A Generalization of f-Divergence Measure to Convex Functions Defined on Linear Spaces

A GENERALIZATION OF f-DIVERGENCE MEASURE TO CONVEX FUNCTIONS DEFINED ON LINEAR SPACES

S.S. DRAGOMIR

Abstract. In this paper we generalise the concept of f-divergence to a convex function defined on a convex cone in a linear space. Some fundamental results are established.

1. Introduction

Given a convex function $f:[0,\infty)\to\mathbb{R}$, the $f-$*divergence functional*

$$I_f(\mathbf{p},\mathbf{q})=\sum_{i=1}^{n}q_if\left(\frac{p_i}{q_i}\right), \tag{1.1}$$

was introduced by Csiszár [3]-[4] as a generalized measure of information, a *"distance function"* on the set of probability distribution $\mathbb{P}^n$. The restriction here to discrete distributions is only for convenience, similar results hold for general distributions. As in Csiszár [3]-[4] , we interpret undefined expressions by

$$f(0)=\lim_{t\to 0+}f(t),\quad 0\,f\left(\tfrac{0}{0}\right)=0,$$
$$0\,f\left(\tfrac{a}{0}\right)=\lim_{\varepsilon\to 0+}\varepsilon f\left(\tfrac{a}{\varepsilon}\right)=a\lim_{t\to\infty}\tfrac{f(t)}{t},\ a>0.$$

The following results were essentially given by Csiszár and Körner [5].

Proposition 1. *(Joint Convexity) If $f:[0,\infty)\to\mathbb{R}$ is convex, then $I_f(\mathbf{p},\mathbf{q})$ is jointly convex in p and q.*

Proposition 2. *(Jensen's inequality) Let $f:[0,\infty)\to\mathbb{R}$ be convex. Then for any $\mathbf{p},\mathbf{q}\in[0,\infty)^n$ with $P_n:=\sum_{i=1}^{n}p_i>0$, $Q_n:=\sum_{i=1}^{n}q_i>0$, we have the inequality*

$$I_f(\mathbf{p},\mathbf{q})\geq Q_nf\left(\frac{P_n}{Q_n}\right). \tag{1.2}$$

If f is strictly convex, equality holds in (1.2) iff

$$\frac{p_1}{q_1}=\frac{p_2}{q_2}=...=\frac{p_n}{q_n}.$$

It is natural to consider the following corollary.

Corollary 1. *(Nonnegativity) Let $f:[0,\infty)\to\mathbb{R}$ be convex and normalised, i.e.,*

$$f(1)=0. \tag{1.3}$$

Date: January 07, 2009.

1991 *Mathematics Subject Classification.* 26D15; 94.

Key words and phrases. Convex functions, Jensen's inequality, f-Divergences.

Then for any $\mathbf{p},\mathbf{q}\in[0,\infty)^n$ *with* $P_n=Q_n$*, we have the inequality*

$$I_f(\mathbf{p},\mathbf{q})\geq 0. \tag{1.4}$$

If f *is strictly convex, equality holds in (1.4) iff*

$$p_i=q_i \text{ for all } i\in\{1,...,n\}.$$

In particular, if $\mathbf{p},\mathbf{q}$ are probability vectors, then Corollary 1 shows that, for strictly convex and normalized $f:[0,\infty)\to\mathbb{R}$ that

$$I_f(\mathbf{p},\mathbf{q})\geq 0 \text{ and } I_f(\mathbf{p},\mathbf{q})=0 \text{ iff } p=q. \tag{1.5}$$

We now give some examples of divergence measures in Information Theory which are particular cases of $f-$divergences.

Kullback-Leibler distance ([14]). The *Kullback-Leibler distance* $D(\cdot,\cdot)$ is defined by

$$D(\mathbf{p},\mathbf{q}):=\sum_{i=1}^n p_i\log\left(\frac{p_i}{q_i}\right).$$

If we choose $f(t)=t\ln t$, $t>0$, then obviously

$$I_f(\mathbf{p},\mathbf{q})=D(\mathbf{p},\mathbf{q}).$$

Variational distance (l_1-distance). The *variational distance* $V(\cdot,\cdot)$ is defined by

$$V(\mathbf{p},\mathbf{q}):=\sum_{i=1}^n|p_i-q_i|.$$

If we choose $f(t)=|t-1|$, $t\in[0,\infty)$, then we have

$$I_f(\mathbf{p},\mathbf{q})=V(\mathbf{p},\mathbf{q}).$$

Hellinger discrimination ([1]). The *Hellinger discrimination* is defined by $\sqrt{2h^2(\cdot,\cdot)}$, where $h^2(\cdot,\cdot)$ is given by

$$h^2(\mathbf{p},\mathbf{q}):=\frac{1}{2}\sum_{i=1}^n\left(\sqrt{p_i}-\sqrt{q_i}\right)^2.$$

It is obvious that if $f(t)=\frac{1}{2}\left(\sqrt{t}-1\right)^2$, then

$$I_f(\mathbf{p},\mathbf{q})=h^2(\mathbf{p},\mathbf{q}).$$

Triangular discrimination ([17]). We define *triangular discrimination* between $\mathbf{p}$ and $\mathbf{q}$ by

$$\Delta(\mathbf{p},\mathbf{q})=\sum_{i=1}^n\frac{|p_i-q_i|^2}{p_i+q_i}.$$

It is obvious that if $f(t)=\frac{(t-1)^2}{t+1}$, $t\in(0,\infty)$, then

$$I_f(\mathbf{p},\mathbf{q})=\Delta(\mathbf{p},\mathbf{q}).$$

Note that $\sqrt{\Delta(\mathbf{p},\mathbf{q})}$ is known in the literature as the Le Cam distance.

