

## Hölder-Like Csiszar's Type Inequalities

George A. Anastassiou  
Department of Mathematical Sciences  
University of Memphis  
Memphis, TN 38152 U.S.A.  
E-mail address: ganastss@memphis.edu

**Abstract** Here we present a set of a large variety of probabilistic inequalities regarding Csiszar's  $f$ -divergence of two probability measures. These are in parallel to Hölder's type inequalities and some of them are attained.

**2000 AMS Subject Classification:** 26D15, 28A25, 60E15.

**Key Words and Phrases:** Csiszar's discrimination and divergence, Hölder inequality, Rogers inequality, probabilistic inequality.

### 1. Background

Here we follow [6].

Let  $f$  be an arbitrary convex function from  $(0, +\infty)$  into  $\mathbb{R}$  which is strictly convex at 1. We agree with the following notational conventions:

$$\begin{aligned} f(0) &= \lim_{u \rightarrow 0^+} f(u), \\ 0 \cdot f\left(\frac{0}{0}\right) &= 0, \end{aligned} \tag{1}$$

$$0f\left(\frac{a}{0}\right) = \lim_{\varepsilon \rightarrow 0^+} \varepsilon f\left(\frac{a}{\varepsilon}\right) = a \lim_{u \rightarrow +\infty} \frac{f(u)}{u}, \quad (0 < a < +\infty).$$

Let  $(X, \mathcal{A}, \lambda)$  be an arbitrary measure space with  $\lambda$  being a finite or  $\sigma$ -finite measure. Let also  $\mu_1, \mu_2$  probability measures on  $X$  such that  $\mu_1, \mu_2 \ll \lambda$  (absolutely continuous).

Denote the Radon–Nikodym derivatives (densities) of  $\mu_i$  with respect to  $\lambda$  by  $p_i(x)$ :

$$p_i(x) = \frac{\mu_i(dx)}{\lambda(dx)}, \quad i = 1, 2.$$

**Definition 1** (see [6]). The quantity

$$\Gamma_f(\mu_1, \mu_2) = \int_X p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) \quad (2)$$

is called the *f-divergence* of the probability measures  $\mu_1$  and  $\mu_2$ . The function  $f$  is called the *base function*. By Lemma 1.1 of [6]  $\Gamma_f(\mu_1, \mu_2)$  is always well-defined and  $\Gamma_f(\mu_1, \mu_2) \geq f(1)$  with equality only for  $\mu_1 = \mu_2$ . Also from [6] it derives that  $\Gamma_f(\mu_1, \mu_2)$  does not depend on the choice of  $\lambda$ . When assuming  $f(1) = 0$  then  $\Gamma_f$  can be considered as the most general measure of difference between probability measures.

The Csiszar's *f-divergence*  $\Gamma_f$  incorporated most of special cases of probability measure distances, e.g. the variation distance,  $\chi^2$ -divergence, information for discrimination or generalized entropy, information gain, mutual information, mean square contingency, etc.  $\Gamma_f$  has tremendous applications to almost all applied sciences where stochastics enters. For more see [1]–[8].

In this note all the base functions  $f$  appearing in the function  $\Gamma_f$  are assumed to have all the above properties of  $f$ . Notice that  $\Gamma_f(\mu_1, \mu_2) \leq \Gamma_{|f|}(\mu_1, \mu_2)$ . For arbitrary functions  $f, g$  we notice that

$$f^r \circ g = (f \circ g)^r, \quad r \in \mathbb{R} \quad \text{and} \quad |f \circ g| = |f| \circ g.$$

Clearly one can have widely products of (strictly) convex functions to be (strictly) convex.

