On f-divergences between Cauchy distributions*

Frank Nielsen
Sony Computer Science Laboratories Inc
E-mail: `Frank.Nielsen@acm.org`

Kazuki Okamura
Shinshu University, School of General Education
E-mail: `kazukio@shinshu-u.ac.jp`

Abstract

We prove that the f-divergences between univariate Cauchy distributions are always symmetric and can be expressed as functions of the chi-squared divergence. We show that this property does not hold anymore for multivariate Cauchy distributions. We then present several metrizations of f-divergences between univariate Cauchy distributions.

Keywords: Cauchy distributions; Complex analysis; Maximal invariant; Information geometry.

1 Introduction

Let $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ be the sets of real numbers, non-negative real numbers, and positive real numbers, respectively. The probability density function of a Cauchy distribution is

$$p_{l,s}(x) := \frac{1}{\pi s\left(1+\left(\frac{x-l}{s}\right)^2\right)} = \frac{s}{\pi(s^2+(x-l)^2)},$$

where $l \in \mathbb{R}$ denotes the location parameter and $s \in \mathbb{R}_{++}$ the scale parameter of the Cauchy distribution, and $x \in \mathbb{R}$. The space of Cauchy distributions form a location-scale family

$$\mathcal{C} = \left\{ p_{l,s}(x) := \frac{1}{s} p\left(\frac{x-l}{s}\right) \ : \ (l,s) \in \mathbb{R} \times \mathbb{R}_{++} \right\},$$

with standard density

$$p(x) := \frac{1}{\pi(1+x^2)}.$$

To measure the dissimilarity between two continuous probability distributions P and Q, we consider the class of statistical f-divergences [5, 20] between their corresponding probability densities functions $p(x)$ and $q(x)$ assumed to be strictly positive on $\mathbb{R}$:

$$I_f(p:q) := \int_{\mathbb{R}} p(x) f\left(\frac{q(x)}{p(x)}\right) \mathrm{d}x,$$

*The second author was supported by JSPS KAKENHI 19K14549.

where $f(u)$ is a convex function on $(0,\infty)$, strictly convex at $u=1$, and satisfying $f(1)=0$ so that $I_f(p:q)\geq f(1)=0$ by Jensen's inequality. The Kullback-Leibler divergence (KLD also called relative entropy) is a f-divergence obtained for $f_{\mathrm{KL}}(u)=-\log u$. In general, the f-divergences are oriented dissimilarities: $I_f(p:q)\neq I_f(q:p)$ (eg., the KLD). The reverse f-divergence $I_f(q:p)$ can be obtained as a forward f-divergence for the conjugate function $f^*(u):=uf\left(\frac{1}{u}\right)$ (convex with $f^*(1)=0$): $I_f(q:p)=I_{f^*}(p:q)$. In general, calculating the definite integrals of f-divergences is non trivial: For example, the formula for the KLD between Cauchy densities was only recently obtained [4]:

$$\begin{aligned}
D_{\mathrm{KL}}(p_{l_1,s_1}:p_{l_2,s_2}) &:= I_{f_{\mathrm{KL}}}(p:q)=\int p_{l_1,s_1}(x)\log\frac{p_{l_1,s_1}(x)}{p_{l_2,s_2}(x)}\mathrm{d}x,\\
&= \log\left(\frac{(s_1+s_2)^2+(l_1-l_2)^2}{4s_1s_2}\right).
\end{aligned}$$

Let $\lambda=(\lambda_1=l,\lambda_2=s)$. Then we can rewrite the KLD formula as

$$D_{\mathrm{KL}}(p_{\lambda_1}:p_{\lambda_2})=\log\left(1+\frac{1}{2}\chi(\lambda_1,\lambda_2)\right), \tag{1}$$

where

$$\chi(\lambda,\lambda'):=\frac{(\lambda_1-\lambda_1')^2+(\lambda_2-\lambda_2')^2}{2\lambda_2\lambda_2'}.$$

We observe that the KLD between Cauchy distributions is symmetric: $D_{\mathrm{KL}}(p_{l_1,s_1}:p_{l_2,s_2})=D_{\mathrm{KL}}(p_{l_2,s_2}:p_{l_1,s_1})$. Let $D_\chi^N(p:q):=\int\frac{(p(x)-q(x))^2}{q(x)}\mathrm{d}x$ and $D_\chi^P(p:q):=\int\frac{(p(x)-q(x))^2}{p(x)}\mathrm{d}x$ denote the Neyman and Pearson chi-squared divergences between densities $p(x)$ and $q(x)$. These divergences are f-divergences [20] for the generators $f_\chi^P(u)=(u-1)^2$ and $f_\chi^N(u)=\frac{1}{u}(u-1)^2$, respectively. The χ^2-divergences between Cauchy densities are symmetric [17]:

$$D_\chi(p_{\lambda_1}:p_{\lambda_2}):=D_\chi^N(p_{\lambda_1}:p_{\lambda_2})=D_\chi^P(p_{\lambda_1}:p_{\lambda_2})=\chi(\lambda_1,\lambda_2),$$

hence the naming of the function $\chi(\cdot,\cdot)$. Notice that we have

$$\chi(p_{\lambda_1}:p_{\lambda_2})=\rho(\lambda_1)\rho(\lambda_2)\,\frac{1}{2}D_E^2(\lambda_1,\lambda_2),$$

where $D_E(\lambda_1,\lambda_2)=\sqrt{(\lambda_2-\lambda_1)^\top(\lambda_2-\lambda_1)}$ and $(\lambda_2-\lambda_1)^\top$ denotes the transpose of the vector $(\lambda_2-\lambda_1)$. That is, the function χ is a conformal half squared Euclidean divergence [22, 21] with conformal factor $\rho(\lambda):=\frac{1}{\lambda_2}$.

In this work, we first prove in §3 that all f-divergences between univariate Cauchy distributions are symmetric (Theorem 1) and can be expressed as a function of the chi-squared divergence (Theorem 2). This property holds only for the univariate case as we report in §4 an example of bivariate Cauchy distributions for which the KLD is asymmetric. In §5, we report a simple proof of the KLD formula of Eq. 1 based on complex analysis. Similarly, we report a proof of the chi-squared divergence between two univariate Cauchy distributions in §6. Finally, we consider metrizations of the KLD and the Bhattacharyya distance in §7.

Let us start by recalling the hyperbolic nature of the Fisher-Rao geometry of location-scale families.

2 Information geometry of location-scale families

The Fisher information matrix [15, 17] of a location-scale family with continuously differentiable standard density $p(x)$ with full support $\mathbb{R}$ is

$$I(\lambda) = \frac{1}{s^2} \left[\begin{array}{cc} a^2 & c \\ c & b^2 \end{array} \right],$$

where

$$a^2 = E_p\left[\left(\frac{p'(x)}{p(x)}\right)^2\right], \tag{2}$$

$$b^2 = E_p\left[\left(1 + x\frac{p'(x)}{p(x)}\right)^2\right], \tag{3}$$

$$c = E_p\left[\frac{p'(x)}{p(x)}\left(1 + x\frac{p'(x)}{p(x)}\right)\right]. \tag{4}$$

When the standard density is even (i.e., $p(x) = p(-x)$), we get a diagonal Fisher matrix that can reparameterize with

$$\theta(\lambda) = \left(\frac{a}{b}\lambda_1, \lambda_2\right)$$

so that the Fisher matrix with respect to θ becomes

$$I_\theta(\theta) = \frac{b^2}{\theta_2^2} \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right].$$

It follows that the Fisher-Rao geometry is hyperbolic with curvature $\kappa = -\frac{1}{b^2} < 0$, and that the Fisher-Rao distance is

$$\rho_p(\lambda_1, \lambda_2) = b\ \rho_U\left(\left(\frac{a}{b}l_1, s_1\right), \left(\frac{a}{b}l_2, s_2\right)\right)$$

where

$$\rho_U(\theta_1, \theta_2) = \mathrm{arccosh}\left(1 + \chi(\theta_1, \theta_2)\right),$$

where $\mathrm{arccosh}(u) = \log(u + \sqrt{u^2 - 1})$ for $u > 1$.

