A GENERALIZED $f$–DIVERGENCE FOR PROBABILITY VECTORS AND APPLICATIONS

S.S. DRAGOMIR

Abstract. A generalized $f$–divergence for probability vectors and some fundamental inequalities are pointed out.

1. Introduction

One of the important issues in many applications of Probability Theory is finding an appropriate measure of distance (or difference or discrimination) between two probability distributions. A number of divergence measures for this purpose have been proposed and extensively studied by Jeffreys [31], Kullback and Leibler [42], Rényi [29], Havrda and Charvat [25], Kapur [44], Sharma and Mittal [6], Burbea and Rao [11], Rao [34], Lin [10], Csiszár [11], Ali and Silvey [52], Vajda [37], Shioya and Dake [21] and others (see for example [44] and the references therein).

These measures have been applied in a variety of fields such as: anthropology [34], genetics [43], finance, economics, and political science [47], [48], [39], biology [19], the analysis of contingency tables [9], approximation of probability distributions [26], [23], signal processing [24], [3] and pattern recognition [8], [53]. A number of these measures of distance are specific cases of Csiszár $f$-divergence and so further exploration of this concept will have a flow on effect to other measures of distance and to areas in which they are applied.

Assume that a set $\chi$ and the $\sigma$–finite measure $\mu$ are given. Consider the set of all probability densities on $\mu$ to be $\Omega := \{ p: \chi \to \mathbb{R}, p(x) \geq 0, \int p(x) \, d\mu(x) = 1 \}$.

The Kullback-Leibler divergence [42] is well known among the information divergences. It is defined as:-

$$ D_{KL}(p, q) := \int \chi p(x) \log \left( \frac{p(x)}{q(x)} \right) \, d\mu(x), \quad p, q \in \Omega, $$

where log is to base 2.

In Information Theory and Statistics, various divergences are applied in addition to the Kullback-Leibler divergence. These are the: variation distance $D_v$, Hellinger distance $D_H$ [31], $\chi^2$–divergence $D_{\chi^2}$, $\alpha$–divergence $D_\alpha$, Bhattacharyya distance $D_B$ [42], Harmonic distance $D_{Ha}$, Jeffreys distance $D_J$ [31], triangular discrimination $D_\Delta$ [33], etc... They are defined as follows:

$$ D_v(p, q) := \int \chi |p(x) - q(x)| \, d\mu(x), \quad p, q \in \Omega; $$

**Date:** March 20, 2000.
**1991 Mathematics Subject Classification.** Primary 94Xxx; Secondary 26D15.
**Key words and phrases.** A Generalized $f$–Divergence, Jensen’s Inequality.
(1.3) \[ D_H (p, q) := \int_{\chi} \left[ \sqrt{p(x)} - \sqrt{q(x)} \right]^2 d\mu(x), \quad p, q \in \Omega; \]

(1.4) \[ D_{\chi^2} (p, q) := \int_{\chi} p(x) \left[ \left( \frac{q(x)}{p(x)} \right)^2 - 1 \right] d\mu(x), \quad p, q \in \Omega; \]

(1.5) \[ D_\alpha (p, q) := \frac{4}{1 - \alpha^2} \left[ 1 - \int_{\chi} \left| \frac{p(x)}{\alpha} \right|^{\frac{1}{1-\alpha}} \left| \frac{q(x)}{\alpha} \right|^{\frac{1}{1-\alpha}} d\mu(x) \right], \quad p, q \in \Omega; \]

(1.6) \[ D_B (p, q) := \int_{\chi} \sqrt{p(x)q(x)} d\mu(x), \quad p, q \in \Omega; \]

(1.7) \[ D_{\alpha} (p, q) := \frac{4}{1 - \alpha^2} \left[ 1 - \int_{\chi} \left| \frac{p(x)}{\alpha} \right|^{\frac{1}{1-\alpha}} \left| \frac{q(x)}{\alpha} \right|^{\frac{1}{1-\alpha}} d\mu(x) \right], \quad p, q \in \Omega; \]

(1.8) \[ D_J (p, q) := \int_{\chi} \left[ p(x) - q(x) \right] \ln \left[ \frac{p(x)}{q(x)} \right] d\mu(x), \quad p, q \in \Omega; \]

(1.9) \[ D_\Delta (p, q) := \int_{\chi} \frac{\left[ p(x) - q(x) \right]^2}{p(x) + q(x)} d\mu(x), \quad p, q \in \Omega. \]

For other divergence measures, see the paper [25] by Kapur or the book on line [5] by Taneja. For a comprehensive collection of preprints available on line, see the RGMIA web site http://rgmia.vu.edu.au/papersinfth.html.

