POSTERIOR MOMENTS OF THE CAUCHY DISTRIBUTION

DAVID R. WOLF
NASA Ames Research Center
MS 269-2 Moffet Field
CA 94035-1000[†]

Abstract. The posterior moments of parameters specifying distributions are minimum mean square Bayesian estimators for the corresponding moments of those parameters, and as such are ubiquitous in the Bayesian approach to statistical inference of distributions. The Cauchy distribution is most notable for its wide tails, decided absence of high-order moments, and non-existence of less-than-data dimension sufficient statistics. Thus it vastly differs qualitatively from the Gaussian distribution, where tails are small, moments of all orders exist, and dimension-two sufficient statistics always exist. In this paper the posterior moments of the position parameter of the Cauchy distribution are found in closed form. (Estimating the other parameter, the width or distance parameter of the Cauchy is done using the same mathematics.) The interplay between the amount of data acquired for the estimation of the position parameter and the existence of higher order moments of the inferred posterior distribution for the postition parameter is made explicit.

Key words: Cauchy distribution, Bayesian estimation, moment analysis, parameter estimation, parameter inference, sufficient statistics

1. Introduction

The Cauchy distribution [1], [2] has a clear physically motivated model as the normalized density function for reception on a line, of particles radiated omnidirectionally from a radiator. If the line is the x-axis and the radiator is at $(position, distance) = (x_0, y_0)$ then this density is proportional to the Cauchy likelihood distribution

$$P(x \mid (x_0, y_0)) = \frac{y_0}{\pi(y_0^2 + (x - x_0)^2)}. \tag{1}$$

Given N independent observations (receptions of particles) at positions $(x_1, \ldots, x_N)$ $=: \boldsymbol{x}$ and given that the prior for the true position of the radiator is given by

[†]Email: greywolf@ptolemy.arc.nasa.gov, Web: http://dino.ph.utexas.edu/ ˜ wolf

$P((x_0, y_0))$ then Bayes' theorem states that the probability density function for the true position of the radiator is proportional to the product of the Cauchy distribution and the prior, $P((x_0, y_0) \mid \boldsymbol{x}) \propto P(\boldsymbol{x} \mid (x_0, y_0))\, P((x_0, y_0))$. Since the observations are independent, $P(\boldsymbol{x} \mid (x_0, y_0)) = \Pi_{i=1}^{N} P(x_i \mid (x_0, y_0))$, with each term in the product having the form of (1). Taking the distance of the line from the radiator y_0 as fixed, we are left with the Cauchy posterior distribution for the position parameter x_0, $P(x_0 \mid \boldsymbol{x}, y_0)$, and this is the distribution discussed in the rest of this paper (usually dropping the redundant y_0 from the notation). However, note that the estimation of y_0 is accomplished by the exact same mathematics, as will be seen later.

Given the Cauchy posterior distribution for the position parameter x_0, we may take moments of it, and these are naturally expressed as a ratio of integrals since the proportionality constant is itself the zeroth moment of the distribution. Thus define

$$I_k[f(x) \mid \boldsymbol{x}] := \int f(x_0)^k \, P(\boldsymbol{x} \mid x_0)\, dx_0 \tag{2}$$

and satisfy oneself that for uniform $P(x_0, y_0)$

$$E[x^k \mid \boldsymbol{x}] := \int x_0^k \, P(x_0 \mid \boldsymbol{x})\, dx_0 = I_k[x \mid \boldsymbol{x}]/I_0[x \mid \boldsymbol{x}]. \tag{3}$$

The notation $I_k := I_k[x \mid \boldsymbol{x}]$ is used later. The use of the first posterior moment of the Cauchy distribution as an estimator for the true value of the position parameter is motivated by the fact that it is the Bayes' minimum mean squared error (mmse) estimator for that position parameter. Similarly, it is left as an exercise to show that the nth Cauchy posterior moment is the Bayes' mmse estimator for the nth power of the position parameter. In particular, the combination of posterior averages $E[x^2 \mid \boldsymbol{x}] - E[x \mid \boldsymbol{x}]^2$ is the mmse estimator for the variance. Thus we are immensely interested in computing these moments.