χ^2-**distance.** We define the χ^2-*distance* (chi-square distance) by

$$D_{\chi^2}(\mathbf{p},\mathbf{q}):=\sum_{i=1}^n\frac{(p_i-q_i)^2}{q_i}.$$

It is clear that if $f(t) = (t-1)^2$, $t \in [0,\infty)$, then

$$I_f(\mathbf{p},\mathbf{q}) = D_{\chi^2}(\mathbf{p},\mathbf{q}).$$

Rényi's divergences ([16]). For $\alpha \in \mathbb{R}\setminus\{0,1\}$, consider

$$\rho_\alpha(\mathbf{p},\mathbf{q}) := \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}.$$

It is obvious that if $f(t) = t^\alpha$ $(t \in (0,\infty))$, then

$$I_f(\mathbf{p},\mathbf{q}) = \rho_\alpha(\mathbf{p},\mathbf{q}).$$

Rényi's divergences $R_\alpha(\mathbf{p},\mathbf{q}) := \frac{1}{\alpha(\alpha-1)}\ln[\rho_\alpha(\mathbf{p},\mathbf{q})]$ have been introduced for all real orders $\alpha \neq 0$, $\alpha \neq 1$ (and continuously extended for $\alpha = 0$ and $\alpha = 1$) in [15], where the reader may find many inequalities valid for these divergences, without, as well as with, some restrictions for $\mathbf{p}$ and $\mathbf{q}$.

For other examples of divergence measures, see the paper [12] and the books [15] and [18], where further references are given.

In this paper we generalize the concept of f-divergence to a convex function defined on a convex cone in a linear space. Some fundamental results are established.

2. The f-Divergence of an n-tuple of Vectors

Firstly, we recall that the subset K in a linear space X is a *cone* if the following two conditions are satisfied:

(i) for any $x, y \in K$ we have $x + y \in K$;

(ii) for any $x \in K$ and any $\alpha \geq 0$ we have $\alpha x \in K$.

For a given n-tuple of vectors $\mathbf{z} = (z_1, ..., z_n) \in K^n$ and a probability distribution $\mathbf{q} = (q_1, ..., q_n) \in \mathbb{P}^n$ with all values nonzero, we can define, for the convex function $f : K \to \mathbb{R}$, the following *f-divergence of* $\mathbf{z}$ *with the distribution* $\mathbf{q}$ (see [8]):

$$I_f(\mathbf{z},\mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{z_i}{q_i}\right). \tag{2.1}$$

It is obvious that if $X = \mathbb{R}$, $K = [0,\infty)$ and $\mathbf{x} = \mathbf{p} \in \mathbb{P}^n$ then we obtain the usual concept of the f-divergence associated with a function $f : [0,\infty) \to \mathbb{R}$.

The following result concerning the mutual convexity of the f-divergence holds.

Theorem 1. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then the function $I_f(\cdot,\cdot)$ is convex on the convex set $K^n \times \mathbb{P}^n$.*

Proof. Let $\mathbf{z} = (z_1, ..., z_n), \mathbf{v} = (v_1, ..., v_n) \in K^n$, $\mathbf{p} = (p_1, ..., p_n), \mathbf{q} = (q_1, ..., q_n) \in \mathbb{P}^n$ two probability distributions with all values nonzero and $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$. Then we have

$$\begin{aligned} I_f[\alpha(\mathbf{v},\mathbf{p}) + \beta(\mathbf{z},\mathbf{q})] &= I_f(\alpha\mathbf{v} + \beta\mathbf{z}, \alpha\mathbf{p} + \beta\mathbf{q}) \\ &= \sum_{i=1}^{n} (\alpha p_i + \beta q_i) f\left(\frac{\alpha v_i + \beta z_i}{\alpha p_i + \beta q_i}\right) \\ &= \sum_{i=1}^{n} (\alpha p_i + \beta q_i) f\left[\left(\frac{\alpha p_i}{\alpha p_i + \beta q_i}\right)\cdot\frac{v_i}{p_i} + \left(\frac{\beta q_i}{\alpha p_i + \beta q_i}\right)\cdot\frac{z_i}{q_i}\right]. \end{aligned} \tag{2.2}$$

Due to the convexity of f, we have

$$f\left[\left(\frac{\alpha p_i}{\alpha p_i+\beta q_i}\right)\cdot\frac{v_i}{p_i}+\left(\frac{\beta q_i}{\alpha p_i+\beta q_i}\right)\cdot\frac{z_i}{q_i}\right] \leq \frac{\alpha p_i}{\alpha p_i+\beta q_i}\cdot f\left(\frac{v_i}{p_i}\right)+\frac{\beta q_i}{\alpha p_i+\beta q_i}\cdot f\left(\frac{z_i}{q_i}\right) \tag{2.3}$$

for each $i\in\{1,...,n\}$.