For the Cauchy family, we have $a^2 = b^2 = \frac{1}{2}$ (curvature $\kappa = -\frac{1}{b^2} = -2$) and the Fisher-Rao distance is

$$\rho_{\mathrm{FR}}(p_{\lambda_1} : p_{\lambda_2}) = \frac{1}{\sqrt{2}} \mathrm{arccosh}(1 + \chi(\lambda_1, \lambda_2)).$$

Notice that if we let $\theta = l + is$ then the metric in the complex upper plane $\mathbb{H}$ is $\frac{|\mathrm{d}\theta|^2}{\mathrm{Im}(\theta)^2}$ where $|x + iy| = \sqrt{x^2 + y^2}$ denotes the complex modulus, and $\theta \in \mathbb{H} := \{x + iy\ :\ x \in \mathbb{R}, y \in \mathbb{R}_{++}\}$.

The information geometry of the wrapped Cauchy family is investigated in [3]. Goto and Umeno [8] regards the Cauchy distribution as an invariant measure of the generalized Boole transforms. The Boole transform $\frac{1}{2}\left(X - \frac{1}{X}\right)$ of a standard Cauchy random variable X yields a standard Cauchy random variable. See [13] for a description of the functions preserving Cauchy distributions.

3 f-divergences between univariate Cauchy distributions

Consider the location-scale non-abelian group LS(2) which can be represented as a matrix group [18]. A group element $g_{l,s}$ is represented by a matrix element $M_{l,s} = \begin{bmatrix} s & l \\ 0 & 1 \end{bmatrix}$ for $(l, s) \in \mathbb{R} \times \mathbb{R}_{++}$. The group operation $g_{l_{12},s_{12}} = g_{l_1,s_1} \times g_{l_2,s_2}$ corresponds to a matrix multiplication $M_{l_{12},s_{12}} = M_{l_1,s_1} \times M_{l_2,s_2}$ (with the group identity element $g_{0,1}$ being the matrix identity). A location-scale family is defined by the action of the location-group on a standard density $p(x) = p_{0,1}(x)$. That is, density $p_{l,s}(x) = g_{l,s}.p(x)$ where '.' denotes the action. We have the following invariance for the f-divergences between any two densities of a location-scale family [18] (including the Cauchy family):

$$I_f(g.p_{l_1,s_1} : g.p_{l_2,s_2}) = I_f(p_{l_1,s_1} : p_{l_2,s_2}), \forall g \in \mathrm{LS}(2).$$

Thus we have

$$I_f(p_{l_1,s_1} : p_{l_2,s_2}) = I_f\left(p : p_{\frac{l_2-l_1}{s_1},\frac{s_2}{s_1}}\right) = I_f\left(p_{\frac{l_1-l_2}{s_2},\frac{s_1}{s_2}} : p\right).$$

Therefore, we may always consider the calculation of the f-divergence between the standard density and another density of the location-scale family. For example, we check that

$$\chi((l_1, s_1), (l_2, s_2)) = \chi\left((0,1), \left(\frac{l_2-l_1}{s_1}, \frac{s_2}{s_1}\right)\right)$$

since $\chi((0,1),(l,s)) = \frac{(s-1)^2+l^2}{2s}$. If we assume that the standard density p is such that $E_p[X] = \int xp(x)\mathrm{d}x = 0$ and $E_p[X^2] = \int x^2 p(x)\mathrm{d}x = 1$ (hence unit variance), then the random variable $Y = \mu + \sigma X$ has mean $E[Y] = \mu$ and standard deviation $\sigma(Y) = \sqrt{E[(Y-\mu)^2]} = \sigma$. However, the expectation and variance of Cauchy distributions are not defined, hence we preferred (l, s) parameterization over the (μ, σ^2) parameterization, where l denotes the median and s the probable error for the Cauchy location-scale family [14].

3.1 f-divergences between Cauchy distributions are symmetric

Let $\|\lambda\| = \sqrt{\lambda_1^2 + \lambda_2^2}$ denote the Euclidean norm of a 2D vector $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$. We state the main theorem:

Theorem 1 *All f-divergences between univariate Cauchy distributions p_λ and $p_{\lambda'}$ with $\lambda = (l, s)$ and $\lambda' = (l', s')$ are symmetric and can be expressed as*

$$I_f(p_\lambda : p_{\lambda'}) = h_f\left(\chi(\lambda, \lambda')\right)$$

where

$$\chi(\lambda, \lambda') := \frac{\|\lambda - \lambda'\|^2}{2\lambda_2\lambda_2'}$$

and $h_f : \mathbb{R}_+ \to \mathbb{R}_+$ is a function (with $h_f(0) = 0$).

The proof does not yield explicit closed-form formula for the f-divergences as it can be in general difficult (e.g., the Jensen-Shannon divergence [7]), and relies on McCullagh's complex parametrization [14] p_θ of the parameter of the Cauchy density $p_{l,s}$ with $\theta = l + is$:

$$p_\theta(x) = \frac{\mathrm{Im}(\theta)}{\pi|x-\theta|^2}.$$

We make use of the special linear group $\mathrm{SL}(2,\mathbb{R})$ for θ the complex parameter:

$$\mathrm{SL}(2,\mathbb{R}) := \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \ : \ a,b,c,d \in \mathbb{R}, ad-bc=1 \right\}.$$

Let $A.\theta := \frac{a\theta+b}{c\theta+d}$ (real linear fractional transformations) be the action of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{R})$. McCullagh proved that if $X \sim \mathrm{Cauchy}(\theta)$ then $A.X \sim \mathrm{Cauchy}\,(A.\theta)$. We can also define an action of $\mathrm{SL}(2,\mathbb{R})$ to the real line $\mathbb{R}$ by $x \mapsto \frac{ax+b}{cx+d}, \quad x \in \mathbb{R}$, where we interpret $-\frac{d}{c} \mapsto \frac{a}{c}$ if $c \neq 0$. We remark that $d \neq 0$ if $c = 0$. This map is bijective between $\mathbb{R}$. We have the following invariance:

Lemma 1 (Invariance of Cauchy f-divergence under $\mathrm{SL}(2,\mathbb{R})$) *For any $A \in \mathrm{SL}(2,\mathbb{R})$ and $\theta \in \mathbb{H}$, we have*

$$I_f(p_{A.\theta_1} : p_{A.\theta_2}) = I_f(p_{\theta_1} : p_{\theta_2}).$$

We prove the invariance by the change of variable in the integral. Let $D(\theta_1 : \theta_2) := I_f\,(p_{\theta_1} : p_{\theta_2})$. We have

$$D(A.\theta_1 : A.\theta_2) = \int_{\mathbb{R}} \frac{\mathrm{Im}(A.\theta_1)}{\pi|x-A.\theta_1|^2} f\left(\frac{\mathrm{Im}(A.\theta_2)|x-A.\theta_1|^2}{\mathrm{Im}(A.\theta_1)|x-A.\theta_2|^2}\right) \mathrm{d}x.$$

