A closed-form formula for the Kullback-Leibler divergence between Cauchy distributions

Frédéric Chyzak[†] and Frank Nielsen[⋆]

[†]INRIA
[†]Université Paris–Saclay, France
[†]frederic.chyzak@inria.fr
[⋆]Sony Computer Science Laboratories, Inc.
[⋆]Tokyo, Japan
[⋆]Frank.Nielsen@acm.org

Abstract

We report a closed-form expression for the Kullback-Leibler divergence between Cauchy distributions which involves the calculation of a parametric definite integral with 6 parameters. The formula shows that the Kullback-Leibler divergence between Cauchy densities is always finite and symmetric.

1 Introduction

The Cauchy distribution $P_{l,s}$ has the following density, $p_{l,s}$, with respect to the Lebesgue measure μ on $(-\infty, \infty)$:

$$p_{l,s}(x) = \frac{\mathrm{d}P_{l,s}}{\mathrm{d}\mu}(x) = \frac{s}{\pi(s^2 + (x - l)^2)}, \tag{1}$$

where l and s denote the location parameter and scale parameter, respectively.

The set of Cauchy densities form a location-scale family [7, 9] with Cauchy standard distribution

$$p(x) = p_{0,1}(x) = \frac{1}{\pi(1 + x^2)}. \tag{2}$$

Indeed, for any density p (called the standard density), one can generate a location-scale family $\{p_{l,s} \mid l \in \mathbb{R}, s > 0\}$ by defining the density

$$p_{l,s}(x) = \frac{1}{s}p\left(\frac{x - l}{s}\right). \tag{3}$$

We want to calculate the Kullback-Leibler (KL) divergence [5] between two Cauchy densities p_{l_1,s_1} and p_{l_2,s_2}:

$$\mathrm{KL}(p_{l_1,s_1} : p_{l_2,s_2}) = \int p_{l_1,s_1}(x) \log \frac{p_{l_1,s_1}}{p_{l_2,s_2}} \mathrm{d}\mu(x). \tag{4}$$

Since the KL divergence is the relative entropy [5] (i.e., the difference between the cross-entropy and the entropy), we have the following identity:

$$\mathrm{KL}(p_{l_1,s_1} : p_{l_2,s_2}) = h^\times(p_{l_1,s_1} : p_{l_2,s_2}) - h(p_{l_1,s_1}), \tag{5}$$

where $h^\times$ denotes the cross-entropy and $h(p_{l_1,s_1}) = h^\times(p_{l_1,s_1} : p_{l_1,s_1})$ the differential entropy [6]. Thus it is enough to derive a closed-form expression of the cross-entropy between two Cauchy densities to get the closed-form formula for the KL divergence between these densities. Observe that in general, the cross-entropy $-\int p(x)\ln q(x)\mathrm{d}\mu(x)$ between two densities p and q may not be finite when the integral diverges. The cross-entropy and entropy for densities belonging to an exponential family have been reported in [10].

2 Closed-form formula

We have

$$
\begin{aligned}
h^\times(p_{l_1,s_1} : p_{l_2,s_2}) &= -\int \frac{s_1}{\pi(s_1^2 + (x-l_1)^2)} \log \frac{s_2}{\pi(s_2^2 + (x-l_2)^2)} \mathrm{d}\mu(x), &(6)\\
&= \log \frac{\pi}{s_2} + \frac{s_1}{\pi} \int \frac{\log(x^2 - 2l_2x + s_2^2 + l_2^2)}{x^2 - 2l_1x + s_1^2 + l_1^2} \mathrm{d}\mu(x), &(7)\\
&= \log \frac{\pi}{s_2} + \frac{s_1}{\pi} A(1, -2l_1, l_1^2 + s_1^2; 1, -2l_2, l_2^2 + s_2^2), &(8)
\end{aligned}
$$

where we define the parametric definite integral A as follows:

$$A(a, b, c; d, e, f) = \int_{-\infty}^{\infty} \frac{\log(dx^2 + ex + f)}{ax^2 + bx + c} \mathrm{d}x, \tag{9}$$

for both positive polynomials $ax^2 + bx + c$ and $dx^2 + ex + f$ on the reals. (For a positive polynomial $ux^2 + vx + w$ on the reals, we mean a polynomial given by coefficients satisfying $u > 0$, $v \in \mathbb{R}$, $w > 0$, and $4uw - v^2 > 0$.) In particular, the polynomials $x^2 - 2l_1x + s_1^2 + l_1^2$ and $x^2 - 2l_2x + s_2^2 + l_2^2$ are both positive.

Using the MGFUN package[1] for the computer algebra system Maple, we calculate the definite integral A (see Appendix A for details), and get

Proposition 1. *We have*

$$A(a, b, c; d, e, f) = \frac{2\pi \left(\log(2af - be + 2cd + \sqrt{4ac - b^2}\sqrt{4df - e^2}) - \log(2a)\right)}{\sqrt{4ac - b^2}}. \tag{10}$$

Thus we get overall the following expression for the Kullback-Leibler divergence between Cauchy densities (see the end of Appendix A):

Theorem 1. *The Kullback-Leibler divergence between Cauchy density p_{l_1,s_1} and p_{l_2,s_2} is*

$$\mathrm{KL}(p_{l_1,s_1} : p_{l_2,s_2}) = \log \frac{(s_1 + s_2)^2 + (l_1 - l_2)^2}{4s_1s_2}. \tag{11}$$

---

[1] https://specfun.inria.fr/chyzak/mgfun.html

Observe that the KL divergence is always finite and symmetric:

$$\mathrm{KL}(p_{l_1,s_1} : p_{l_2,s_2}) = \mathrm{KL}(p_{l_2,s_2} : p_{l_1,s_1}). \tag{12}$$

The symmetric property does not hold in general for location-scale families (e.g., normal family). In particular, when $l_1 = l_2 = l$ (scale family), we have

$$\mathrm{KL}(p_{l,s_1} : p_{l,s_2}) = 2 \log \frac{s_1 + s_2}{2\sqrt{s_1 s_2}}, \tag{13}$$

recovering the result stated in [8].

When $s_1 = s_2 = s$ (location family), we have

$$\mathrm{KL}(p_{l_1,s} : p_{l_2,s}) = \log \left(1 + \frac{(l_1 - l_2)^2}{4s^2} \right). \tag{14}$$

The cross-entropy between two location-scale Cauchy distributions is

$$h^\times(p_{l_1,s_1} : p_{l_2,s_2}) = \log \frac{\pi((s_1 + s_2)^2 + (l_1 - l_2)^2)}{s_2}. \tag{15}$$

It follows that the differential entropy is $h(p_{l,s}) = h^\times(p_{l,s} : p_{l,s}) = \log 4\pi s$, in accordance with [6].

By the same change of variable in the KL definition Eq. 4 as the one used in the appendix, we find the following key property [9] of the KL divergence between location-scale densities:

$$\mathrm{KL}(p_{l_1,s_1} : p_{l_2,s_2}) = \mathrm{KL}(p_{0,1} : p_{\frac{l_2 - l_1}{s_1}, \frac{s_2}{s_1}}). \tag{16}$$

Note that this property holds for any f-divergence in general, that is for integrals

$$\int f\left(\frac{p(x)}{q(x)} \right) q(x) \mathrm{d}\mu(x)$$

given for two densities by a fixed function f.