Now, on multiplying (2.3) with $\alpha p_i+\beta q_i>0$, summing over i from 1 to n and utilising (2.2) we get that

$$I_f\left[\alpha(\mathbf{v},\mathbf{p})+\beta(\mathbf{z},\mathbf{q})\right]\leq\alpha I_f(\mathbf{v},\mathbf{p})+\beta I_f(\mathbf{z},\mathbf{q})$$

proving the desired result. □

Now, for a given n-tuple of vectors $\mathbf{x}=(x_1,...,x_n)\in K^n$, a probability distribution $\mathbf{q}\in\mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1,...,n\}$ we have

$$\mathbf{q}_J:=\left(Q_J,\bar{Q}_J\right)\in\mathbb{P}^2$$

where $Q_J:=\sum_{j\in J}q_j, \bar{Q}_J:=1-Q_J$ and

$$\mathbf{x}_J:=\left(X_J,\bar{X}_J\right)\in K^2$$

where, as above,

$$X_J:=\sum_{i\in J}x_i, \quad\text{and}\quad \bar{X}_J:=X_{\bar{J}}.$$

It is obvious that

$$I_f(\mathbf{x}_J,\mathbf{q}_J)=Q_Jf\left(\frac{X_J}{Q_J}\right)+\bar{Q}_Jf\left(\frac{\bar{X}_J}{\bar{Q}_J}\right).$$

The following inequality for the f-divergence of an n-tuple of vectors in a linear space holds [8]:

Theorem 2. *Let $f:K\to\mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x}=(x_1,...,x_n)\in K^n$, a probability distribution $\mathbf{q}\in\mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1,...,n\}$ we have*

$$I_f(\mathbf{x},\mathbf{q})\geq\max_{\emptyset\neq J\subset\{1,...,n\}}I_f(\mathbf{x}_J,\mathbf{q}_J)\geq I_f(\mathbf{x}_J,\mathbf{q}_J) \geq\min_{\emptyset\neq J\subset\{1,...,n\}}I_f(\mathbf{x}_J,\mathbf{q}_J)\geq f(X_n) \tag{2.4}$$

where $X_n:=\sum_{i=1}^n x_i$.

We observe that, for a given n-tuple of vectors $\mathbf{x}=(x_1,...,x_n)\in K^n$, a sufficient condition for the positivity of $I_f(\mathbf{x},\mathbf{q})$ for any probability distribution $\mathbf{q}\in\mathbb{P}^n$ with all values nonzero is that $f(X_n)\geq 0$. In the scalar case and if $\mathbf{x}=\mathbf{p}\in\mathbb{P}^n$, then a sufficient condition for the positivity of the f-divergence $I_f(\mathbf{p},\mathbf{q})$ is that $f(1)\geq 0$.

The case of functions of a real variable that is of interest for applications is incorporated in [8]:

Corollary 2. *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function. Then for any $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ we have*

$$I_f(\mathbf{p},\mathbf{q})\geq\max_{\emptyset\neq J\subset\{1,...,n\}}\left[Q_Jf\left(\frac{P_J}{Q_J}\right)+(1-Q_J)f\left(\frac{1-P_J}{1-Q_J}\right)\right](\geq 0). \tag{2.5}$$

Remark 1. *For various applications of the inequality (2.5) to particular divergence measures of interest in applications, see* [8]*. In order to give an example, we point out the following result*

$$J(p,q) \geq \ln\left(\max_{\emptyset\neq J\subset\{1,\dots,n\}}\left\{\left[\frac{(1-P_J)Q_J}{(1-Q_J)P_J}\right]^{(Q_J-P_J)}\right\}\right) \geq \max_{\emptyset\neq J\subset\{1,\dots,n\}}\left[\frac{(Q_J-P_J)^2}{P_J+Q_J-2P_JQ_J}\right] \geq 0, \tag{2.6}$$

where the Jeffreys divergence is defined as

$$J(p,q) := \sum_{j=1}^{n} q_j\cdot\left(\frac{p_j}{q_j}-1\right)\ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n}(p_j-q_j)\ln\left(\frac{p_j}{q_j}\right), \tag{2.7}$$

which is an f-divergence for $f(t)=(t-1)\ln t$, $t>0$.

3. Some Upper and Lower Bounds

Let K be a convex subset of the real linear space X and let $f: K \to \mathbb{R}$ be a convex mapping. Here we consider the following well-known form of Jensen's discrete inequality:

$$f\left(\frac{1}{P_I}\sum_{i\in I}p_ix_i\right) \leq \frac{1}{P_I}\sum_{i\in I}p_if(x_i),$$

where I denotes a finite subset of the set $\mathbb{N}$ of natural numbers, $x_i \in K$, $p_i \geq 0$ for $i \in I$ and $P_I := \sum_{i\in I} p_i > 0$.

Let us fix $I \in \mathcal{P}_f(\mathbb{N})$ (the class of finite parts of $\mathbb{N}$) and $x_i \in K$ $(i \in I)$. Now consider the functional $J : S_+(I) \to \mathbb{R}$ given by

$$J_I(\mathbf{p}) := \sum_{i\in I}p_if(x_i) - P_If\left(\frac{1}{P_I}\sum_{i\in I}p_ix_i\right) \geq 0$$

where $S_+(I) := \left\{\mathbf{p} = (p_i)_{i\in I}\middle|\, p_i \geq 0,\ i\in I \text{ and } P_I > 0\right\}$ and f is convex on K.