Since $A \in \mathrm{SL}(2,\mathbb{R})$, we have

$$\mathrm{Im}(A.\theta_i) = \frac{\mathrm{Im}(\theta_i)}{|c\theta_i+d|^2}, \quad i \in \{1,2\}.$$

If $x = A.y$ then $\mathrm{d}x = \frac{\mathrm{d}y}{|cy+d|^2}$, and

$$|A.y - A.\theta_i|^2 = \frac{|y-\theta_i|^2}{|cy+d|^2\ |c\theta_i+d|^2}, \quad i \in \{1,2\}.$$

Hence we get:

$$\begin{aligned} \int_{\mathbb{R}} f\left(\frac{\mathrm{Im}(A.\theta_2)|x-A.\theta_1|^2}{\mathrm{Im}(A.\theta_1)|x-A.\theta_2|^2}\right)\frac{\mathrm{Im}(A.\theta_2)}{\pi|x-A.\theta_1|^2}\mathrm{d}x &= \int_{\mathbb{R}} f\left(\frac{\mathrm{Im}(\theta_2)|y-\theta_1|^2}{\mathrm{Im}(\theta_1)|y-\theta_2|^2}\right)\frac{\mathrm{Im}(\theta_2)}{\pi|y-\theta_2|^2}\mathrm{d}y, \\ &= I_f\,(p_{\theta_1} : p_{\theta_2}). \end{aligned}$$

QED.

Let us notice that the Cauchy family is the only univariate location-scale family that is also closed by inversion [11]: That is, if $X \sim \mathrm{Cauchy}(l,s)$ then $\frac{1}{X} \sim \mathrm{Cauchy}(l',s')$. Therefore our results are specific to the Cauchy family and not to any other location-scale family.

We now prove Theorem 1 using the notion of maximal invariants of Eaton [6] (Chapter 2) that will be discussed in §3.2.

Let us rewrite the function χ with complex arguments as:

$$\chi(z,w) := \frac{|z-w|^2}{2\,\mathrm{Im}(z)\mathrm{Im}(w)}, \quad z,w \in \mathbb{C}. \tag{5}$$

Proposition 1 (McCullagh [14]) *The function* χ *defined in Eq. 5 is a* maximal invariant *for the action of the special linear group* $\mathrm{SL}(2,\mathbb{R})$ *to* $\mathbb{H}\times\mathbb{H}$ *defined by*

$$A.(z,w) := \left(\frac{az+b}{cz+d}, \frac{aw+b}{cw+d}\right), \quad A = \left[\begin{array}{cc} a & b \\ c & d \end{array}\right] \in \mathrm{SL}(2,\mathbb{R}),\ z,w \in \mathbb{H}.$$

That is, we have

$$\chi(A.z, A.w) = \chi(z,w), \quad A \in \mathrm{SL}(2,\mathbb{R}),\ z,w \in \mathbb{H},$$

and it holds that for every $z,w,z',w' \in \mathbb{H}$ *satisfying that* $\chi(z',w') = \chi(z,w)$*, there exists* $A \in \mathrm{SL}(2,\mathbb{R})$ *such that* $(A.z, A.w) = (z', w')$*.*

By Lemma 1 and Theorem 2.3 of [6], there exists a unique function $h_f : [0,\infty) \to [0,\infty)$ such that $h_f(\chi(z,w)) = D(z,w)$ for all $z,w \in \mathbb{H}$.

Theorem 2 *The* f*-divergence between two univariate Cauchy densities is symmetric and expressed as a function of the chi-squared divergence:*

$$I_f(p_{\theta_1} : p_{\theta_2}) = I_f(p_{\theta_2} : p_{\theta_1}) = h_f(\chi(\theta_1,\theta_1)), \quad \theta_1,\theta_2 \in \mathbb{H}. \tag{6}$$

Therefore we have proven that the f-divergences between univariate Cauchy densities are all symmetric. For example, we have

$$h_{\mathrm{KL}}(u) = \log\left(1 + \frac{1}{2}u\right).$$

Let us consider another illustrating example: The LeCam triangular divergence [12] defined by

$$D_{\mathrm{LC}}[p:q] := \int \frac{(p(x)-q(x))^2}{p(x)+q(x)}\mathrm{d}x.$$

This divergence is a symmetric f-divergence obtained for the generator $f_{\mathrm{LC}}(u) = \frac{(u-1)^2}{1+u}$. The triangular divergence is a bounded divergence since $f(0) = f^*(0) = 1 < \infty$, and its square root $\sqrt{D_{\mathrm{LC}}[p:q]}$ yields a metric distance. The LeCam triangular divergence between a Cauchy standard density $p_{0,1}$ and a Cauchy density $p_{l,s}$ is

$$D_{\mathrm{LC}}[p_{0,1} : p_{l,s}] = 2 - 4\sqrt{\frac{s}{l^2+s^2+2s+1}} \leq 2.$$

Since $\chi[p_{0,1} : p_{l,s}] = \frac{l^2+(s-1)^2}{2s}$, we can express the triangular divergence using the χ-squared divergence as

$$D_{\mathrm{LC}}[p_{l_1,s_1}, p_{l_2,s_2}] = 2 - 4\sqrt{\frac{1}{2(\chi[p_{l_1,s_1}, p_{l_2,s_2}]+2)}}.$$

Thus we have the function:

$$h_{f_{\mathrm{LC}}}(u) = 2 - 4\sqrt{\frac{1}{2(u+2)}}.$$

Note that since $I_f(p_{\theta_2} : p_{\theta_1}) = h_f(\chi(\theta_1,\theta_1)$, Lemma 1 can *a posteriori* be checked for the chi-squared divergence: For any $A \in \mathrm{SL}(2,\mathbb{R})$ and $\theta \in \mathbb{H}$, we have

$$\chi(p_{A.\theta_1} : p_{A.\theta_2}) = \chi(p_{\theta_1} : p_{\theta_2}),$$

and therefore for any f-divergence, since we have $I_f(p_{A.\theta_1}:p_{A.\theta_2})=I_f(p_{\theta_1}:p_{\theta_2})$ since

$$I_f(p_{A.\theta_2}:p_{A.\theta_1})=h_f(\chi(A.\theta_1,A.\theta_1))=h_f(\chi(\theta_1,\theta_1))=I_f(p_{\theta_2}:p_{\theta_1}).$$

To prove that $\chi(p_{A.\theta_1}:p_{A.\theta_2})=\chi(p_{\theta_1}:p_{\theta_2})$, let us first recall that $\mathrm{Im}(A.\theta)=\frac{\mathrm{Im}(\theta)}{|c\theta+d|^2}$ and $|A.\theta_1-A.\theta_2|^2=\frac{|\theta_1-\theta_2|^2}{|c\theta_1+d|^2\ |c\theta_2+d|^2}$. Thus we have

$$\begin{aligned}
\chi(A.\theta_1,A.\theta_2) &= \frac{|A.\theta_1-A.\theta_2|^2}{2\,\mathrm{Im}(A.\theta_1)\mathrm{Im}(A.\theta_2)},\\
&= \frac{|\theta_1-\theta_2|^2|c\theta_1+d|^2\ |c\theta_2+d|^2}{|c\theta_1+d|^2\ |c\theta_2+d|^2\ 2\,\mathrm{Im}(\theta_1)\mathrm{Im}(\theta_2)},\\
&= \frac{|\theta_1-\theta_2|^2}{2\,\mathrm{Im}(\theta_1)\mathrm{Im}(\theta_2)}=\chi(\theta_1,\theta_2).
\end{aligned}$$

Alternatively, we may also define a bivariate function $g_f(l,s)$ so that using the action of the location-scale group, we have:

$$h_f(\chi(\theta_1,\theta_2))=g_f\left(\frac{l_1-l_2}{s_2},\frac{s_1}{s_2}\right),$$

where $\theta_1=l_1+is_1$ and $\theta_2=l_2+is_2$. When the function h_f is not explicitly known, we may estimate the f-divergences using Monte Carlo importance samplings [18].