Indeed, we check that

$$
\begin{aligned}
\mathrm{KL}(p_{0,1} : p_{\frac{l_2 - l_1}{s_1}, \frac{s_2}{s_1}}) &= \log \frac{\left(1 + \frac{s_2}{s_1}\right)^2 + \left(0 - \frac{l_2 - l_1}{s_1}\right)^2}{4 \frac{s_2}{s_1}}, & (17) \\
&= \log \frac{(s_1 + s_2)^2 + (l_1 - l_2)^2}{4 s_1 s_2}, & (18) \\
&= \mathrm{KL}(p_{l_1,s_1} : p_{l_2,s_2}). & (19)
\end{aligned}
$$

3

A A Maple-assisted calculation of the definite integral A and of the KL divergence

To the best of our knowledge, the closed-form formula for the definite integral A is only given with no explicit proof in the literature, in an almost equivalent form [11, Eq. 19, page 514] as

$$\int_{-\infty}^{\infty} \log(a^2 - 2abx + x^2)\frac{\mathrm{d}x}{x^2 + z^2} = \frac{\pi}{z}\log(z^2 + 2az\sqrt{1 - b^2} + a^2), \quad a > 0,\ \Re z > 0,\ b \in (-1, 1].$$

For real z, this integral is a specialization of our more general integral A. The implicit proof there is, following an algorithm by Marichev described in [1], to express the integrand as a product of so-called Meijer G-functions, to use one of the formulas in [12, Sec. 2.24], providing a "closed-form" formula as another Meijer G-function, to identify this formula as the right-hand side of (10).

   For an explicit proof, we prefer another approach, named creative telescoping, that makes algorithmic use of derivation under the integral sign. The algorithm that we use can be viewed as an extension [3] applicable to integrands satisfying higher-order linear differential equations of an earlier algorithm for hyperexponential integrands [2], that is, satisfying first-order equations. (See also [4] for an introduction to developments of creative telescoping.)

   The calculation starts with proving the special case $(a, b, c) = (1, 0, 1)$. The integrand is then

$$\phi(d, e, f; x) = \frac{\ln(dx^2 + ex + f)}{x^2 + 1}.$$

It satisfies second-order homogeneous linear differential equations, one for each of the variables d, e, f, x, as well as relations between the first-order derivatives. Chyzak's implementation of [3] as the MGFUN package for Maple readily derives the nonhomogeneous relation

$$\mathcal{L}\frac{\mathrm{d}\phi}{\mathrm{d}d}(d, e, f; x) = \frac{\mathrm{d}\psi}{\mathrm{d}x}(d, e, f; x), \tag{20}$$

where we define a linear differential operator $\mathcal{L}$ by

$$\mathcal{L}y = (2df + e^2 + 6f^2)(4df - e^2)(d^2 - 2df + e^2 + f^2)\frac{\mathrm{d}^3 y}{\mathrm{d}d^3}$$

$$+ (60d^3f^2 + 24d^2e^2f + 132d^2f^3 - 6de^4 - 60de^2f^2 - 252df^4 + 18e^4f + 108e^2f^3 + 60f^5)\frac{\mathrm{d}^2 y}{\mathrm{d}d^2}$$

$$+ (96d^2f^2 + 60de^2f + 336df^3 - 6e^4 - 84e^2f^2 - 240f^4)\frac{\mathrm{d}y}{\mathrm{d}d} + 24(df + e^2 + 5f^2)fy,$$

as well as

$$\psi(d, e, f; x) = \frac{-2xP(d, e, f; x)}{(dx^2 + ex + f)^2}\frac{\mathrm{d}\phi}{\mathrm{d}d}(d, e, f; x)$$

for the polynomial

$$P(d, e, f; x) = e(d^2f^2 - 4de^2f - 9df^3 - e^4 - 7e^2f^2 - 6f^4)x^5$$
$$- 3f(3de^2f + 4df^3 + 2e^4 + 11e^2f^2 + 4f^4)x^4 + e(d^2f^2 - 4de^2f - 18df^3 - e^4 - 16e^2f^2 - 51f^4)x^3$$
$$- f(9de^2f + 16df^3 + 6e^4 + 37e^2f^2 + 32f^4)x^2 - 9ef^2(df + e^2 + 5f^2)x - 4f^3(df + e^2 + 5f^2).$$