We observe that $S_+(I)$ is a cone and the functional J_I is nonnegative, superadditive [10] and positive homogeneous on $S_+(I)$.

We have the following inequalities that are of interest in their turn as well (see [9]):

Lemma 1. *If $\mathbf{p},\mathbf{q} \in S_+(I)$ and $M \geq m \geq 0$ such that $M\mathbf{p} \geq \mathbf{q} \geq m\mathbf{p}$, i.e., $Mp_i \geq q_i \geq mp_i$ for each $i \in I$, then:*

$$M\left[\sum_{i\in I}p_if(x_i) - P_If\left(\frac{1}{P_I}\sum_{i\in I}p_ix_i\right)\right] \geq \sum_{i\in I}q_if(x_i) - Q_If\left(\frac{1}{Q_I}\sum_{i\in I}q_ix_i\right) \geq m\left[\sum_{i\in I}p_if(x_i) - P_If\left(\frac{1}{P_I}\sum_{i\in I}p_ix_i\right)\right] \quad (\geq 0). \tag{3.1}$$

and

$$\left[\frac{1}{P_I}\sum_{i\in I}p_i f(x_i) - f\left(\frac{1}{P_I}\sum_{i\in I}p_i x_i\right)\right]^{MP_I} \tag{3.2}$$
$$\geq \left[\frac{1}{Q_I}\sum_{i\in I}q_i f(x_i) - f\left(\frac{1}{Q_I}\sum_{i\in I}q_i x_i\right)\right]^{Q_I}$$
$$\geq \left[\frac{1}{P_I}\sum_{i\in I}p_i f(x_i) - f\left(\frac{1}{P_I}\sum_{i\in I}p_i x_i\right)\right]^{mP_I}.$$

respectively.

We may state the following result:

Theorem 3. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Consider an n-tuple of vectors $\mathbf{z} = (z_1, ..., z_n) \in K^n$ and two probability distribution $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ with all values nonzero and satisfying the condition*

$$Rp_i \geq q_i \geq rp_i \text{ for each } i \in \{1, ..., n\}, \tag{3.3}$$

where $R \geq 1 \geq r > 0$.

If we define the vector

$$\mathbf{y} = \left(\frac{p_1}{q_1}z_1, ..., \frac{p_n}{q_n}z_n\right) \in K^n,$$

then we have the inequalities

$$R\left[I_f(\mathbf{y},\mathbf{p}) - f(Y_n)\right] \geq I_f(\mathbf{z},\mathbf{q}) - f(Z_n) \geq r\left[I_f(\mathbf{y},\mathbf{p}) - f(Y_n)\right] (\geq 0) \tag{3.4}$$

and the inequalities

$$\left[I_f(\mathbf{y},\mathbf{p}) - f(Y_n)\right]^R \geq I_f(\mathbf{z},\mathbf{q}) - f(Z_n) \geq \left[I_f(\mathbf{y},\mathbf{p}) - f(Y_n)\right]^r (\geq 0) \tag{3.5}$$

respectively, where $Z_n := \sum_{i=1}^n z_i$ and $Y_n := \sum_{i=1}^n y_i = \sum_{i=1}^n \frac{p_i}{q_i}\cdot z_i \in K$.

The proof follows from Lemma 1 applied for $M = R, m = r$ and $x_i = \frac{z_i}{q_i}$ where $i \in \{1, ..., n\}$.

Corollary 3. *Let $f : [0,\infty) \to \mathbb{R}$ be a normalized convex function. For two probability distributions $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ with all values nonzero assume that there exists the constants $R \geq 1 \geq r > 0$ satisfying the condition (3.3).*

If $\mathbf{s} = (s_1, ..., s_n) \in \mathbb{P}^n$ is such that the vector

$$\mathbf{y} = \left(\frac{p_1}{q_1}s_1, ..., \frac{p_n}{q_n}s_n\right) \in \mathbb{R}^n_+ \tag{3.6}$$

is a probability distribution, then we have the inequalities

$$RI_f(\mathbf{y},\mathbf{p}) \geq I_f(\mathbf{s},\mathbf{q}) \geq rI_f(\mathbf{y},\mathbf{p}) \tag{3.7}$$

and the inequalities

$$\left[I_f(\mathbf{y},\mathbf{p})\right]^R \geq I_f(\mathbf{z},\mathbf{q}) \geq \left[I_f(\mathbf{y},\mathbf{p})\right]^r. \tag{3.8}$$

Remark 2. *It is natural to ask if we can find probability distributions* $\mathbf{p},\mathbf{q},\mathbf{s}\in\mathbb{P}^n$ *such that* $\mathbf{y}$ *defined by (3.6) is a probability distribution as well.*

Let consider the simplest example, namely for $n=2$*. In this case for, say* $\mathbf{p}=(0.1,0.9)$*,* $\mathbf{q}=(0.2,0.8)$ *and* $\mathbf{s}=(s_1,s_2)\in\mathbb{P}^2$ *we have* $\mathbf{y}=\left(\frac{1}{2}s_1,\frac{9}{8}s_2\right)$ *which should satisfy the condition that* $\frac{1}{2}s_1+\frac{9}{8}s_2=1$ *for some* $s_1,s_2\in[0,1]$ *with* $s_1+s_2=1$*. We observe that this system of equations has the unique solution* $s_1=\frac{1}{5}$ *and* $s_2=\frac{4}{5}$*, showing that* $(s_1,s_2)\in\mathbb{P}^2$*.*