In the rest of this brief paper we show how to compute the posterior moments of the Cauchy distribution in closed form, provide explicit indications of how the existence of these moments is tied to the number of observations, discuss the use of non-uniform priors, discuss non-integer moments, and finally discuss the moment generating function. In the next section, the moments are computed when the prior is "uniform" on the *infinite* line (by treating the prior as a infinitesimal constant). Following that we discuss the computation of these moments when the prior is uniform on any *finite* interval of the line. The infinite interval solution is discussed as a two-sided limit of a large finite interval and explicit agreement with the first (infinite line) solution is shown. The final sections present comments on using non-uniform priors, and the moment generation function. For simplicity, we only consider moments of the position parameter x (moments of the distance parameter y are similar). The presentation is terse, and references to the standard mathematical literature are minimal, hopefully allowing this paper to serve as a handbook for the Cauchy distribution in Bayesian statistics.

2. Infinite integration interval

When the interval of integration is $(-\infty, \infty)$ then the problem of finding moments may be viewed as a contour integration problem. The zeros in the denominator of the Cauchy distribution are given by $x_i^\pm := x_i \pm iy_i$. Thus I_k is given by (same normalization for all k for fixed N)

$$I_k \propto \int_{-\infty}^{\infty} \prod_{i=1}^{N} x^k \frac{1}{y^2 + (x - x_i)^2}\, dx = \oint x^k \prod_{i=1}^{N} x^k \frac{1}{y^2 + (x - x_i)^2}\, dx = 2\pi i \sum_{i=1}^{N} R_i \tag{4}$$

with the residues given by

$$R_i = \frac{(x_i^+)^k}{x_i^+ - x_i^-} \prod_{k \neq i} \frac{1}{(x_i^+ - x_k^+)(x_i^+ - x_k^-)} \tag{5}$$

and integer k, $0 \leq k < 2N - 1$.

3. Finite integration interval

When the interval of integration is a finite (a, b) the residue theorem no longer applies. However, it is possible to expand the Cauchy likelihood via partial fractions to find

$$\prod_{i=1}^{N} \frac{1}{y^2 + (x - x_i)^2} = \sum_{i=1}^{N} \frac{C_i^+}{x - x_i^+} + \frac{C_i^-}{x - x_i^-} \tag{6}$$

where

$$C_i^\pm := \frac{\pm}{x_i^+ - x_i^-} \prod_{k \neq i} \frac{1}{(x_i^\pm - x_k^+)(x_i^\pm - x_k^-)}. \tag{7}$$

Now proceed with the integration as usual, noting that [1]

$$\int \frac{x^k}{x - z}\, dx = \sum_{i=1}^{k} \frac{x^i z^{k-i}}{i} + z^k log(x - z). \tag{8}$$

The result is that the desired integral is proportional to

$$I_k = \sum_{i=1}^{N} (T_i^+ + T_i^-) \mid_a^b \tag{9}$$

where

$$T_i^\pm := \sum_{i=1}^{N} C_i^\pm \left(\sum_{m=1}^{k} \frac{1}{m} x^m (x_i^\pm)^{k-m} + (x_i^\pm)^k log(x - x_i^\pm) \right) \tag{10}$$

[1] This follows immediately from the representation $\frac{x^k}{x-z} = \sum_{i=1}^{k} x^i z^{k-i-1} + \frac{z^k}{x-z}$

which simplifies to

$$I_k = \sum_{i=1}^{N} (U_i^+ + U_i^-) \mid_a^b \tag{11}$$

for [2] $0 \leq k < 2N-1$, with

$$U_i^{\pm} := C_i^{\pm} (x_i^{\pm})^k log(x - x_i^{\pm}). \tag{12}$$

4. Large integration interval

As the interval of integration becomes larger we expect the result of the section 3 for the finite interval, to approach the infinite-line result of section 2. To show this, define $r_i^{\pm}, \theta_i^{\pm}$ via $r_i^{\pm}\theta_i^{\pm} = x - x_i^{\pm}$. Then with this notation the result (11) of section 3 is

$$I_k = \sum_{i=1}^{N} C_i^+ (x_i^+)^k (log(r_i^+) + i\theta_i^+) + C_i^- (x_i^-)^k (log(r_i^-) + i\theta_i^-) \mid_a^b . \tag{13}$$

Letting $b \rightarrow \infty, a \rightarrow -\infty$ [3] and using the result of the previous footnote gives the large interval result (with the same conditions on k as in section 3)

$$I_k = 2\pi i Im[C_i^+ (x_i^+)^k]. \tag{14}$$

Using definitions (5) and (7) this is immediately seen to be the result (4) of section 2.

5. Examples

Here we give three examples which i) show how a posterior moment may fail to exist for too few observations, using the context of section 2, ii) give the first posterior moments for the 2 and 3 observation cases, and iii) give the estimator for the variance in the 2 observation case.