Csiszár \( f \)-divergence is defined as follows [11]

(1.10) \[ D_f (p, q) := \int_{\chi} p(x) f \left[ \frac{q(x)}{p(x)} \right] d\mu(x), \quad p, q \in \Omega, \]

where \( f \) is convex on \((0, \infty)\). It is assumed that \( f(u) \) is zero and strictly convex at \( u = 1 \). By appropriately defining this convex function, various divergences are derived. All the above distances (1.1) – (1.9), are particular instances of Csiszár \( f \)-divergence. There are also many others which are not in this class (see for example [14] or [5]). For the basic properties of Csiszár \( f \)-divergence see [11]-[11].

In [1], Lin and Wong (see also [10]) introduced the following divergence

(1.11) \[ D_{LW} (p, q) := \int_{\chi} p(x) \log \left[ \frac{p(x)}{\frac{1}{2} p(x) + \frac{1}{2} q(x)} \right] d\mu(x), \quad p, q \in \Omega. \]

This can be represented as follows, using the Kullback-Leibler divergence:

\[ D_{LW} (p, q) = D_{KL} \left( p, \frac{1}{2} p + \frac{1}{2} q \right). \]

Lin and Wong have established the following inequalities

(1.12) \[ D_{LW} (p, q) \leq \frac{1}{2} D_{KL} (p, q); \]

(1.13) \[ D_{LW} (p, q) + D_{LW} (q, p) \leq D_{v} (p, q) \leq 2; \]
(1.14) \[ D_{LW}(p,q) \leq 1. \]

In \[ 37 \], Shioya and Da-te improved \[ 1.12 \] - \[ 1.14 \] by showing that

\[ D_{LW}(p,q) \leq \frac{1}{2} D_v(p,q) \leq 1. \]

For classical and new results in comparing different kinds of divergence measures, see the papers \[ 31 \] - \[ 38 \] where further references are given.

2. The Generalized \( f \)-Divergence and Some Inequalities

Let \( (p_1, ..., p_n) \) and \( (q_1, ..., q_n) \) be such that \( p_i, q_i \in \Omega \) (\( i = 1, ..., n \)). Define the \( f \)-divergence between two vectors by

\[ (2.1) \quad \tilde{D}_f(p_1, ..., p_n, q_1, ..., q_n) \]

\[ = \int_x \cdots \int_x p_1(x_1) \cdots p_n(x_n) f \left[ \frac{q_1(x_1)}{p_1(x_1)} + \cdots + \frac{q_n(x_n)}{p_n(x_n)} \right] d\mu(x_1) \cdots d\mu(x_n) \]

provided that the multiple integral exists and is finite and \( f : [0, \infty) \rightarrow \mathbb{R} \) is a given mapping.

Usually, the mapping is assumed to be convex and normalised.

The following simple result holds.

**Theorem 1.** Assume that the mapping \( f : [0, \infty) \rightarrow \mathbb{R} \) is convex. Then we have the inequalities

\[ (2.2) \quad f(1) \leq \tilde{D}_f(p_1, ..., p_{n+1}, q_1, ..., q_{n+1}) \leq \tilde{D}_f(p_1, ..., p_n, q_1, ..., q_n) \leq \cdots \leq \tilde{D}_f(p_1, p_2, q_1, q_2) \leq \tilde{D}_f(p_1, q_1), \]

provided that \( p_i, q_i \in \Omega, i \in \mathbb{N} \) and \( D_f(\cdot, \cdot) \) is the usual \( f \)-Csiszár divergence.