4

Observe that (20) is really a differential relation on the rational function

$$\frac{\mathrm{d}\phi}{\mathrm{d}d}(d,e,f;x) = \frac{x^2}{(dx^2 + ex + f)(x^2 + 1)},$$

which is indefinitely differentiable with respect to x on the integration interval $\mathbb{R}$, and that

$$\lim_{x \to +\infty} \psi(d,e,f;x) = \lim_{x \to -\infty} \psi(d,e,f;x) = -2e(d^2f^2 - 4de^2f - 9df^3 - e^4 - 7e^2f^2 - 6f^4)/d^3.$$

As a consequence, (20) can be integrated over $\mathbb{R}$, leading to

$$\mathcal{L}\frac{\mathrm{d}A}{\mathrm{d}d}(1,0,1;d,e,f) = 0.$$

The generation and analysis leading to the previous homogeneous differential equation is done in Maple as follows.

```
libname := "./Mgfun.mla", libname:
assms := a>0, b :: real, c>0, 4*a*c-b^2 > 0, d>0, e :: real, f>0, 4*d*f-e^2 > 0;
phi := ln(d*x^2+e*x+f)/(x^2+1);
ct := Mgfun:-creative_telescoping(phi, d::diff, x::diff)[1];
L := collect(subs(diff(_F(d), d) = y(d), ct[1]), diff, factor);
map(p -> collect(p, x, factor), ct[2]);
P := collect(ct[2] / (-2*x*diff(_f(d,x),d)) * (d*x^2+e*x+f)^2, x, factor);
diff(phi, d);
psi := factor(eval(ct[2], _f(d,x) = phi));
limit(psi, x=+infinity);
limit(psi, x=-infinity);
%% - %;
```

Now, note that the Cauchy–Lipshitz theorem applies to the differential equation $\mathcal{L}y = 0$ at $d = 0$. Maple can indeed find algebraic expressions for the integrals

$$\frac{\mathrm{d}^i A}{\mathrm{d}d^i}(1,0,1;d,e,f) = \int_{-\infty}^{\infty} \frac{\mathrm{d}^i \phi}{\mathrm{d}d^i}(d,e,f;x)\mathrm{d}x$$

when $1 \le i \le 4$. Such expressions are in $\mathbb{Q}(\pi, d, e, f, \sqrt{4df - e^2})$. Maple can then solves the differential equation $\mathcal{L}y = 0$ under the constraints $\frac{\mathrm{d}^i y}{\mathrm{d}d^i}(d,e,f) = \frac{\mathrm{d}^i A}{\mathrm{d}d^i}(1,0,1;d,e,f)$, thus proving

$$\frac{\mathrm{d}A}{\mathrm{d}d}(1,0,1;d,e,f) = \pi\frac{(d-f)(4df - e^2) + (-2df + e^2 + 2f^2)\sqrt{4df - e^2}}{(d^2 - 2df + e^2 + f^2)(4df - e^2)}. \tag{21}$$

Maple can then compute a primitive with respect to d, which provides $A(1,0,1;d,e,f)$

$$A(1,0,1;d,e,f) = \frac{\pi}{2}(\ln G_1 - \ln G_2 + \ln G_3) + K(e,f) \tag{22}$$

where

$$G_1(d,e,f) = d + f + \sqrt{4df - e^2}, \tag{23}$$
$$G_2(d,e,f) = d + f - \sqrt{4df - e^2}, \tag{24}$$
$$G_3(d,e,f) = d^2 - 2df + e^2 + f^2 = (d-f)^2 + e^2 = G_1(d,e,f)G_2(d,e,f) \tag{25}$$

and K is a function K still to be determined.