4. Other Bounds in Terms of Gâteau Derivatives

Assume that $f:X\to\mathbb{R}$ is a *convex function* on the real linear space X. Since for any vectors $x,y\in X$ the function $g_{x,y}:\mathbb{R}\to\mathbb{R}$, $g_{x,y}(t):=f(x+ty)$ is convex it follows that the following limits exist

$$\nabla_{+(-)}f(x)(y):=\lim_{t\to 0+(-)}\frac{f(x+ty)-f(x)}{t} \tag{4.1}$$

and they are called the *right(left) Gâteaux derivatives* of the function f in the point x over the direction y.

It is obvious that for any $t>0>s$ we have

$$\begin{aligned}\frac{f(x+ty)-f(x)}{t}&\geq\nabla_+f(x)(y)=\inf_{t>0}\left[\frac{f(x+ty)-f(x)}{t}\right]\\&\geq\sup_{s<0}\left[\frac{f(x+sy)-f(x)}{s}\right]=\nabla_-f(x)(y)\geq\frac{f(x+sy)-f(x)}{s}\end{aligned} \tag{4.2}$$

for any $x,y\in X$ and, in particular,

$$\nabla_-f(u)(u-v)\geq f(u)-f(v)\geq\nabla_+f(v)(u-v) \tag{4.3}$$

for any $u,v\in X$. We call this *the gradient inequality* for the convex function f. It will be used frequently in the sequel in order to obtain various results related to Jensen's inequality.

The following properties are also of importance:

$$\nabla_+f(x)(-y)=-\nabla_-f(x)(y), \tag{4.4}$$

and

$$\nabla_{+(-)}f(x)(\alpha y)=\alpha\nabla_{+(-)}f(x)(y) \tag{4.5}$$

for any $x,y\in X$ and $\alpha\geq 0$.

The right Gâteaux derivative is *subadditive* while the left one is *superadditive*, i.e.,

$$\nabla_+f(x)(y+z)\leq\nabla_+f(x)(y)+\nabla_+f(x)(z) \tag{4.6}$$

and

$$\nabla_-f(x)(y+z)\geq\nabla_-f(x)(y)+\nabla_-f(x)(z) \tag{4.7}$$

for any $x,y,z\in X$.

Some natural examples can be provided by the use of normed spaces.

Assume that $(X,\|\cdot\|)$ is a real normed linear space. The function $f:X\to\mathbb{R}$, $f(x):=\frac{1}{2}\|x\|^2$ is a convex function which generates *the superior* and *the inferior* *semi-inner products*

$$\langle y,x\rangle_{s(i)} := \lim_{t\to 0+(-)} \frac{\|x+ty\|^2 - \|x\|^2}{t}.$$

For a comprehensive study of the properties of these mappings in the Geometry of Banach Spaces see the monograph [7].

For the convex function $f_p : X \to \mathbb{R}$, $f_p(x) := \|x\|^p$ with $p > 1$, we have

$$\nabla_{+(-)} f_p(x)(y) = \begin{cases} p\|x\|^{p-2}\langle y,x\rangle_{s(i)} & \text{if } x \neq 0 \\ \\ 0 & \text{if } x = 0 \end{cases}$$

for any $y \in X$.

If $p = 1$, then we have

$$\nabla_{+(-)} f_1(x)(y) = \begin{cases} \|x\|^{-1}\langle y,x\rangle_{s(i)} & \text{if } x \neq 0 \\ \\ +(-)\|y\| & \text{if } x = 0 \end{cases}$$

for any $y \in X$.

This class of functions will be used to illustrate the inequalities obtained in the general case of convex functions defined on an entire linear space.

The following result holds:

Lemma 2. *Let $f : X \to \mathbb{R}$ be a convex function. Then for any $x, y \in X$ and $t \in [0,1]$ we have*

$$\begin{aligned} (4.8)\quad & t(1-t)[\nabla_- f(y)(y-x) - \nabla_+ f(x)(y-x)] \\ & \geq tf(x) + (1-t)f(y) - f(tx+(1-t)y) \\ & \geq t(1-t)[\nabla_+ f(tx+(1-t)y)(y-x) - \nabla_- f(tx+(1-t)y)(y-x)] \geq 0. \end{aligned}$$

Proof. Utilising the gradient inequality (4.3) we have

$$f(tx+(1-t)y) - f(x) \geq (1-t)\nabla_+ f(x)(y-x) \tag{4.9}$$

and

$$f(tx+(1-t)y) - f(y) \geq -t\nabla_- f(y)(y-x). \tag{4.10}$$

If we multiply (4.9) with t and (4.10) with $1-t$ and add the resultant inequalities we obtain

$$\begin{aligned} & f(tx+(1-t)y) - tf(x) - (1-t)f(y) \\ & \geq (1-t)t\nabla_+ f(x)(y-x) - t(1-t)\nabla_- f(y)(y-x) \end{aligned}$$

which is clearly equivalent with the first part of (4.8).