**Proof.** Using Jensen’s inequality for multiple integrals, we obtain

\[ \int_x \cdots \int_x p_1(x_1) \cdots p_n(x_n) f \left[ \frac{1}{n} \left( \frac{q_1(x_1)}{p_1(x_1)} + \cdots + \frac{q_n(x_n)}{p_n(x_n)} \right) \right] d\mu(x_1) \cdots d\mu(x_n) \]

\[ \geq f \left[ \int_x \cdots \int_x p_1(x_1) \cdots p_n(x_n) \cdot \frac{1}{n} \left( \frac{q_1(x_1)}{p_1(x_1)} + \cdots + \frac{q_n(x_n)}{p_n(x_n)} \right) d\mu(x_1) \cdots d\mu(x_n) \right] \]

\[ = f(1) \]

because a simple calculation shows that

\[ \frac{1}{n} \int_x \cdots \int_x p_1(x_1) \cdots p_n(x_n) \left( \frac{q_1(x_1)}{p_1(x_1)} + \cdots + \frac{q_n(x_n)}{p_n(x_n)} \right) d\mu(x_1) \cdots d\mu(x_n) \]

\[ = \frac{1}{n} \left[ \int_x q_1(x_1) d\mu(x_1) \int_x p_2(x_2) d\mu(x_2) \cdots \int_x p_n(x_n) d\mu(x_n) \right. \]

\[ + \int_x p_1(x_1) d\mu(x_1) \int_x q_n(x_n) d\mu(x_n) \right] = 1 \]

and the first inequality in \[ 2.2 \] is proved.
If we apply Jensen’s inequality for \( y_1, \ldots, y_{n+1} \), we may write

\[
(2.3) \quad \frac{1}{n+1} \left[ f(y_1) + \ldots + f(y_{n+1}) \right] \geq f \left( \frac{y_1 + \ldots + y_{n+1}}{n+1} \right).
\]

Choose

\[
y_1 = z_1 + z_2 + \ldots + z_{n-1} + z_n,
\]
\[
y_2 = \frac{z_1 + z_3 + \ldots + z_n + z_{n+1}}{n},
\]
\[
\vdots
\]
\[
y_{n+1} = \frac{z_{n+1} + z_1 + \ldots + z_{n-1}}{n}.
\]

Then

\[
\frac{1}{n+1} (y_1 + \ldots + y_{n+1}) = \frac{n(z_1 + z_2 + \ldots + z_{n+1})}{n(n+1)} = \frac{z_1 + z_2 + \ldots + z_n + z_{n+1}}{n+1}
\]

and then, by (2.3) we have the inequality

\[
(2.4) \quad \frac{1}{n+1} \left[ f \left( \frac{z_1 + z_2 + \ldots + z_{n-1} + z_n}{n} \right) + \ldots + f \left( \frac{z_{n+1} + z_1 + \ldots + z_{n-1}}{n} \right) \right]
\geq f \left( \frac{z_1 + z_2 + \ldots + z_n + z_{n+1}}{n+1} \right)
\]

for all \( z_1, \ldots, z_{n+1} \in [0, \infty) \).

If we put in (2.4) \( z_i = \frac{p_i(x_i)}{q_i(x_i)} \) (\( i = 1, \ldots, n+1 \)) and if we multiply with \( p_1(x_1) \ldots p_{n+1}(x_{n+1}) \geq 0 \) and then integrate over \( \chi^{n+1} \), we may write

\[
(2.5) \quad \frac{1}{n+1} \int_{\chi} \ldots \int_{\chi} p_1(x_1) \ldots p_{n+1}(x_{n+1})
\times f \left[ \frac{1}{n} \left( \frac{p_1(x_1)}{q_1(x_1)} + \ldots + \frac{p_n(x_n)}{q_n(x_n)} \right) \right] d\mu(x_1) \ldots d\mu(x_{n+1})
\]
\[
\quad + \int_{\chi} \ldots \int_{\chi} p_1(x_1) \ldots p_{n+1}(x_{n+1})
\times f \left[ \frac{1}{n} \left( \frac{p_{n+1}(x_{n+1})}{q_{n+1}(x_{n+1})} + \frac{p_1(x_1)}{q_1(x_1)} + \ldots + \frac{p_{n-1}(x_{n-1})}{q_{n-1}(x_{n-1})} \right) \right] d\mu(x_1) \ldots d\mu(x_{n+1})
\geq \int_{\chi} \ldots \int_{\chi} p_1(x_1) \ldots p_{n+1}(x_{n+1})
\times f \left[ \frac{1}{n+1} \left( \frac{p_1(x_1)}{q_1(x_1)} + \ldots + \frac{p_{n+1}(x_{n+1})}{q_{n+1}(x_{n+1})} \right) \right] d\mu(x_1) \ldots d\mu(x_{n+1})
\]