5

```
evala(Normal(dsolve({
  eval(ct[1], diff(_F(d), d) = y(d)),
  seq(diff(y(d), [d$i]) =
        evala(Normal(int(diff(phi, [d$(i+1)]), x=-infinity..infinity)
          assuming assms)), i=0..3)
})));
A := simplify(int(op([1,2], %), d)) assuming assms;
```

Note that an alternative, more direct derivation up to this point is by observing that $\phi(d, e, f; x)$ and $\frac{\mathrm{d}\phi}{\mathrm{d}d}(d, e, f; x)$ are both integrable on $\mathbb{R}$, so that one has

$$A(1, 0, 1; d, e, f) = \int \left(\int_{-\infty}^{\infty} \frac{\mathrm{d}\phi}{\mathrm{d}d}(d, e, f; x)\mathrm{d}x \right) \mathrm{d}t.$$

In Maple, this leads to the same formula as with the method of creative telescoping, using the following commands:

```
assms := a>0, b :: real, c>0, 4*a*c-b^2 > 0, d>0, e :: real, f>0, 4*d*f-e^2 > 0;
phi := ln(d*x^2+e*x+f)/(x^2+1);
A := simplify(int(int(diff(phi, d), x=-infinity..infinity), d)) assuming assms;
```

In a first step, the previous, one-line derivation calculates a primitive with respect to x:

$$B(d, e, f; x) = \int_0^x \frac{\mathrm{d}\phi}{\mathrm{d}d}(d, e, f; y)\mathrm{d}y =$$

$$\frac{2\sqrt{4df - e^2}\left((d - f)\arctan x - \frac{e}{2}\ln(dx^2 + ex + f) + \frac{e}{2}\ln(x^2 + 1)\right) - 4(df - \frac{e^2}{2} - f^2)\arctan\frac{2dx+e}{\sqrt{4df-e^2}}}{(2d^2 - 4df + 2e^2 + 2f^2)\sqrt{4df - e^2}}.$$

Then, the integral A can be derived as a difference of limits of B at $x = \pm\infty$. For the primitivation, Maple uses the Risch algorithm for determining if a primitive of a so-called Liouvillian function can be expressed as a Liouvillian function [13, 14]. This includes the case of rational expressions in logarithms, like ϕ, and for ϕ a primitive is in the class. The primitive B can be obtained in Maple by the following command:

```
B := simplify(int(diff(phi, d), x)) assuming assms;
```

As G_1 and G_3 are positive by construction when $d > e^2/4/f$, so is G_2, and thus, $\ln G_3 = \ln G_1 + \ln G_2$, so that $A(1, 0, 1; d, e, f) = \pi \ln G_1 + K(e, f)$. To fix the integration constant K, we successively take derivatives with respect to f and e, and use the fact that Maple is, like above for $\frac{\mathrm{d}\phi}{\mathrm{d}d}$, able to integrate $\frac{\mathrm{d}\phi}{\mathrm{d}f}$ and $\frac{\mathrm{d}\phi}{\mathrm{d}e}$ with respect to x. This proves that K is in fact independent of f and e. Choosing next $f = d$ and $e = 2d$ proves $K = 0$, and the final formula for the special case $(a, b, c) = (1, 0, 1)$ is:

$$A(1, 0, 1; d, e, f) = \pi \ln(d + f + \sqrt{4df - e^2}). \tag{26}$$

```
(d+f+sqrt(4*d*f-e^2)) * (d+f-sqrt(4*d*f-e^2));
A := simplify(subs(expand(%) = %, A)) assuming op(1, %) > 0, op(2, %) > 0;
simplify(int(diff(phi, f), x=-infinity..infinity) - diff(A, f)) assuming assms;
simplify(int(diff(phi, e), x=-infinity..infinity) - diff(A, e)) assuming assms;
{f = d, e = 2*d};
simplify(int(eval(phi, %), x=-infinity..infinity) - eval(A, %)) assuming assms;
EQN := Int(phi, x=-infinity..infinity) = A;
```