By the gradient inequality we also have

$$(1-t)\nabla_- f(tx+(1-t)y)(y-x) \geq f(tx+(1-t)y) - f(x)$$

and

$$-t\nabla_+ f(tx+(1-t)y)(y-x) \geq f(tx+(1-t)y) - f(y)$$

which by the same procedure as above yields the second part of (4.8). □

Theorem 4. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. If $\mathbf{z} = (z_1, ..., z_n)$, $\mathbf{v} = (v_1, ..., v_n) \in K^n$, $\mathbf{p} = (p_1, ..., p_n)$, $\mathbf{q} = (q_1, ..., q_n) \in \mathbb{P}^n$ are two probability distributions with all values nonzero and $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$, then we have*

$$\begin{aligned} (4.11) \qquad & \alpha\beta \sum_{i=1}^{n} \frac{p_i q_i}{\alpha p_i + \beta q_i} \left[\nabla_- f\left(\frac{z_i}{q_i}\right)\left(\frac{z_i}{q_i} - \frac{v_i}{p_i}\right) - \nabla_+ f\left(\frac{v_i}{p_i}\right)\left(\frac{z_i}{q_i} - \frac{v_i}{p_i}\right)\right] \\ & \geq \alpha I_f(\mathbf{v}, \mathbf{p}) + \beta I_f(\mathbf{z}, \mathbf{q}) - I_f[\alpha(\mathbf{v}, \mathbf{p}) + \beta(\mathbf{z}, \mathbf{q})] \\ & \geq \alpha\beta \sum_{i=1}^{n} \frac{p_i q_i}{\alpha p_i + \beta q_i} \\ & \times \left[\nabla_+ f\left(\frac{\alpha v_i + \beta z_i}{\alpha p_i + \beta q_i}\right)\left(\frac{z_i}{q_i} - \frac{v_i}{p_i}\right) - \nabla_- f\left(\frac{\alpha v_i + \beta z_i}{\alpha p_i + \beta q_i}\right)\left(\frac{z_i}{q_i} - \frac{v_i}{p_i}\right)\right] \\ & \geq 0 \end{aligned}$$

Proof. If we write the inequality (4.8) for

$$x = \frac{v_i}{p_i}, y = \frac{z_i}{q_i} \text{ and } t = \frac{\alpha p_i}{\alpha p_i + \beta q_i}$$

then we get

$$\begin{aligned} (4.12) \qquad & \frac{\alpha\beta p_i q_i}{(\alpha p_i + \beta q_i)^2} \left[\nabla_- f\left(\frac{z_i}{q_i}\right)\left(\frac{z_i}{q_i} - \frac{v_i}{p_i}\right) - \nabla_+ f\left(\frac{v_i}{p_i}\right)\left(\frac{z_i}{q_i} - \frac{v_i}{p_i}\right)\right] \\ & \geq \frac{\alpha p_i}{\alpha p_i + \beta q_i} f\left(\frac{v_i}{p_i}\right) + \frac{\beta q_i}{\alpha p_i + \beta q_i} f\left(\frac{z_i}{q_i}\right) - f\left(\frac{\alpha v_i + \beta z_i}{\alpha p_i + \beta q_i}\right) \\ & \geq \frac{\alpha\beta p_i q_i}{(\alpha p_i + \beta q_i)^2} \\ & \times \left[\nabla_+ f\left(\frac{\alpha v_i + \beta z_i}{\alpha p_i + \beta q_i}\right)\left(\frac{z_i}{q_i} - \frac{v_i}{p_i}\right) - \nabla_- f\left(\frac{\alpha v_i + \beta z_i}{\alpha p_i + \beta q_i}\right)\left(\frac{z_i}{q_i} - \frac{v_i}{p_i}\right)\right] \\ & \geq 0, \end{aligned}$$

for each $i \in \{1, ..., n\}$.

Now, if we multiply (4.12) by $\alpha p_i + \beta q_i > 0$ and sum over i from 1 to n we derive the desired result (4.11). □

It is natural now to consider the corresponding result for convex functions of a real variable.

Corollary 4. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function. If $\mathbf{z} = (z_1, ..., z_n)$, $\mathbf{v} = (v_1, ..., v_n)$, $\mathbf{p} = (p_1, ..., p_n)$, $\mathbf{q} = (q_1, ..., q_n) \in \mathbb{P}^n$ are probability distributions with all values nonzero and $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$, then we have*

$$\begin{aligned} (4.13) \qquad & \alpha\beta \sum_{i=1}^{n} \frac{\det\begin{bmatrix} z_i & v_i \\ q_i & p_i \end{bmatrix}}{\alpha p_i + \beta q_i} \left[f'_-\left(\frac{z_i}{q_i}\right) - f'_+\left(\frac{v_i}{p_i}\right)\right] \\ & \geq \alpha I_f(\mathbf{v}, \mathbf{p}) + \beta I_f(\mathbf{z}, \mathbf{q}) - I_f[\alpha(\mathbf{v}, \mathbf{p}) + \beta(\mathbf{z}, \mathbf{q})] \\ & \geq \alpha\beta \sum_{i=1}^{n} \frac{\det\begin{bmatrix} z_i & v_i \\ q_i & p_i \end{bmatrix}}{\alpha p_i + \beta q_i} \left[f'_+\left(\frac{\alpha v_i + \beta z_i}{\alpha p_i + \beta q_i}\right) - f'_-\left(\frac{\alpha v_i + \beta z_i}{\alpha p_i + \beta q_i}\right)\right] \\ & \geq 0. \end{aligned}$$

