\right)^2\right\}\right] + L_M(\theta; t)\left(-\frac{n}{\theta_2^2}\right).\end{aligned}$$

On evaluating this expression at the marginal maximum likelihood estimate $\theta = t$ and dividing through by L_M, the trace of the marginal observed information matrix is obtained in the form

$$(27)\quad -\nabla^2 l_M(t; t) = \frac{n}{t_2^2} - E_A\Big[\frac{L(t; t, a)}{L_M(t; t)}\{(\frac{\partial l(t)}{\partial\theta_1})^2 + (\frac{\partial l(t)}{\partial\theta_2})^2\}\Big] \le \frac{n}{t_2^2}.$$

Equality is achieved if and only if t is a stationary point of both l and l_M.

In terms of the density $p_n(\chi)$ and for any equivariant estimator, this result implies

$$-g_n(0) = -\frac{\partial \log p_n(0)}{\partial \chi} \le n,$$

equality being achieved only by the maximum likelihood estimator. In other words, the maximum likelihood estimator achieves maximum local concentration as measured by the invariant $\chi(t;\theta)$. It should be borne in mind that, for large n, $\chi(T;\theta) = O_p(n^{-1})$. In large samples, it is the local behaviour of $p_n(\cdot)$ near the origin that dominates.

Note that the exact expressions (16) and (17) satisfy the condition that $-g_n(0) = n$, but approximation (18) gives $-g_n(0) = n - 1$.

2.9. *Marginal likelihood ratio statistic*. The marginal likelihood ratio statistic is $G = 2\log p_n(0) - 2\log p_n(\chi)$, a function of the invariant statistic. The exact distribution can be obtained for $n = 2$ and $n = 3$ using (16) and (17). The first three cumulants are 3.877, 15.676 and 126.690 for $n = 3$, and 3.217, 10.080 and 62.357 for $n = 4$. However, approximation (18) gives the very simple result that the rth cumulant of G is $2^r(r-1)!((n-1)/(n-2))^r$. Consequently, ignoring the error in (18), the Bartlett adjusted statistic $(n-2)G/(n-1)$ has rth cumulant equal to $2^r(r-1)!$, which is exactly the rth cumulant of χ_2^2 on two degrees of freedom. Note that, even for $n = 3$ and 4, the approximate cumulants based on (18) are reasonably accurate, and the Bartlett adjustment is quite effective.

3. Miscellaneous topics.

3.1. *Invariant tests*. Let $Y_1, \dots, Y_n$ be independent and identically distributed with distribution F. Suppose that it is required to test the null hypothesis that F is in the two-parameter Cauchy family $C(\cdot)$. The alternative hypothesis includes all continuous distributions on the real line. If $Y \sim C(\cdot)$, then $gY \sim C(\cdot)$ for all $g \in \mathscr{G}$. Likewise, the alternative hypothesis is preserved by $\mathscr{G}$. Consequently, a plausible argument can be made for treating Y and gY as equally consistent or equally inconsistent with the null hypothesis. This criterion of invariance leads to the selection of a test statistic that is constant on orbits, that is, $T(gy) = T(y)$ for all $g \in \mathscr{G}$ and for all $y \in R^n$.

The easiest way to describe the class of invariant test statistics is to begin with the configuration statistic $a_j = (y_j - \hat{\theta}_1)/\hat{\theta}_2$ and to transform to the unit circle via $w_j = (1 + ia_j)/(1 - ia_j)$. The maximal invariant under $\mathscr{G}$ is the vector w modulo rotations. All invariant test statistics are of the form $T(w)$ such that, for every real α,

$$T(\exp(i\alpha)(w_1, \dots, w_n)) = T(w_1, \dots, w_n).$$

The Fourier sum $T_r = \Sigma w_j^r$ is not itself invariant, but $|T_r|$ is invariant. The Fourier sums can be used to generate other invariant statistics that are also symmetric functions of y. One simple example is

$$T_{rst} = T_r T_s T_t,$$

with $r + s + t = 0$. Note that $T_{-r} = \overline{T}_r$, and $T_1 = 0$ is a consequence of the likelihood equation.

Some of these statistics have simple statistical interpretations. For example, $|T_2|$ is a measure of bimodality of the configuration on the circle; $|T_3|$ is a measure of circular trimodality and so on.

It appears to be difficult, perhaps even impossible, to construct an invariant test that is sensitive to skewness or asymmetry of any form in F. Certainly no such test exists for $n = 3$ because any three distinct points on the real line can be transformed to any other three distinct points by an element of $\mathscr{G}$. For $n = 4$ the only invariant is the cross-ratio, which does not measure skewness. I have been unable to prove or even formulate adequately the claim of nonexistence of such a test for general $n > 4$, but it is difficult to imagine what form such a test statistic could take as a function of $w_1, \ldots, w_n$. Skewness on the real line does not translate directly to skewness on the circle. For example, a large value of $|T_3|$ does not, of itself, suggest asymmetry.