3.2 Maximal invariants (proof of Proposition 1)

First, let us show that

Lemma 2 *For every $(z,w)\in\mathbb{H}^2$, there exist $\lambda\geq 1$ and $A\in\mathrm{SL}(2,\mathbb{R})$ such that $(A.z,A.w)=(\lambda i,i)$.*

Since the special orthogonal group $\mathrm{SO}(2,\mathbb{R})$ is the isotropy subgroup of $\mathrm{SL}(2,\mathbb{R})$ for i and the action is transitive, it suffices to show that for every $z\in\mathbb{H}$ there exist $\lambda\geq 1$ and $A\in\mathrm{SO}(2,\mathbb{R})$ such that $\lambda i=A.z$.

Since we have that for every $\lambda>0$,

$$\begin{bmatrix} 0 & -1\\ 1 & 0\end{bmatrix}.\lambda i=\frac{i}{\lambda},$$

it suffices to show that for every $z\in\mathbb{H}$ there exist $\lambda>0$ and $A\in\mathrm{SO}(2,\mathbb{R})$ such that $\lambda i=A.z$.

We have that

$$\begin{bmatrix} \cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}.z=\frac{\frac{|z|^2-1}{2}\sin 2\theta+\mathrm{Re}(z)\cos 2\theta+i\mathrm{Im}(z)}{|z\sin\theta+\cos\theta|^2},$$

Therefore for some θ, we have

$$\frac{|z|^2-1}{2}\sin 2\theta+\mathrm{Re}(z)\cos 2\theta=0.$$

QED.

By this lemma, we have that for some $\lambda, \lambda' \geq 1$ and $A, A' \in \mathrm{SL}(2, \mathbb{R})$,

$$(\lambda i, i) = (A.z, A.w),\ (\lambda' i, i) = (A'.z', A'.w'),$$

$$F(z, w) = F(\lambda i, i) = \frac{(\lambda-1)^2}{4\lambda} = \frac{1}{4}\left(\lambda + \frac{1}{\lambda} - 2\right),$$

and

$$F(z', w') = F(\lambda' i, i) = \frac{(\lambda'-1)^2}{4\lambda'} = \frac{1}{4}(\lambda' + \frac{1}{\lambda'} - 2).$$

If $F(z', w') = F(z, w)$, then, $\lambda = \lambda'$ and hence $(A.z, A.w) = (A'.z', A'.w')$. QED.

4 Asymmetric KLD between multivariate Cauchy distributions

The probability density function of a d-dimensional Cauchy distribution with parameters $\mu \in \mathbb{R}^d$ and Σ be a $d \times d$ positive-definite symmetric matrix is defined by:

$$p_{\mu,\Sigma}(x) := \frac{C_d}{(\det \Sigma)^{1/2}} \left(1 + (x-\mu)' \Sigma^{-1} (x-\mu)\right)^{-(d+1)/2},\ x \in \mathbb{R}^d,$$

where C_d is a normalizing constant. Contrary to the univariate Cauchy distribution, we have the following:

Proposition 2 *There exist two bivariate Cauchy densities p_{μ_1,Σ_1} and p_{μ_2,Σ_2} such that $D_{\mathrm{KL}}\left(p_{\mu_1,\Sigma_1} : p_{\mu_2,\Sigma_2}\right) \neq D_{\mathrm{KL}}\left(p_{\mu_2,\Sigma_2} : p_{\mu_1,\Sigma_1}\right)$.*

We let $d = 2$. By the change of variable in the integral [18], we have

$$D_{\mathrm{KL}}\left(p_{\mu_1,\Sigma_1} : p_{\mu_2,\Sigma_2}\right) = D_{\mathrm{KL}}\left(p_{0,I_2} : p_{\Sigma_1^{-1/2}(\mu_2-\mu_1), \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}}\right),$$

where I_2 denotes the unit 2×2 matrix.

Let

$$\mu_1 = 0, \Sigma_1 = I_2,\ \mu_2 = (0,1)^\top, \Sigma_2 = \begin{bmatrix} n & 0 \\ 0 & \frac{1}{n} \end{bmatrix},$$

where n is a natural number. We will show that $D_{\mathrm{KL}}\left(p_{\mu_1,\Sigma_1} : p_{\mu_2,\Sigma_2}\right) \neq D_{\mathrm{KL}}\left(p_{\mu_2,\Sigma_2} : p_{\mu_1,\Sigma_1}\right)$ for sufficiently large n. Then,

$$D_{\mathrm{KL}}\left(p_{\mu_1,\Sigma_1} : p_{\mu_2,\Sigma_2}\right) = \frac{3C_2}{2} \int_{\mathbb{R}^2} \frac{\log(1 + x_1^2/n + n x_2^2) - \log(1 + x_1^2 + x_2^2)}{(1 + x_1^2 + x_2^2)^{3/2}} \mathrm{d}x_1 \mathrm{d}x_2$$

and

$$\begin{aligned} D_{\mathrm{KL}}\left(p_{\mu_2,\Sigma_2} : p_{\mu_1,\Sigma_1}\right) &= D_{\mathrm{KL}}\left(p_{0,I_2} : p_{-\Sigma_1^{-1/2}\mu_1, \Sigma_1^{-1}}\right), \\ &= \frac{3C_2}{2} \int_{\mathbb{R}^2} \frac{\log(1 + x_1^2/n + n(x_2 + \sqrt{n})^2) - \log(1 + x_1^2 + x_2^2)}{(1 + x_1^2 + x_2^2)^{3/2}} \mathrm{d}x_1 \mathrm{d}x_2. \end{aligned}$$

Hence it suffices to show that

$$\int_{\mathbb{R}^2} \frac{\log(1 + x_1^2/n + n(x_2 + \sqrt{n})^2) - \log(1 + x_1^2/n + nx_2^2)}{(1 + x_1^2 + x_2^2)^{3/2}} \mathrm{d}x_1 \mathrm{d}x_2 \neq 0$$

for some n.

Since $\log(1 + x) \leq x$, it suffices to show that

$$\int_{\mathbb{R}^2} \frac{-n^2 + n - 2x_2(n + \sqrt{n})}{(1 + x_1^2 + x_2^2)^{3/2}(1 + x_1^2/n + n(x_2 + \sqrt{n})^2)} \mathrm{d}x_1 \mathrm{d}x_2 < 0$$

for some n.

$$\begin{aligned}&\int_{\mathbb{R}^2} \frac{-2x_2(n + \sqrt{n})}{(1 + x_1^2 + x_2^2)^{3/2}(1 + x_1^2/n + n(x_2 + \sqrt{n})^2)} \mathrm{d}x_1 \mathrm{d}x_2, \\ &\leq 4n \int_{\mathbb{R}^2} \frac{|x_2|}{(1 + x_1^2 + x_2^2)^{3/2}(1 + x_1^2/n + n(x_2 + \sqrt{n})^2)} \mathrm{d}x_1 \mathrm{d}x_2.\end{aligned}$$

We have that

$$\begin{aligned}&n \int_{|x_2 + \sqrt{n}| > n^{1/3}} \frac{|x_2|}{(1 + x_1^2 + x_2^2)^{3/2}(1 + x_1^2/n + n(x_2 + \sqrt{n})^2)} \mathrm{d}x_1 \mathrm{d}x_2, \\ &\leq \int_{|x_2 + \sqrt{n}| > n^{1/3}} \frac{|x_2|}{(1 + x_1^2 + x_2^2)^{3/2}(x_2 + \sqrt{n})^2} dx_1 dx_2.\end{aligned}$$

We have that $\limsup_{n\to\infty} \sup_{x;|x+\sqrt{n}|>n^{1/3}} \frac{|x|}{(x+\sqrt{n})^2} = 0$.