5.1. EXAMPLE 1 - FIRST MOMENT OF THE 1 OBSERVATION POSTERIOR

Here $k = 1 = 2N-1$ so we expect that the moment will not exist, since $k < 2N-1$ was required, see section 3. Clearly, the numerator term for $E[x \mid \boldsymbol{x}]$ with $a \rightarrow -\infty, b \rightarrow \infty$ does not exist because the integrand of the numerator is logarithmically divergent for $k = 1$. However it is still interesting to see explicitly how this integral is ill defined in the context of the results of section 3. The expansion of the integrand of I_1 is there given by $C_1^+ x/(x - x_1^+) + C_1^- x/(x - x_1^-)$, and this integrates to $(x_1^+ log(x - x_1^+) - x_1^- log(x - x_1^-))/(2iy)$. Converting to polar

[2] Consider the leading order coefficient of x in the numerator of the expansion of $\prod_{i=1}^{N} \frac{1}{y^2 + (x - x_i)^2}$. In particular, $\sum_{i=1}^{N} C_i^+ (x_i^+)^k + C_i^- (x_i^-)^k = 0$ for $k = 0, 1, \ldots, 2(N-1)$.

[3] Write $log(r^{\pm}) = log(r(1 + \epsilon^{\pm}))$ and note that $\epsilon^{\pm} \rightarrow 0$ for $b \rightarrow \infty, a \rightarrow -\infty$, respectively.

form gives $(x_1^+ log(r_1^+) - x_1^- log(r_1^-) + x_1^+ i\theta_i^+ - x_1^- i\theta_i^-)/(2iy)$. As can be immediately seen, the only case when cancellation of the logarithms in the numerator is to be expected for $a \to -\infty, b \to \infty$ independently is when $y = 0$, yet we must also divide by that $y = 0$. Thus the integral is ill defined for this case, and the moment fails to exist. We may take the principle value of the integral as the object of interest. In that case, the desired moment is simply x_1.

5.2. EXAMPLE 2, FIRST MOMENTS OF THE 2 AND 3 OBSERVATION POSTERIORS

For two observations, the result is trivially given by $E[x \mid (x_1, x_2)] = (x_1 + x_2)/2$. However, the pattern set here (the average) does not continue; for 3 observations the first posterior moment is given by $E[x \mid (x_1, x_2, x_3)] =$

$$\frac{8y^2(x_1 + x_2 + x_3) + x_1^2(x_2 + x_3) + x_2^2(x_1 + x_3) + x_3^2(x_1 + x_2) - 6x_1x_2x_3}{2(12y^2 + x_1^2 + x_2^2 + x_3^2 - (x_1x_2 + x_2x_3 + x_3x_1))} \tag{15}$$

Interestingly, when $y = 0$ this is not the average. Another interesting observation is that the difference of this and the average of the three points has a numerator given by the symmetric form $(-2x_1 + x_2 + x_3)(-2x_2 + x_3 + x_1)(-2x_3 + x_1 + x_2)$. This pattern, too, does not continue. For example, for $N = 2$ the corresponding difference is zero.

5.3. EXAMPLE 3, ESTIMATOR FOR THE STANDARD DEVIATION WITH TWO OBSERVATIONS.

The second moment of the two observation posterior is given by $y^2 + (x_1^2 + x_2^2)/2$. This, along with the first moment result from example 2 above gives

$$E[(x - E[x \mid \boldsymbol{x}])^2 \mid \boldsymbol{x}] = y^2 + (x_1 - x_2)^2/4. \tag{16}$$

6. Non-uniform priors

Briefly, for an infinite interval of integration, note that when the prior is any finite sum of the form $\sum_{i=0}^{n} a_i x^i$ then moments up to order $2(N-1) - n$ exist if this quantity is non-negative. For a finite interval of integration then any series which when multiplied by the likelihood is properly integrated term-by-term may be used as a prior. Rational functions where the numerator degree minus the denominator degree is less than or equal to $2(N-1)$ may be used as priors.