\[
= \bar{D}_f(p_1, \ldots, p_{n+1}, q_1, \ldots, q_{n+1})
\]
A GENERALIZED $f$–DIVERGENCE

and as

$$
\frac{1}{n+1} \left[ \int_{\chi} \ldots \int_{\chi} p_1(x_1) \ldots p_{n+1}(x_{n+1}) \right. \\
\times f \left[ \frac{1}{n} \left( \frac{p_1(x_1)}{q_1(x_1)} + \ldots + \frac{p_n(x_n)}{q_n(x_n)} \right) \right] d\mu(x_1) \ldots d\mu(x_{n+1}) \\
= \ldots = \int_{\chi} \ldots \int_{\chi} p_1(x_1) \ldots p_{n+1}(x_{n+1}) \\
\times f \left[ \frac{1}{n} \left( \frac{p_{n+1}(x_{n+1})}{q_{n+1}(x_{n+1})} + \ldots + \frac{p_{n-1}(x_{n-1})}{q_{n-1}(x_{n-1})} \right) \right] d\mu(x_1) \ldots d\mu(x_{n+1}) \\
= \bar{D}_f(p_1, \ldots, p_n, q_1, \ldots, q_n),
$$

then, from (2.5), we deduce the second inequality for all $n$, and the theorem is thus proved. 

For $f : [0, \infty) \to \mathbb{R}$, $f(x) = (x - 1)^2$, we obtain the $\chi^2$–distance

$$
D_{\chi^2}(p_1, q_1) = \int_{\chi} \frac{q_1(x) - p_1(x)}{p_1(x)} d\mu(x), \ p, q \in \Omega.
$$

The following corollary holds.

**Corollary 1.** Let $p_i, q_i \in \Omega$ ($i \in \mathbb{N}$) and define the $\chi^2$–divergence by

(2.6) \( \bar{D}_{\chi^2}(p_1, \ldots, p_n; q_1, \ldots, q_n) \)

\[
= \int_{\chi} \ldots \int_{\chi} p_1(x_1) \ldots p_n(x_n) \left[ \frac{q_1(x_1) + \ldots + q_n(x_n)}{p_1(x_1) + \ldots + p_n(x_n)} - 1 \right]^2 d\mu(x_1) \ldots d\mu(x_n).
\]

Then

(2.7) \( \bar{D}_{\chi^2}(p_1, \ldots, p_n; q_1, \ldots, q_n) = \frac{1}{n^2} \sum_{i=1}^{n} D_{\chi^2}(p_i, q_i) \)

and

(2.8) \( 0 \leq \bar{D}_{\chi^2}(p_1, \ldots, p_{n+1}; q_1, \ldots, q_{n+1}) \leq \bar{D}_{\chi^2}(p_1, \ldots, p_n; q_1, \ldots, q_n) \)

\[
\leq \ldots \leq \bar{D}_{\chi^2}(p_1, p_2, q_1, q_2) \leq \bar{D}_{\chi^2}(p_1, q_1)
\]

for all $n \in \mathbb{N}$, $n \geq 1$. 
Proof. We observe that
\[
\bar{D}_\chi^2 (p_1, ..., p_n; q_1, ..., q_n)
= \frac{1}{n^2} \int_X \cdots \int_X p_1 (x_1) \cdots p_n (x_n)
\times \left[ \left( \frac{q_1 (x_1)}{p_1 (x_1)} - 1 \right) + \cdots + \left( \frac{q_n (x_n)}{p_n (x_n)} - 1 \right) \right]^2
d\mu (x_1) \cdots d\mu (x_n)
\]
\[
= \frac{1}{n^2} \int_X \cdots \int_X p_1 (x_1) \cdots p_n (x_n) \left[ \sum_{i=1}^n \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right)^2 \right]
+ 2 \sum_{1 \leq i < j \leq n} \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right) \left( \frac{q_j (x_j)}{p_j (x_j)} - 1 \right)
d\mu (x_1) \cdots d\mu (x_n)
\]
\[
= \frac{1}{n^2} \sum_{i=1}^n \int_X p_1 (x_1) \cdots p_n (x_n) \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right)^2
d\mu (x_1) \cdots d\mu (x_n)
+ 2 \sum_{1 \leq i < j \leq n} \int_X p_1 (x_1) \cdots p_n (x_n)
\times \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right) \left( \frac{q_j (x_j)}{p_j (x_j)} - 1 \right)
d\mu (x_1) \cdots d\mu (x_n).
\]