## 2. Results

We start with

**Theorem 1.** *Let  $p, q > 1$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ . Then*

$$\Gamma_{|f_1 f_2|}(\mu_1, \mu_2) \leq (\Gamma_{|f_1|^p}(\mu_1, \mu_2))^{1/p} (\Gamma_{|f_2|^q}(\mu_1, \mu_2))^{1/q}. \quad (3)$$

Equality in (3) is true when  $|f_1|^p = c|f_2|^q$  for  $c > 0$ . More precisely, inequality (3) holds as equality iff there are  $a, b \in \mathbb{R}$  not both zero such that

$$p_2 \cdot \left( (a|f_1|^p - b|f_2|^q) \left( \frac{p_1}{p_2} \right) \right) = 0, \quad \text{a.e. } [\lambda].$$

**Proof.** Here we use the Hölder's inequality. We observe that

$$\begin{aligned} \Gamma_{|f_1 f_2|}(\mu_1, \mu_2) &= \int_X p_2 |f_1| \left( \frac{p_1}{p_2} \right) \cdot |f_2| \left( \frac{p_1}{p_2} \right) d\lambda \\ &= \int_X \left( p_2^{1/p} |f_1| \left( \frac{p_1}{p_2} \right) \right) \left( p_2^{1/q} |f_2| \left( \frac{p_1}{p_2} \right) \right) d\lambda \\ &\leq \left( \int_X p_2 |f_1|^p \left( \frac{p_1}{p_2} \right) d\lambda \right)^{1/p} \left( \int_X p_2 |f_2|^q \left( \frac{p_1}{p_2} \right) d\lambda \right)^{1/q} \\ &= (\Gamma_{|f_1|^p}(\mu_1, \mu_2))^{1/p} (\Gamma_{|f_2|^q}(\mu_1, \mu_2))^{1/q}. \quad \square \end{aligned}$$

**Corollary 1.** *It holds*

$$\Gamma_{|f_1 f_2|}(\mu_1, \mu_2) \leq (\Gamma_{f_1^2}(\mu_1, \mu_2))^{1/2} (\Gamma_{f_2^2}(\mu_1, \mu_2))^{1/2}. \quad (4)$$

**Proof.** By (3) setting  $p = q = 2$ . □

We continue with a generalization of Theorem 1.

**Theorem 2.** *Let  $a_1, a_2, \dots, a_n > 1$ ,  $n \in \mathbb{N}$ :  $\sum_{i=1}^n \frac{1}{a_i} = 1$ . Then*

$$\Gamma_{\left| \prod_{i=1}^n f_i \right|}(\mu_1, \mu_2) \leq \prod_{i=1}^n (\Gamma_{|f_i|^{a_i}}(\mu_1, \mu_2))^{1/a_i}. \quad (5)$$

**Proof.** Here we use the generalized Hölder's inequality, see p. 186, [11] and p. 200, [10]. We see that

$$\begin{aligned} \Gamma_{\left| \prod_{i=1}^n f_i \right|}(\mu_1, \mu_2) &= \int_X p_2 \left| \prod_{i=1}^n f_i \left( \frac{p_1}{p_2} \right) \right| d\lambda = \int_X \prod_{i=1}^n \left( p_2^{1/a_i} \left| f_i \left( \frac{p_1}{p_2} \right) \right| \right) d\lambda \\ &\leq \prod_{i=1}^n \left( \int_X p_2 |f_i|^{a_i} \left( \frac{p_1}{p_2} \right) d\lambda \right)^{1/a_i} = \prod_{i=1}^n (\Gamma_{|f_i|^{a_i}}(\mu_1, \mu_2))^{1/a_i}. \quad \square \end{aligned}$$

It follows the counterpart of Theorem 1.

**Theorem 3.** *Let  $0 < p < 1$  and  $q < 0$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ . We assume that  $p_2 > 0$  a.e.  $[\lambda]$ . Then we have*

$$\Gamma_{|f_1 f_2|}(\mu_1, \mu_2) \geq (\Gamma_{|f_1|^p}(\mu_1, \mu_2))^{1/p} \Gamma_{|f_2|^q}(\mu_1, \mu_2)^{1/q} \quad (6)$$

unless  $\Gamma_{|f_2|^q}(\mu_1, \mu_2) = 0$ . Inequality (6) holds as an equality iff there exist  $A, B \in \mathbb{R}_+$  ( $A + B \neq 0$ ) such that

$$Ap_2^{1/p} \left| f_1 \left( \frac{p_1}{p_2} \right) \right| = B \left( p_2^{1/q} \left| f_2 \left( \frac{p_1}{p_2} \right) \right| \right)^{1/p-1}, \quad \text{a.e. } [\lambda].$$