6

Dealing with the general case next amounts to reducing it to the previous formula, by a suitable change of integration parameter. Namely, after setting

$$D = \frac{d(4ac - b^2)}{4a^2}, \quad E = \frac{(ae - bd)\sqrt{4ac - b^2}}{2a^2}, \quad F = \frac{4a^2f - 2abe + b^2d}{4a^2}, \quad K = \frac{2}{\sqrt{4ac - b^2}}, \quad (27)$$

one has

$$A(a, b, c; d, e, f) = K A(1, 0, 1; D, E, F).$$

This is done in Maple as follows:

```
phi := ln(d*x^2+e*x+f)/(a*x^2+b*x+c);
alt_phi := eval(factor(normal(
  subs(x = (x*sqrt(4*a*c-b^2)-b)/2/a, phi) * sqrt(4*a*c-b^2)/2/a)),
  ln = (e -> ln(collect(e, x, normal)))));
K := eval(alt_phi, ln=1)*(x^2+1);
alt_phi := alt_phi / K;
op(select(has, alt_phi, ln));
local D := coeff(%, x, 2);
E := coeff(%%, x, 1);
F := coeff(%%%, x, 0);
GEN_EQN := Int(phi, x=-infinity..infinity) =
  K * simplify(eval(op(2, EQN), {d = D, e = E, f = F})) assuming assms;
```

The cross-entropy, then our expression (11) for the KL divergence, follow by appealing to the same kind of simplification as above in, respectively, (15) and (5):

```
GEN_A := op(2, GEN_EQN);
h_cross := simplify(log(Pi/s2) + s1/Pi * eval(GEN_A,
  {a=1, b=-2*l1, c=l1^2+s1^2, d=1, e=-2*l2, f=l2^2+s2^2})) assuming s1>0, s2>0;
KL := simplify(h_cross - eval(h_cross, {l2=l1, s2=s1})) assuming s1>0, s2>0;
```

As a final note, we remark that we have used the Maple syntax `...` `assuming assms` to allow the automatic use of analytic formulas for logarithms and square roots, both in the procedures `ln` and `sqrt` for mathematical functions, and in the procedures `int` and `simplify` for integration and simplification. We have found this more efficient than using the `assume(...)` facility, which would have attached properties to the names `a`, . . . , `f`, `x`.

References

[1] V. S. Adamchik and O. I. Marichev. The algorithm for calculating integrals of hypergeometric type functions and its realization in REDUCE system. In *ISSAC'90*, pages 212–224. ACM, New York, 1990.

[2] Gert Almkvist and Doron Zeilberger. The method of differentiating under the integral sign. *Journal of Symbolic Computation*, 10(6):571–591, 1990.

[3] Frédéric Chyzak. An extension of Zeilberger's fast algorithm to general holonomic functions. *Discrete Mathematics*, 217(1-3):115–134, 2000.

[4] Frédéric Chyzak. *The ABC of Creative Telescoping: Algorithms, Bounds, Complexity*. Memoir of accreditation to supervise research (HDR), Université d'Orsay, April 2014.

[5] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[6] Joseph Victor Michalowicz, Jonathan M Nichols, and Frank Bucholtz. *Handbook of differential entropy*. CRC Press, 2013.

[7] M. K. Murray and J. W. Rice. *Differential Geometry and Statistics*, volume 48. CRC Press, 1993.

[8] Frank Nielsen. On the Jensen-Shannon symmetrization of distances relying on abstract means. *Entropy*, 21(5), 2019.

[9] Frank Nielsen. On the Kullback-Leibler divergence between location-scale densities. *CoRR*, abs/1904.10428, 2019.

[10] Frank Nielsen and Richard Nock. Entropies and cross-entropies of exponential families. In *2010 IEEE International Conference on Image Processing*, pages 3621–3624. IEEE, 2010.

[11] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev. *Integrals and Series. Vol. 1: Elementary functions*. Gordon & Breach, New York, 1986.

[12] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev. *Integrals and series. Vol. 3: More special functions*. Gordon & Breach, New York, 1990.

[13] Robert H. Risch. The problem of integration in finite terms. *Transactions of the American Mathematical Society*, 139:167–189, 1969.

[14] Robert H. Risch. The solution of the problem of integration in finite terms. *Bulletin of the American Mathematical Society*, 76:605–608, 1970.