Hence,

$$\limsup_{n\to\infty} n \int_{|x_2 + \sqrt{n}| > n^{1/3}} \frac{|x_2|}{(1 + x_1^2 + x_2^2)^{3/2}(1 + x_1^2/n + n(x_2 + \sqrt{n})^2)} \mathrm{d}x_1 \mathrm{d}x_2 = 0.$$

We have that by Fatou's lemma [10] (p. 93),

$$\begin{aligned}&\liminf_{n\to\infty} \int_{|x_2 + \sqrt{n}| > n^{1/3}} \frac{n^2 - n}{(1 + x_1^2 + x_2^2)^{3/2}(1 + x_1^2/n + n(x_2 + \sqrt{n})^2)} \mathrm{d}x_1 \mathrm{d}x_2 \\ &\geq \liminf_{n\to\infty} \int_{x_1^2 + x_2^2 \leq 1} \frac{n^2 - n}{(1 + x_1^2 + x_2^2)^{3/2}(1 + x_1^2 + 2n(x_2^2 + n))} \mathrm{d}x_1 \mathrm{d}x_2 \\ &\geq \frac{1}{2} \int_{x_1^2 + x_2^2 \leq 1} \frac{1}{(1 + x_1^2 + x_2^2)^{3/2}} \mathrm{d}x_1 \mathrm{d}x_2 > 0.\end{aligned}$$

Hence,

$$\begin{aligned}&\liminf_{n\to\infty} \int_{|x_2 + \sqrt{n}| > n^{1/3}} \frac{n^2 - n - 4|x_2|n}{(1 + x_1^2 + x_2^2)^{3/2}(1 + x_1^2/n + n(x_2 + \sqrt{n})^2)} dx_1 dx_2 \\ &\geq \frac{1}{2} \int_{x_1^2 + x_2^2 \leq 1} \frac{1}{(1 + x_1^2 + x_2^2)^{3/2}} \mathrm{d}x_1 \mathrm{d}x_2 > 0.\end{aligned}$$

We have that for large n,

$$\int_{|x_2+\sqrt{n}|\le n^{1/3}} \frac{n^2-n-4|x_2|n}{(1+x_1^2+x_2^2)^{3/2}(1+x_1^2/n+n(x_2+\sqrt{n})^2)}\mathrm{d}x_1\mathrm{d}x_2 \ge 0.$$

Thus we have that

$$\liminf_{n\to\infty}\int_{\mathbb{R}^2} \frac{n^2-n-4|x_2|n}{(1+x_1^2+x_2^2)^{3/2}(1+x_1^2/n+n(x_2+\sqrt{n})^2)}\mathrm{d}x_1\mathrm{d}x_2 > 0.$$

QED.

5 Revisiting the KLD between Cauchy densities

We shall prove the following result [4] using complex analysis:

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

$$D_{\mathrm{KL}}(p_{l_1,s_1}:p_{l_2,s_2}) = \frac{s_1}{\pi}\int_{\mathbb{R}}\frac{\log((z-l_2)^2+s_2^2)}{(z-l_1)^2+s_1^2}\mathrm{d}z$$
$$-\frac{s_1}{\pi}\int_{\mathbb{R}}\frac{\log((z-l_1)^2+s_1^2)}{(z-l_1)^2+s_1^2}\mathrm{d}z+\log\frac{s_1}{s_2}. \tag{7}$$

As a function of z,

$$\frac{\log(z-l_2+is_2)}{z-l_1+is_1}$$

is holomorphic on the upper-half plane $\{x+yi : y>0\}$. By the Cauchy integral formula [16], we have that for sufficiently large R,

$$\frac{1}{2\pi i}\int_{C_R^+}\frac{\log(z-l_2+is_2)}{(z-l_1)^2+s_1^2}\mathrm{d}z = \frac{\log(l_1-l_2+i(s_2+s_1))}{2s_1 i},$$

where

$$C_R^+ := \{z : |z| = R, \mathrm{Im}(z)>0\}\cup\{z : \mathrm{Im}(z)=0, |\mathrm{Re}(z)|\le R\}.$$

Hence, by $R\to+\infty$, we get

$$\frac{s_1}{\pi}\int_{\mathbb{R}}\frac{\log(z-l_2+is_2)}{(z-l_1)^2+s_1^2}\mathrm{d}z = \log(l_1-l_2+i(s_2+s_1)). \tag{8}$$

As a function of z,

$$\frac{\log(z-l_2-is_2)}{z-l_1-is_1}$$

is holomorphic on the lower-half plane $\{x+yi : y<0\}$. By the Cauchy integral formula again, we have that for sufficiently large R,

$$\frac{1}{2\pi i}\int_{C_R^-}\frac{\log(z-l_2-is_2)}{(z-l_1)^2+s_1^2}\mathrm{d}z = \frac{\log(l_1-l_2-i(s_2+s_1))}{-2s_1 i},$$

where

$$C_R^- := \{z : |z| = R, \mathrm{Im}(z) < 0\} \cup \{z : \mathrm{Im}(z) = 0, |\mathrm{Re}(z)| \leq R\}.$$

Hence, by $R \to +\infty$, we get

$$\frac{s_1}{\pi} \int_{\mathbb{R}} \frac{\log(z - l_2 - is_2)}{(z - l_1)^2 + s_1^2} \mathrm{d}z = \log(l_1 - l_2 - i(s_2 + s_1)). \tag{9}$$

By Eq. 8 and Eq. 9, we have that

$$\frac{s_1}{\pi} \int_{\mathbb{R}} \frac{\log((z - l_2)^2 + s_2^2)}{(z - l_1)^2 + s_1^2} \mathrm{d}z = \log\left((l_1 - l_2)^2 + (s_1 + s_2)^2\right). \tag{10}$$

In the same manner, we have that

$$\frac{s_1}{\pi} \int_{\mathbb{R}} \frac{\log((z - l_1)^2 + s_1^2)}{(z - l_1)^2 + s_1^2} \mathrm{d}z = \log(4s_1^2). \tag{11}$$

By substituting Eq. 10 and Eq. 11 into Eq. 7, we obtain the formula Eq. 1. QED.

6 Revisiting the chi-squared divergence between Cauchy densities

Proposition 3

$$D_\chi^N(p_{l_1,s_1} : p_{l_2,s_2}) = \frac{(l_1 - l_2)^2 + (s_1 - s_2)^2}{2s_1 s_2}. \tag{12}$$

We first remark that

$$D_\chi^N(p_{l_1,s_1} : p_{l_2,s_2}) = \int_{\mathbb{R}} \frac{p_{l_2,s_2}^2(x)}{p_{l_1,s_1}(x)} \mathrm{d}x - 1$$

Let $F(z) := \frac{(z-l_1)^2+s_1^2}{(z-l_2+is_2)^2}$. Then, this is holomorphic on the upper-half plane $\mathbb{H}$, and,

$$\frac{p_{l_2,s_2}(x)^2}{p_{l_1,s_1}(x)} = \frac{s_2^2}{\pi s_1} \frac{F(x)}{(x - l_2 - is_2)^2}.$$

By the Cauchy integral formula [16], we have that for sufficiently large R,

$$\frac{1}{2\pi i} \int_{C_R^+} \frac{F(z)}{(z - l_2 - is_2)^2} \mathrm{d}z = F'(l_2 + is_2),$$

where $C_R^+ := \{z : |z| = R, \mathrm{Im}(z) > 0\} \cup \{z : \mathrm{Im}(z) = 0, |\mathrm{Re}(z)| \leq R\}$.