**Proof.** It is based on Hölder's inequality for  $0 < p < 1$ , see p. 191 of [10]. Indeed we have

$$\begin{aligned} \Gamma_{|f_1 f_2|}(\mu_1, \mu_2) &= \int_X p_2 \left| f_1 \left( \frac{p_1}{p_2} \right) \right| \left| f_2 \left( \frac{p_1}{p_2} \right) \right| d\lambda \\ &= \int_X \left( p_2^{1/p} \left| f_1 \left( \frac{p_1}{p_2} \right) \right| \right) \left( p_2^{1/q} \left| f_2 \left( \frac{p_1}{p_2} \right) \right| \right) d\lambda \\ &\geq \left( \int_X p_2 |f_1|^p \left( \frac{p_1}{p_2} \right) d\lambda \right)^{1/p} \left( \int_X p_2 |f_2|^q \left( \frac{p_1}{p_2} \right) d\lambda \right)^{1/q} \\ &= (\Gamma_{|f_1|^p}(\mu_1, \mu_2))^{1/p} (\Gamma_{|f_2|^q}(\mu_1, \mu_2))^{1/q}. \quad \square \end{aligned}$$

Next we have

**Proposition 1.** Let  $f_2 \left( \frac{p_1}{p_2} \right) \in L_\infty(X)$ . Then

$$\Gamma_{|f_1 f_2|}(\mu_1, \mu_2) \leq \Gamma_{|f_1|}(\mu_1, \mu_2) \cdot \left( \text{esssup} \left| f_2 \left( \frac{p_1}{p_2} \right) \right| \right). \quad (7)$$

**Proof.** We notice that

$$\begin{aligned} \Gamma_{|f_1 f_2|}(\mu_1, \mu_2) &= \int_X p_2 \cdot |f_1| \cdot |f_2| \left( \frac{p_1}{p_2} \right) d\lambda = \int_X \left( p_2 |f_1| \left( \frac{p_1}{p_2} \right) \right) |f_2| \left( \frac{p_1}{p_2} \right) d\lambda \\ &\leq \left( \int_X p_2 |f_1| \left( \frac{p_1}{p_2} \right) d\lambda \right) \cdot \text{esssup} \left( |f_2| \left( \frac{p_1}{p_2} \right) \right) \\ &= \Gamma_{|f_1|}(\mu_1, \mu_2) \cdot \text{esssup} \left( |f_2| \left( \frac{p_1}{p_2} \right) \right). \quad \square \end{aligned}$$

In the following we need from [9], p. 120, see also [12]

**Theorem 4.** Let  $f$  and  $g$  be positive, measurable functions on  $X$  for a measure space  $(X, \mathcal{F}, \mu)$ . Let  $0 < t < r < m < \infty$ . Then it holds:

(i) The Rogers inequality

$$\left( \int_X f g^r d\mu \right)^{m-t} \leq \left( \int_X f g^t d\mu \right)^{m-r} \left( \int_X f g^m d\mu \right)^{r-t} \quad (8)$$

provided that the integrals on the right are finite.

(ii) For  $m > 1$ , the "historical Hölder's inequality"

$$\left( \int_X fg \, d\mu \right)^m \leq \left( \int_X f \, d\mu \right)^{m-1} \left( \int_X fg^m \, d\mu \right) \quad (9)$$

provided that the integrals on the right are finite.

Using the last results we present our final result.

**Theorem 5.** Let  $f_1, f_2$  with  $|f_1|, |f_2| > 0$  and  $p_2(x) > 0$  a.e.  $[\lambda]$ . Let  $0 < t < r < m < \infty$ . Then it holds

i)

$$(\Gamma_{|f_1||f_2|^r}(\mu_1, \mu_2))^{m-t} \leq (\Gamma_{|f_1||f_2|^t}(\mu_1, \mu_2))^{m-r} (\Gamma_{|f_1||f_2|^m}(\mu_1, \mu_2))^{r-t}. \quad (10)$$

ii) For  $m > 1$ ,

$$(\Gamma_{|f_1||f_2|}(\mu_1, \mu_2))^m \leq (\Gamma_{|f_1|}(\mu_1, \mu_2))^{m-1} \Gamma_{|f_1||f_2|^m}(\mu_1, \mu_2). \quad (11)$$