7. Non-integer moments

For a finite interval of integration and uniform prior, the binomial series [4] may be used to develop an approximation in the following way. Let q be a non-integer real,

[4] The binomial series for non-integer α is given by $(1+x)^\alpha = \sum_{i=0}^{\infty} \frac{1}{i!} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-i)} x^i$, where $|x| < 1$.

and consider I_q. Make the expansion of the denominator of the Cauchy likelihood as done in section 3. In each of the terms make the substitutions $u_i^{\pm} = x - x_i^{\pm}$ respectively. Now, expand each of the resulting numerators $(u_i^{\pm} + x_i^{\pm})^q$ using the binomial series appropriate for the limit under consideration (depending upon whether $\mid u_i^{\pm} \mid > x_i^{\pm}$ or $\mid u_i^{\pm} \mid < x_i^{\pm}$ at the limits $x = a, b$ as appropriate). When it converges, then the series representing the integration may be truncated at whatever order is needed for the desired accuracy.

8. Moment generating function

The kth moment is given in terms of the kth derivative of the moment generating function by $E[x^k \mid \boldsymbol{x}] = \partial_\alpha^k E[e^{\alpha x} \mid \boldsymbol{x}] \mid_{\alpha=0}$. The moment generating function for the N-observation Cauchy distribution may be found in closed form. From (6)

$$I_k[e^{\alpha x} \mid \boldsymbol{x}] = \int_a^b \left(\sum_{i=1}^N \frac{C_i^+ e^{\alpha x}}{x - x_i^+} + \frac{C_i^- e^{\alpha x}}{x - x_i^-} \right) dx. \tag{17}$$

After making appropriate variable changes each of the terms in the sum on the right is of the form [5]

$$Ce^{\alpha x_0} \int_{\alpha(a-x_0)}^{\alpha(b-x_0)} \frac{e^u}{u} du = Ce^{\alpha x_0}[Ei(\alpha(b - x_0)) - Ei(\alpha(a - x_0))]. \tag{18}$$

Making the substitutions for the moment generating function yields

$$E[e^{\alpha x} \mid \boldsymbol{x}] = \frac{\sum_{i=1}^N Re(C_i^+ e^{\alpha x_i^+}[Ei(\alpha(b - x_i^+)) - Ei(\alpha(a - x_i^+))])}{\sum_{i=1}^N Re(C_i^+ s)} \tag{19}$$

Clearly, all moments are easily accessible since $\partial_x Ei(x) = e^x/x$. As a check of the first moment

$$\lim_{\alpha \to 0} [Ei(\alpha q) - Ei(\alpha p)] = log(p) - log(q). \tag{20}$$

9. Acknowledgements

I thank Kenneth Hanson for suggesting this problem. Part of this work was sponsored by the United States Department of Energy under contract number W-7405-ENG-36. Part of this work was sponsored by NASA.

A. Mathematica program for finding Cauchy posterior moments

This Mathematica program finds the kth posterior moment of the position of the Cauchy distribution, given the N observations $(x_1, \ldots, x_N)$, the distance $y = y_0$ and "uniform" prior on the line.

[5] The Exponential Integral is given by $Ei(z) = \int^z e^u/u du$ where the path of integration does not cut $[0, \infty)$ [3]. It may be written as $Ei(z) = \gamma + log(-z) + \sum_{i=1}^\infty z^i/(i!i)$ in the plane cut along the positive real axis.

```
kthMoment[k_Integer,N_Integer] := kthIntegral[k,N] / kthIntegral[0,N],
```

where

```
kthIntegral[k_Integer,N_Integer] :=
Block[ {i,j,Result},
        For[i=1,i<=N,i++,
            xp[i] = x[i] + I y;
            xm[i] = x[i] - I y;
        ];

        For[i=1,i<=N,i++,
            Cp[i] =  1/(xp[i]-xm[i]);
            Cm[i] = -1/(xp[i]-xm[i]);
            For[j=1,j<=N,j++,
                If[ j!=i,
                    Cp[i] *= 1/(xp[i]-xp[j]) 1/(xp[i]-xm[j]);
                    Cm[i] *= 1/(xm[i]-xp[j]) 1/(xm[i]-xm[j]);
                ];
            ];
        ];

        Result = 0;

        For[i=1,i<=N,i++,
            Result += Cp[i] xp[i]^k - Cm[i] xm[i]^k;
        ];

        Return[Result];
]
```

References

1. K. M. Hanson and D. R. Wolf, “Estimators for the cauchy distribution,” in *Maximum Entropy and Bayesian Methods in Science and Engineering*, G. Heidbreder, ed., pp. 255–263, Kluwer Academic, Dordrecht, 1996.
2. S. F. Gull, “Bayesian inductive inference and maximum entropy,” in *Maximum Entropy and Bayesian Methods in Science and Engineering*, G. Erickson and C. R. Smith, eds., pp. 53–74, Kluwer Academic, Dordrecht, 1988.
3. M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover, NY, 1972.