Since

$$F'(z) = 2\frac{(z - s_1)(z - l_2 + is_2) - (z - l_1)^2 - s_1^2}{(z - l_2 + is_2)^3},$$

we have that

$$\int_{C_R^+} \frac{F(z)}{(z - l_2 - is_2)^2} \mathrm{d}z = \frac{\pi}{2} \frac{(l_1 - l_2)^2 + s_1^2 + s_2^2}{s_2^3}.$$

Now, by $R \to \infty$, we obtain the formula Eq. 12. QED.

7 Metrization of f-divergences between Cauchy densities

7.1 Metrization of the KLD between Cauchy densities

Recall that f-divergences can always be symmetrized by taking the generator $s(u) = f(u)+uf(1/u)$. Metrizing f divergences consists in finding the largest exponent α such that I_s^α is a metric distance satisfying the triangle inequality [9, 23, 24]. For example, the square root of the Jensen-Shannon divergence [7] yields a metric distance which is moreover Hilbertian [1], i.e., meaning that there is an embedding $\phi(\cdot)$ into a Hilbert space $\mathcal{H}$ such that $D_{\mathrm{JS}}(p:q) = \|\phi(p)-\phi(q)\|_{\mathcal{H}}$.

Proposition 4 *Let $d(\theta_1,\theta_2) := D_{\mathrm{KL}}(p_{\theta_1} : p_{\theta_2})^\alpha$ for $0<\alpha\leq 1$. Then d is a metric on $\mathbb{H}$ if and only if $0<\alpha\leq 1/2$.*

We proceed as in [17] by letting

$$t(u) := \log\left(\frac{1+\cosh(\sqrt{2}u)}{2}\right), u \geq 0.$$

Let us consider the properties of $F(u) := t(u)^\alpha/u$.

$$F'(u) = -2\frac{t(u)^{\alpha-1}}{u^2}G(u/\sqrt{2}),$$

where

$$G(w) := (2+e^{2w}+e^{-2w})\log\left(\frac{e^w+e^{-w}}{2}\right) - \alpha w(e^{2w}-e^{-2w}).$$

If we let $x := e^w$, then,

$$G(w) = (x+x^{-1})\left((x+x^{-1})\log(\frac{x^2+1}{2x}) - \alpha(x-x^{-1})\log x\right).$$

Let

$$H(x) := x\left((x+x^{-1})\log(\frac{x^2+1}{2x}) - \alpha(x-x^{-1})\log x\right).$$

Then, $H(1)=0$ and

$$H'(x) = 4\left(x\log(\frac{x^2+1}{2}) - (1+\alpha)x\log x + \frac{x^3}{x^2+1} - \alpha x\right).$$

Let

$$I(x) := x\log(\frac{x^2+1}{2}) - (1+\alpha)x\log x + \frac{x^3}{x^2+1} - \alpha x.$$

Then, $I(1) = 1/2-\alpha$ and

$$I'(x) = \log(\frac{x^2+1}{2}) - (1+\alpha)\log x + \frac{x^2(3x^2+5)}{(x^2+1)^2} - (1+2\alpha).$$

Consider the case that $\alpha > 1/2$. Then, $I(x) < 0$ for every $x > 1$ which is sufficiently close to 1. Hence, $G(w) < 0$ for every $w > 0$ which is sufficiently close to 0. Hence, $F'(u) > 0$ for every $u > 0$ which is sufficiently close to 0. This means that F is strictly increasing near the origin.

Hence there exists $u_0 > 0$ such that

$$2t(u_0)^\alpha < t(2u_0)^\alpha.$$

Take $x_0, z_0 \in \mathbb{H}$ such that $\rho_{\mathrm{FR}}(x_0, z_0) = 2u_0$, where ρ_{FR} is the Fisher metric distance on $\mathbb{H}$. By considering the geodesic between x_0 and z_0, we can take $y_0 \in \mathbb{H}$ such that $\rho_{\mathrm{FR}}(x_0, y_0) = \rho_{\mathrm{FR}}(y_0, z_0) = u_0$.

Finally we consider the case that $\alpha = 1/2$. Let

$$J(x) := (x^2+1)^2 \log(\frac{x^2+1}{2}) - \frac{3}{2}(x^2+1)^2 \log x + x^2(3x^2+5) - 2(x^2+1)^2.$$

Then, $J(1) = 0$. If we let $y := x^2$, then,

$$J(x) = (y+1)^2 \log\left(\frac{y+1}{2}\right) - \frac{3}{4}(y+1)^2 \log y + (y^2+y-2).$$

Let $K(y) := J(\sqrt{y})$. Then,

$$\begin{aligned} K'(y) &= 2(y+1)(\log(\frac{y+1}{2}) + 1) - \frac{3}{2}(y+1)\log y - \frac{3(y+1)^2}{4y} + (2y+1), \\ &= y + (y+1)\left(2\log(y+1) - \frac{3}{2}\log y + \frac{9}{4} - \frac{3}{4y} - 2\log 2\right). \end{aligned}$$

If $y > 1$, then,

$$2\log(y+1) > \frac{3}{2}\log y$$

and

$$\frac{9}{4} - \frac{3}{4y} - 2\log 2 > \frac{3}{2} - 2\log 2 > 0.$$

Then, $J(x) > J(1) = 0$ for every $x > 1$. Hence, $I(x) > I(1) = 0$ for every $x > 1$. Hence, $G(w) > 0$ for every $w > 0$. Hence, $F'(u) < 0$ for every $u > 0$. This means that F is strictly decreasing on $[0, \infty)$. Thus we proved that $D_{\mathrm{KL}}(p_{\theta_1} : p_{\theta_2})^{1/2}$ gives a distance, hence $D_{\mathrm{KL}}(p_{\theta_1} : p_{\theta_2})^\alpha$ is also a distance for every $\alpha \in (0, 1/2)$. QED.

7.2 Metrization of the Bhattacharyya divergence between Cauchy densities

The Bhattacharyya divergence [2] is defined by

$$D_{\mathrm{Bhat}}[p, q] := -\log \int \sqrt{p(x)q(x)}\mathrm{d}x.$$

It is easy to see that $D_{\mathrm{Bhat}}[p, q] = 0$ iff $p = q$, and $D_{\mathrm{Bhat}}[p, q] = D_{\mathrm{Bhat}}[q, p]$.

Proposition 5 *Let $\{p_\theta\}_{\theta \in \mathbb{H}}$ be the family of the univariate Cauchy distributions. Then, $\sqrt{D_{\mathrm{Bhat}}[p_{\theta_1}, p_{\theta_2}]}$ is a distance.*

For exponential families, see Prop. 2 [17, 19]. We can also show that $D_{\mathrm{Bhat}}[p_{\theta_1}, p_{\theta_2}]^\alpha$ is not a metric if $\alpha > 1/2$ in the same manner as in the proof of Proposition 4.

We show the triangle inequality. We follow the idea in the proof of Theorem 3 in [17]. We construct the metric transform $t_{\mathrm{FR}\to\mathrm{Bhat}}$ and show that $t_{\mathrm{FR}\to\mathrm{Bhat}}(s)$ is increasing and $\sqrt{t_{\mathrm{FR}\to\mathrm{Bhat}}(s)}/s$ is decreasing.