Remark 3. *It is obvious that for differentiable convex functions on $(0,\infty)$ the lower bound vanishes and the inequality (4.13) becomes:*

$$0 \leq \alpha I_f(\mathbf{v},\mathbf{p}) + \beta I_f(\mathbf{z},\mathbf{q}) - I_f[\alpha(\mathbf{v},\mathbf{p}) + \beta(\mathbf{z},\mathbf{q})] \leq \alpha\beta \sum_{i=1}^{n} \frac{\det\begin{bmatrix} z_i & v_i \\ q_i & p_i \end{bmatrix}}{\alpha p_i + \beta q_i} \left[f'\left(\frac{z_i}{q_i}\right) - f'\left(\frac{v_i}{p_i}\right)\right] \tag{4.14}$$

that can be used for particular divergence measures.

Indeed, if we consider the normalised convex function $f(t) = (t-1)^2$, $t \in [0,\infty)$, then

$$I_f(\mathbf{p},\mathbf{q}) = D_{\chi^2}(\mathbf{p},\mathbf{q})$$

where, as in the introduction, the χ^2-distance (chi-square distance) is defined by

$$D_{\chi^2}(\mathbf{p},\mathbf{q}) := \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}.$$

It is clear that the inequality (4.14) becomes then

$$0 \leq \alpha D_{\chi^2}(\mathbf{v},\mathbf{p}) + \beta D_{\chi^2}(\mathbf{z},\mathbf{q}) - D_{\chi^2}[\alpha(\mathbf{v},\mathbf{p}) + \beta(\mathbf{z},\mathbf{q})] \leq 2\alpha\beta \sum_{i=1}^{n} \frac{\det^2\begin{bmatrix} z_i & v_i \\ q_i & p_i \end{bmatrix}}{p_i q_i (\alpha p_i + \beta q_i)}. \tag{4.15}$$

The Kullback-Leibler distance $D(\cdot,\cdot)$ is defined by

$$D(\mathbf{p},\mathbf{q}) := \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right).$$

If we choose $f(t) = t\ln t$, $t > 0$, then obviously

$$I_f(\mathbf{p},\mathbf{q}) = D(\mathbf{p},\mathbf{q})$$

and the inequality (4.14) becomes then

$$0 \leq \alpha D(\mathbf{v},\mathbf{p}) + \beta D(\mathbf{z},\mathbf{q}) - D[\alpha(\mathbf{v},\mathbf{p}) + \beta(\mathbf{z},\mathbf{q})] \leq \alpha\beta \ln\left\{\prod_{i=1}^{n}\left[\left(\frac{z_i p_i}{q_i v_i}\right)^{\frac{p_i z_i - q_i v_i}{\alpha p_i + \beta q_i}}\right]\right\}. \tag{4.16}$$

Similar results could be obtained for other particular instances of divergence measures, however the details are omitted.

In what follows we provide some lower and upper bounds for the nonnegative difference $I_f(\mathbf{x},\mathbf{q}) - I_f(\mathbf{x}_J,\mathbf{q}_J)$ where J is a nonempty subset of $\{1,...,n\}$ and

$$I_f(\mathbf{x}_J,\mathbf{q}_J) = Q_J f\left(\frac{X_J}{Q_J}\right) + \bar{Q}_J f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right).$$

For a nonempty subset K of $\{1,...,n\}$ we also use the notation

$$I_{f,K}(\mathbf{x},\mathbf{q}) := \sum_{i\in K} q_i f\left(\frac{x_i}{q_i}\right).$$

Theorem 5. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1, ..., n\}$ we have*

$$(4.17)\quad I_{\nabla_- f(\cdot)\left(\cdot - \frac{X_J}{Q_J}\right), J}(\mathbf{x}, \mathbf{q}) + I_{\nabla_- f(\cdot)\left(\cdot - \frac{\bar{X}_J}{\bar{Q}_J}\right), \bar{J}}(\mathbf{x}, \mathbf{q}) \geq I_f(\mathbf{x}, \mathbf{q}) - I_f(\mathbf{x}_J, \mathbf{q}_J)$$
$$\geq I_{\nabla_+ f\left(\frac{X_J}{Q_J}\right)\left(\cdot - \frac{X_J}{Q_J}\right), J}(\mathbf{x}, \mathbf{q}) + I_{\nabla_+ f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right)\left(\cdot - \frac{\bar{X}_J}{\bar{Q}_J}\right), \bar{J}}(\mathbf{x}, \mathbf{q}) \geq 0$$