However, for any \( i \), we have
\[
\int_X \cdots \int_X p_1 (x_1) \cdots p_n (x_n) \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right)^2
d\mu (x_1) \cdots d\mu (x_n)
= \int_X p_1 (x_1) d\mu (x_1) \cdots \int_X p_1 (x_i) d\mu (x_i) \cdots \int_X p_n (x_n) d\mu (x_n)
= \int_X p_1 (x_1) \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right)^2 d\mu (x_1) = \bar{D}_\chi^2 (p_1, q_i)
\]
and (for \( i \neq j \))
\[
\int_X \cdots \int_X p_1 (x_1) \cdots p_n (x_n) \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right) \left( \frac{q_j (x_j)}{p_j (x_j)} - 1 \right)
d\mu (x_1) \cdots d\mu (x_n)
= \int_X p_1 (x_1) d\mu (x_1) \int_X p_1 (x_i) \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right) d\mu (x_i)
\cdots \int_X p_2 (x_j) \left( \frac{q_j (x_j)}{p_j (x_j)} - 1 \right) d\mu (x_j)
\cdots \int_X p_n (x_n) d\mu (x_n) = 0
\]
as
\[
\int_X p_1 (x_i) \left( \frac{q_i (x_i)}{p_i (x_i)} - 1 \right) d\mu (x_i) = \int_X p_2 (x_j) \left( \frac{q_j (x_j)}{p_j (x_j)} - 1 \right) d\mu (x_j) = 0
\]
and then the representation (2.7) is proved.

The sequence of inequalities in (2.8) follows by (2.2) and we omit the details. \( \blacksquare \)
For $f: [0, \infty) \rightarrow \mathbb{R}$, $f(x) = e^x$, define the exp-divergence by

$$D_{\exp} (p_1, q_1) := \int_x p_1(x) \exp \left( \frac{q_1(x)}{p_1(x)} \right) d\mu(x), \quad p, q \in \Omega.$$ (2.9)

The following corollary holds.

**Corollary 2.** Let $p_i, q_i \in \Omega \ (i \in \mathbb{N})$ and define the exp-divergence by

$$\bar{D}_{\exp} (p_1, ..., p_n; q_1, ..., q_n) = \prod_{i=1}^{n} D_{\exp} (p_i, q_i)$$ (2.10)

Then

$$D_{\exp} (p_1, ..., p_n; q_1, ..., q_n) = \prod_{i=1}^{n} D_{\exp} (p_i, q_i)$$ (2.11)

and

$$1 \leq D_{\exp} (p_1, ..., p_{n+1}; q_1, ..., q_{n+1}) \leq D_{\exp} (p_1, ..., p_n; q_1, ..., q_n) \leq \cdots \leq \bar{D}_{\exp} (p_1, p_2, q_1, q_2) \leq D_{\exp} (p_1, q_1),$$ (2.12)

for all $n \geq 1$.

**Proof.** We observe that

$$\bar{D}_{\exp} (p_1, ..., p_n; q_1, ..., q_n) = \int_x \prod_{i=1}^{n} p_i(x_i) \exp \left( \frac{q_i(x_i)}{p_i(x_i)} \right) d\mu(x_1) ... d\mu(x_n)$$

$$= \prod_{i=1}^{n} \int_x p_i(x_i) \exp \left[ \frac{q_i(x_i)}{n p_i(x_i)} \right] d\mu(x_i)$$

$$= \prod_{i=1}^{n} \int_x p_i(x_i) \left( \exp \frac{q_i(x_i)}{p_i(x_i)} \right)^{\frac{1}{n}} d\mu(x_i) = \prod_{i=1}^{n} D_{\exp} (p_i, q_i)$$

and the identity (2.11) is proved.

The sequence of inequalities (2.12) follows by (2.1).

**References**


A GENERALIZED $f$–DIVERGENCE


School of Communications and Informatics, Victoria University of Technology, PO Box 14428, Melbourne City MC, Victoria 8001, Australia.

E-mail address: sever.dragomir@vu.edu.au