Let

$$\chi(z, w) := \frac{|z-w|^2}{2\mathrm{Im}(z)\mathrm{Im}(w)}.$$

Let ρ_{FR} be the Fisher-Rao distance. Then, by following the argument in the proof of [17, Theorem 3],

$$\chi(z, w) = F(\rho_{\mathrm{FR}}(z, w)),$$

where we let

$$F(s) := \cosh(\sqrt{2}s) - 1.$$

Let

$$I(z, w) := \int \sqrt{p_z(x)p_w(x)}dx.$$

Then, by the invariance of the f-divergences,

$$I(A.z, A.w) = I(z, w).$$

Hence we have that for some function J, $J(\chi(z, w)) = I(z, w)$. Hence,

$$\sqrt{D_{\mathrm{Bhat}}[p_{\theta_1}, p_{\theta_2}]} = \sqrt{-\log J\left(F(\rho_{\mathrm{FR}}(\theta_1, \theta_2))\right)}.$$

We have that

$$t_{\mathrm{FR}\to\mathrm{Bhat}}(s) = -\log J(F(s)).$$

It holds that for every $a \in (0, 1)$,

$$J\left(\chi(ai, i)\right) = I(ai, i).$$

By the change-of-variable $x = \tan\theta$ in the integral of $I(ai, i)$, it is easy to see that

$$I(ai, i) = \frac{2\sqrt{a}K(1-a^2)}{\pi},$$

where K is the elliptic integral of the first kind. It is defined by[1]

$$K(t) := \int_0^{\pi/2} \frac{1}{\sqrt{1 - t\sin^2\theta}}d\theta, \ 0 \le t < 1.$$

Hence,

$$J\left(\frac{(1-a)^2}{2a}\right) = \frac{2\sqrt{a}K(1-a^2)}{\pi}.$$

Since

$$F(s) = \cosh(\sqrt{2}s) - 1 = \frac{(1-e^{-\sqrt{2}s})^2}{2e^{-\sqrt{2}s}},$$

[1] This is a little different from the usual definition.

we have that

$$J(F(s)) = \frac{2e^{-s/\sqrt{2}}K(1-e^{-2\sqrt{2}s})}{\pi}.$$

Since the above function is decreasing with respect to s, $t_{\mathrm{FR}\to\mathrm{Bhat}}(s)$ is increasing.

Furthermore, we have that

$$\frac{\sqrt{t_{\mathrm{FR}\to\mathrm{Bhat}}(s)}}{s} = \sqrt{-\frac{1}{s^2}\log\left(\frac{2e^{-s/\sqrt{2}}K(1-e^{-2\sqrt{2}s})}{\pi}\right)}. \tag{13}$$

This function is decreasing with respect to s. See Figure 1. QED.

It holds that

$$\lim_{s\to+0}\frac{\sqrt{t_{\mathrm{FR}\to\mathrm{Bhat}}(s)}}{s} = \frac{1}{8}, \lim_{s\to+\infty}\frac{\sqrt{t_{\mathrm{FR}\to\mathrm{Bhat}}(s)}}{s} = 0.$$

We remark that the squared Hellinger distance $H^2[p:q] := \frac{1}{2}\int\left(\sqrt{p(x)}-\sqrt{q(x)}\right)\mathrm{d}x$ (a f-divergence for $f_{\mathrm{Hellinger}}(u) = \frac{1}{2}(\sqrt{u}-1)^2$) satisfies that

$$\begin{aligned}
H^2(p_{\theta_1}:p_{\theta_2}) &= 1-\exp\left(-D_{\mathrm{Bhat}}[p_{\theta_1},p_{\theta_2}]\right) = 1-J(F(\rho_{\mathrm{FR}}(\theta_1,\theta_2))) \\
&= 1-\frac{2e^{-\rho_{\mathrm{FR}}(\theta_1,\theta_2)/\sqrt{2}}K(1-e^{-2\sqrt{2}\rho_{\mathrm{FR}}(\theta_1,\theta_2)})}{\pi} \\
&= 1-\frac{2K\left(1-\left(1+\chi(\theta_1,\theta_2)+\sqrt{\chi(\theta_1,\theta_2)(2+\chi(\theta_1,\theta_2))}\right)^{-2}\right)}{\pi\sqrt{1+\chi(\theta_1,\theta_2)+\sqrt{\chi(\theta_1,\theta_2)(2+\chi(\theta_1,\theta_2))}}}.
\end{aligned}$$

The Hellinger distance $H(p_{\theta_1}:p_{\theta_2})$ is known to be a metric distance. Notice that $h_{f_{\mathrm{Hellinger}}}(u) = 1-\frac{2K\left(1-\left(1+u+\sqrt{u(2+u)}\right)^{-2}\right)}{\pi\sqrt{1+u+\sqrt{u(2+u)}}}$ and we check that $h_{f_{\mathrm{Hellinger}}}(0)=0$ since $K(0)=\frac{\pi}{2}$.

In practice, we can calculate efficiently $K(t)$ using the arithmetic–geometric mean (AGM): $K(t) = \frac{\pi}{2\mathrm{AGM}(1,\sqrt{1-t^2})}$ where $\mathrm{AGM}(a,b) = \lim_{n\to\infty}a_n = \lim_{n\to\infty}g_n$ with $a_0=a$, $g_0=b$, $a_{n+1} = \frac{a_n+g_n}{2}$ and $g_{n+1}=\sqrt{a_n g_n}$.

More generally, let $\mathrm{BC}_\alpha[p:q] := \int_{\mathbb{R}} p(x)^\alpha q(x)^{1-\alpha}\mathrm{d}x$ denote the α-skewed Bhattacharyya coefficient for $\alpha\in\mathbb{R}\backslash\{0,1\}$. The α-skewed Bhattacharyya divergence is defined by

$$D_{\mathrm{Bhat},\alpha}[p,q] := -\log\mathrm{BC}_\alpha[p:q] = -\log\int_{\mathbb{R}} p(x)^\alpha q(x)^{1-\alpha}\mathrm{d}x.$$

Using a computer algebra system[2], we can compute the α-skewed Bhattacharyya coefficients for integers α in closed form. For example, we find the following closed-form for the definite integrals:

$$\begin{aligned}
\mathrm{BC}_2[p:p_{l,s}] &= \frac{s^2+l^2+1}{2s}, \\
\mathrm{BC}_3[p:p_{l,s}] &= \frac{3s^4+\left(6l^2+2\right)s^2+3l^4+6l^2+3}{8s^2}, \\
\mathrm{BC}_4[p:p_{l,s}] &= \frac{5s^6+\left(15l^2+3\right)s^4+\left(15l^4+18l^2+3\right)s^2+5l^6+15l^4+15l^2+5}{16s^3}, \\
\mathrm{BC}_5[p:p_{l,s}] &= \frac{35s^8+\left(140l^2+20\right)s^6+\left(210l^4+180l^2+18\right)s^4+\left(140l^6+300l^4+180l^2+20\right)s^2+35l^8+140l^6+210l^4+140l^2+35}{128s^4}
\end{aligned}$$