3.2. *The configuration statistic.* The configuration statistic W as defined in the preceding section is a vector whose components are unit complex numbers adding to zero. The distribution of W does not depend on the parameter, but W is not invariant under $\mathscr{G}$. The orbit of W under $\mathscr{G}$ is the set of all componentwise rotations $\exp(i\alpha)W$ of W, all components rotated equally. Since the distribution of W is unaffected by g, one is tempted to conclude that $W \sim \exp(i\alpha)W$, that is, that the joint distribution of W is invariant under rotation. This claim is false. The effect on W of first applying a transformation g to Y is a rotation, but the angle of rotation $\hat{\alpha}_g$ depends on $\hat{\theta}$. The correct conclusion, that $\exp(i\hat{\alpha}_g)W \sim W$ for each $g \in \mathscr{G}$, does not imply $\exp(i\alpha)W \sim W$ for any fixed $\alpha \neq 0$. In fact, for $n = 3$, the density of $a_1\sqrt{3}$ at x can be obtained by contour integration in the form

$$\frac{1}{\pi^2(x^2 - 1)}\left(\log\left|\frac{x^2 - 1}{4}\right| + x\log\left|\frac{1 + x}{1 - x}\right|\right),$$

which has singularities at $x = \pm 1$ or $a = \pm 1/\sqrt{3}$. The density of W_1 has period $2\pi/3$ with singularities at -1, $-\omega$ and $-\omega^2$ and is clearly not uniform.

If Z has the circular Cauchy distribution (5) with parameter ψ, $(Z - \psi)/(\overline{\psi}Z - 1)$ is uniformly distributed on the unit circle. However, the circular configuration vector with components $W^* = (Z - \hat{\psi})/(\overline{\hat{\psi}}Z - 1)$ is not an ancillary statistic: its distribution depends on ψ. In general, although the configuration W^* is a rotation of W, the angle of rotation depends on $\hat{\psi}$, which induces a dependence on the parameter.

3.3. *Regression models*. The family of linear regression models for the median is not closed under Möbius transformation, so the results given here have no relevance for ordinary Cauchy regression models. However, it should be possible to extend some of the results of Section 2 to the invariant regression model described below.

Let $y_1, \dots, y_n$ be the observed values of independent Cauchy random variables Y_j, and let $z_1, \dots, z_n$ be given complex-valued nonstochastic covariates satisfying $\Im(z_j) \neq 0$. In the fractional linear regression model, the conditional distribution of Y_j given the covariates is Cauchy with parameter θ_j satisfying

$$\theta_j = \frac{\beta_{00} z_j + \beta_{01}}{\beta_{10} z_j + \beta_{11}} = \beta \circ z_j, \tag{28}$$

where the parameter β is a 2×2 real matrix with unit determinant, that is, $\beta \in \mathrm{SL}(2, R)$. In other words, Y_j has median θ_{1_j} and probable error $|\theta_{2_j}|$ given by the real and imaginary parts of (28). Thus, we may write $Y_j \sim C(\beta \circ z_j)$ for the conditional distributions of the components of Y. This family of regression models is closed under fractional linear transformation in the sense that, for each matrix $g \in \mathrm{SL}(2, R)$,

$$g \circ Y_j \sim C(g \circ \theta_j) \equiv C(g \beta \circ z_j),$$

where $g \circ \theta$ denotes fractional linear transformation and $g \beta$ denotes matrix multiplication.

Regrettably, bona fide applications of (28) are not easy to envisage.

Acknowledgments. I am grateful to T. Lyons, P. Mykland, M. Stein, R. A. Wijsman and the referees for helpful comments on several parts of this paper.

REFERENCES

COPAS, J. (1975). On the unimodality of the likelihood function for the Cauchy distribution. *Biometrika* **62** 701–704.

EATON, M. L. (1989). *Group Invariance Applications in Statistics*. IMS, Hayward, CA.

FERGUSON, T. (1978). Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4. *J. Amer. Statist. Assoc.* **73** 211–213.

KNIGHT, F. B. (1976). A characterization of the Cauchy type. *Proc. Amer. Math. Soc.* **55** 130–135.

KNIGHT, F. B. and MEYER, P. A. (1976). Charactérisation de la loi de Cauchy. *Z. Wahrsch. Verw. Gebiete* **34** 129–134.

KOLASSA, J. and MCCULLAGH, P. (1990). Edgeworth series for lattice distributions. *Ann. Statist.* **18** 981–985.

MARDIA, K. V. (1972). *Statistics of Directional Data*. Academic Press, New York.

McCullagh, P. (1992). Conditional inference and Cauchy models. *Biometrika* **79** 247–259.

McCullagh, P. (1993). On the choice of ancillary in the Cauchy location-scale problem. In *Statistics and Probability: A R. R. Bahadur Festschrift* (J. K. Ghosh, S. K. Mitra, K. R. Partasarathy and B. L. S. Prakasa Rao, eds.) 445–462. Wiley Eastern, New Delhi.

Rogers, L. C. G. and Williams, D. (1986). *Diffusions, Markov Processes, and Martingales* **2**. Wiley, New York.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill, New York.

Wold, H. (1934). Sheppard's correction formulae in several variables. *Skand. Akturarietidskrift* **17** 248–255.

Department of Statistics
University of Chicago
5734 University Avenue
Chicago, Illinois 60637