**Proof.** i) By applying (8) we get

$$\begin{aligned} (\Gamma_{|f_1||f_2|^r}(\mu_1, \mu_2))^{m-t} &= \left( \int_X p_2 |f_1| \left( \frac{p_1}{p_2} \right) |f_2|^r \left( \frac{p_1}{p_2} \right) d\lambda \right)^{m-t} \\ &\leq \left( \int_X p_2 |f_1| \left( \frac{p_1}{p_2} \right) |f_2|^t \left( \frac{p_1}{p_2} \right) d\lambda \right)^{m-r} \\ &\quad \cdot \left( \int_X p_2 |f_1| \left( \frac{p_1}{p_2} \right) |f_2|^m \left( \frac{p_1}{p_2} \right) d\lambda \right)^{r-t} \\ &= (\Gamma_{|f_1||f_2|^t}(\mu_1, \mu_2))^{m-r} (\Gamma_{|f_1||f_2|^m}(\mu_1, \mu_2))^{r-t}. \end{aligned}$$

ii) By applying (9) we obtain

$$\begin{aligned} (\Gamma_{|f_1||f_2|}(\mu_1, \mu_2))^m &= \left( \int_X p_2 |f_1| \left( \frac{p_1}{p_2} \right) |f_2| \left( \frac{p_1}{p_2} \right) d\lambda \right)^m \\ &\leq \left( \int_X p_2 |f_1| \left( \frac{p_1}{p_2} \right) d\lambda \right)^{m-1} \left( \int_X p_2 |f_1| \left( \frac{p_1}{p_2} \right) |f_2|^m \left( \frac{p_1}{p_2} \right) d\lambda \right) \\ &= (\Gamma_{|f_1|}(\mu_1, \mu_2))^{m-1} \Gamma_{|f_1||f_2|^m}(\mu_1, \mu_2). \quad \square \end{aligned}$$

**Note.** Let  $X$  be a finite or countable discrete set,  $\mathcal{A}$  be its power set  $\mathcal{P}(X)$  and  $\lambda$  has mass 1 for each  $x \in X$ , then  $\Gamma_f$  in (2) becomes a finite or infinite sum, respectively. As a result all the above presented integral inequalities are discretized and become summation inequalities.

## References

- [1] Anastassiou, G.A., Basic optimal approximation of Csiszar's  $f$ -divergence, submitted, 2004.
- [2] Anastassiou, G.A., Fractional and other approximation of Csiszar's  $f$ -divergence, submitted, 2004.
- [3] Anastassiou, G.A., Representations and estimates to Csiszar's  $f$ -divergence, submitted, 2004.
- [4] Anastassiou, G.A., Higher order optimal approximation of Csiszar's  $f$ -divergence, to appear, *Nonlinear Analysis*, 2005.
- [5] Csiszar, I., Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, *Magyar Tud. Akad. Mat. Kutato Int. Közl.* **8** (1963), 85–108.
- [6] Csiszar, I., Information-type measures of difference of probability distributions and indirect observations, *Studia Math. Hungarica* **2** (1967), 299–318.
- [7] Csiszar, I., On topological properties of  $f$ -divergences, *Studia Scientiarum Mathematicarum Hungarica*, **2** (1967), 329–339.
- [8] Dragomir, S.S., *Inequalities for Csiszar  $f$ -Divergence in Information Theory*, edited monograph, Victoria University, Melbourne, Australia, 2000. On line: <http://rgmia.vu.edu.au>
- [9] Dudley, R.M., *Real Analysis and Probability*, Wadsworth & Brooks/Cole, Pacific Grove, CA, 1989.
- [10] Hewitt, E. and Stromberg, K., *Real and Abstract Analysis*, Springer-Verlag, #25, New York, Berlin, 1965.
- [11] Kingman, J.F.C. and Taylor, S.J., *Introduction to Measure and Probability*, Cambridge University Press, Cambridge, 1966.
- [12] Rogers, Leonard James, An extension of a certain theorem in inequalities, *Messenger of Math.* **17** (1888), 145–150.