[2] `https://maxima.sourceforge.io/`

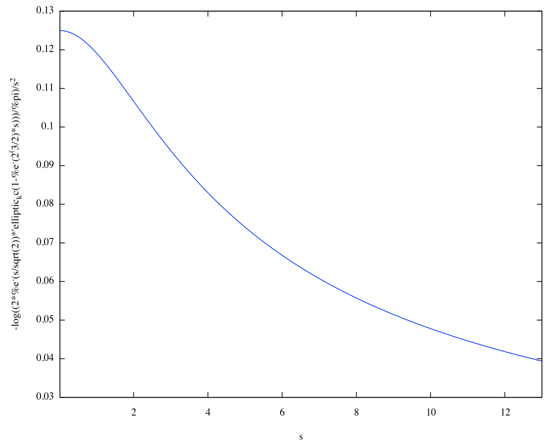

Figure 1: Graph of $\frac{\sqrt{t_{\mathrm{FR}\to\mathrm{Bhat}}(s)}}{s}$

Furthermore, we give some remarks about the complete elliptic integrals of the first and second kinds. Let K and E be the complete elliptic integrals of the first and second kinds respectively. We let[3]

$$E(t) := \int_0^{\pi/2} \sqrt{1 - t\sin^2\theta} d\theta.$$

The following expansion by C. F. Gauss is well-known:

$$1 - \frac{E(x)}{K(x)} = \frac{x}{2} + \sum_{n\geq 1} 2^{n-1}(a_n - b_n)^2, x \in (0, 1),$$

where $(a_0, b_0) = (1, \sqrt{1-x})$ and $(a_{n+1}, b_{n+1}) = \left(\frac{a_n+b_n}{2}, \sqrt{a_n b_n}\right)$, $n \geq 0$. By investigating of the behaviors of $\frac{\sqrt{t_{\mathrm{FR}\to\mathrm{Bhat}}(s)}}{s}$ in Eq. 13, we get some approximation formulae of $1 - \frac{E(x)}{K(x)}$. By numerical computations, it holds that

$$1 - \frac{E(x)}{K(x)} = \frac{x}{2} + \frac{x^2}{16} + \frac{x^3}{32} + \frac{41}{2048}x^4 + \frac{59}{4096}x^5 + \frac{727}{65536}x^6 + O(x^7),$$

$$x\left(\frac{3}{2} + 4\frac{\log(2K(x)/\pi)}{\log(1-x)}\right) = \frac{x}{2} + \frac{x^2}{16} + \frac{x^3}{32} + \frac{251}{12288}x^4 + \frac{123}{8192}x^5 + \frac{34781}{2949120}x^6 + O(x^7)$$

and

$$\frac{x\left(4 - x - \sqrt{(4-3x)^2 + 4(2-x)(1-x)\log(1-x)}\right)}{4x + 2(x-1)\log(1-x)}$$
$$= \frac{x}{2} + \frac{x^2}{16} + \frac{x^3}{32} + \frac{49}{3072}x^4 + \frac{41}{6144}x^5 + \frac{259}{491520}x^6 + O(x^7).$$

They are very close to each other if $x > 0$ is close to 0.

[3] This is also a little different from the usual definition.

f-divergence name	$f(u)$	$h_f(u)$ for $I_f[p_{\lambda_1} : p_{\lambda_2}] = h_f(\chi[p_{\lambda_1} : p_{\lambda_2}])$
Kullback-Leibler divergence	$-\log u$	$\log(1 + \frac{1}{2}u)$
LeCam divergence	$\frac{(u-1)^2}{1+u}$	$2 - 4\sqrt{\frac{1}{2(u+2)}}$
squared Hellinger	$\frac{1}{2}(\sqrt{u} - 1)^2$	$1 - \frac{2K\left(1-\left(1+u+\sqrt{u(2+u)}\right)^{-2}\right)}{\pi\sqrt{1+u+\sqrt{u(2+u)}}}$

Table 1: Closed-form f-divergences between two univariate Cauchy densities expressed as a function h_f of the chi-squared divergence $\chi[p_{\lambda_1} : p_{\lambda_2}]$. The square root of the KLD, LeCam and squared Hellinger divergences between Cauchy densities yields metric distances.

Table 1 summarizes the symmetric closed-form f-divergences $I_f[p_\lambda : p_{\lambda'}] = h_f(\chi[p_\lambda : p_{\lambda'}])$ between two univariate Cauchy densities p_λ and $p_{\lambda'}$ that we obtained as a function h_f of the chi-squared divergence $\chi[p_\lambda : p_{\lambda'}] = \frac{\|\lambda-\lambda'\|^2}{2\lambda_2\lambda_2'}$ (with $h_f(0) = 0$).

References

[1] Sreangsu Acharyya, Arindam Banerjee, and Daniel Boley. Bregman divergences and triangle inequality. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 476–484. SIAM, 2013.

[2] Anil Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.

[3] Limei Cao, Didong Li, Erchuan Zhang, Zhenning Zhang, and Huafei Sun. A statistical cohomogeneity one metric on the upper plane with constant negative curvature. *Advances in Mathematical Physics*, 2014.

[4] Frédéric Chyzak and Frank Nielsen. A closed-form formula for the Kullback-Leibler divergence between Cauchy distributions. *arXiv preprint arXiv:1905.10965*, 2019.

[5] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observation. *studia scientiarum Mathematicarum Hungarica*, 2:229–318, 1967.

[6] Morris L Eaton. *Group invariance applications in statistics*. Institute of Mathematical Statistics Hayward, California, 1989.

[7] Bent Fuglede and Flemming Topsoe. Jensen-Shannon divergence and Hilbert space embedding. In *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, page 31. IEEE, 2004.

[8] Shin-itiro Goto and Ken Umeno. Maps on statistical manifolds exactly reduced from the perron-frobenius equations for solvable chaotic maps. *Journal of Mathematical Physics*, 59(3):032701, 2018.

[9] Peter Kafka, Ferdinand Österreicher, and István Vincze. On powers of f-divergences defining a distance. *Studia Sci. Math. Hungar*, 26(4):415–422, 1991.

[10] Srinivasan Kesavan. *Measure and Integration.* Springer, 2019.

[11] Frank B Knight. A characterization of the Cauchy type. *Proceedings of the American Mathematical Society*, 55(1):130–135, 1976.

[12] Lucien Le Cam. *Asymptotic methods in statistical decision theory.* Springer Science & Business Media, 2012.

[13] Gérard Letac. Which functions preserve Cauchy laws? *Proceedings of the American Mathematical Society*, 67(2):277–286, 1977.

[14] Peter McCullagh. On the distribution of the Cauchy maximum-likelihood estimator. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 440(1909):475–479, 1993.

[15] Ann F. S. Mitchell. Statistical manifolds of univariate elliptic distributions. *International Statistical Review/Revue Internationale de Statistique*, pages 1–16, 1988.

[16] Tristan Needham. *Visual complex analysis.* Oxford University Press, 1998.

[17] Frank Nielsen. On Voronoi diagrams on the information-geometric Cauchy manifolds. *Entropy*, 22(7):713, 2020.

[18] Frank Nielsen. On information projections between multivariate elliptical and location-scale families. Technical report, arXiv, 2021. 2101.03839.

[19] Frank Nielsen and Sylvain Boltz. The Burbea-Rao and Bhattacharyya centroids. *IEEE Transactions on Information Theory*, 57(8):5455–5466, 2011.

[20] Frank Nielsen and Richard Nock. On the chi square and higher-order chi distances for approximating f-divergences. *IEEE Signal Processing Letters*, 21(1):10–13, 2013.

[21] Frank Nielsen and Richard Nock. Total Jensen divergences: definition, properties and clustering. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2016–2020. IEEE, 2015.

[22] Richard Nock, Frank Nielsen, and Shun-ichi Amari. On conformal divergences and their population minimizers. *IEEE Transactions on Information Theory*, 62(1):527–538, 2015.

[23] Ferdinand Österreicher and Igor Vajda. A new class of metric divergences on probability spaces and its applicability in statistics. *Annals of the Institute of Statistical Mathematics*, 55(3):639–653, 2003.

[24] Igor Vajda. On metric divergences of probability measures. *Kybernetika*, 45(6):885–900, 2009.