Proof. Utilising the gradient inequality we have, for a given nonempty set J of $\{1, ..., n\}$ with $J \neq \{1, ..., n\}$, that

$$(4.18)\quad \nabla_- f\left(\frac{x_i}{q_i}\right)\left(\frac{x_i}{q_i} - \frac{X_J}{Q_J}\right) \geq f\left(\frac{x_i}{q_i}\right) - f\left(\frac{X_J}{Q_J}\right)$$
$$\geq \nabla_+ f\left(\frac{X_J}{Q_J}\right)\left(\frac{x_i}{q_i} - \frac{X_J}{Q_J}\right)$$

for any $i \in J$. If we multiply (4.18) with $q_i \geq 0$ and sum over $i \in J$, we get

$$(4.19)\quad I_{\nabla_- f(\cdot)\left(\cdot - \frac{X_J}{Q_J}\right), J}(\mathbf{x}, \mathbf{q}) \geq I_{f,J}(\mathbf{x}, \mathbf{q}) - Q_J f\left(\frac{X_J}{Q_J}\right)$$
$$\geq I_{\nabla_+ f\left(\frac{X_J}{Q_J}\right)\left(\cdot - \frac{X_J}{Q_J}\right), J}(\mathbf{x}, \mathbf{q}) \geq 0.$$

From the gradient inequality we also have

$$(4.20)\quad \nabla_- f\left(\frac{x_j}{q_j}\right)\left(\frac{x_j}{q_j} - \frac{\bar{X}_J}{\bar{Q}_J}\right) \geq f\left(\frac{x_j}{q_j}\right) - f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right)$$
$$\geq \nabla_+ f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right)\left(\frac{x_j}{q_j} - \frac{\bar{X}_J}{\bar{Q}_J}\right)$$

for any $j \in \bar{J}$. If we multiply (4.18) with $q_j \geq 0$ and sum over $j \in \bar{J}$, we get

$$(4.21)\quad I_{\nabla_- f(\cdot)\left(\cdot - \frac{\bar{X}_J}{\bar{Q}_J}\right), \bar{J}}(\mathbf{x}, \mathbf{q}) \geq I_{f,\bar{J}}(\mathbf{x}, \mathbf{q}) - \bar{Q}_J f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right)$$
$$\geq I_{\nabla_+ f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right)\left(\cdot - \frac{\bar{X}_J}{\bar{Q}_J}\right), \bar{J}}(\mathbf{x}, \mathbf{q}) \geq 0.$$

Now, if we sum the inequalities (4.19) with (4.21) and take into account that

$$I_{f,J}(\mathbf{x}, \mathbf{q}) + I_{f,\bar{J}}(\mathbf{x}, \mathbf{q}) = I_f(\mathbf{x}, \mathbf{q})$$

and

$$Q_J f\left(\frac{X_J}{Q_J}\right) + \bar{Q}_J f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right) = I_f(\mathbf{x}_J, \mathbf{q}_J)$$

then we get the desired result (4.17). □

The case of functions of a real variable that is of interest for applications is incorporated in :

Corollary 5. *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function. Then for any $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ and $\emptyset\neq J\subset\{1,...,n\}$ we have*

$$
\begin{aligned}
(4.22)\qquad & I_{f'_-(\cdot)\left(\cdot-\frac{P_J}{Q_J}\right),J}(\mathbf{p},\mathbf{q})+I_{f'_-(\cdot)\left(\cdot-\frac{1-P_J}{1-Q_J}\right),\bar{J}}(\mathbf{p},\mathbf{q}) \\
& \geq I_f(\mathbf{p},\mathbf{q})-Q_J f\left(\frac{P_J}{Q_J}\right)-(1-Q_J)f\left(\frac{1-P_J}{1-Q_J}\right) \\
& \geq I_{f'_+\left(\frac{X_J}{Q_J}\right)\left(\cdot-\frac{P_J}{Q_J}\right),J}(\mathbf{p},\mathbf{q})+I_{f'_+\left(\frac{1-P_J}{1-Q_J}\right)\left(\cdot-\frac{1-P_J}{1-Q_J}\right),\bar{J}}(\mathbf{p},\mathbf{q})\geq 0.
\end{aligned}
$$

Remark 4. *If one chooses different convex functions generating particular divergence measures such as the Kullback-Leibler, Jeffreys or Hellinger divergences, that one can obtain some particular results of interest. However the details are not presented here.*

References

[1] R. BERAN, Minimum Hellinger distance estimates for parametric models, *Ann. Statist.,* **5** (1977), 445-463.

[2] A. BHATTACHARYYA, On a measure of divergence between two statistical populations defined by their probability distributions, *Bull. Calcutta Math. Soc.*, **35** (1943), 99-109.

[3] I. CSISZÁR, Information measures: A critical survey, *Trans. 7th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Processes and 8th European Meeting of Statist.,* Volume B, Academia Prague, 1978, 73-86.

[4] I. CSISZÁR, Information-type measures of difference of probability functions and indirect observations, *Studia Sci. Math. Hungar,* **2** (1967), 299-318.

[5] I. CSISZÁR and J. KÖRNER, *Information Theory: Coding Theorem for Discrete Memoryless Systems,* Academic Press, New York, 1981.

[6] D. DACUNHA-CASTELLE, Ecole d'été de Probabilités de Saint-Fleour, III-1997, Berlin, Heidelberg: Springer 1978.

[7] S.S. DRAGOMIR, *Semi-inner Products and Applications*, Nova Science Publishers Inc., NY, 2004.

[8] S.S. DRAGOMIR, A new refinement of Jensen's inequality in linear spaces with applications, Preprint *RGMIA Res. Rep. Coll.* **12**(2009), Supplement, Article 6. `[Online http://www.staff.vu.edu.au/RGMIA/v12(E).asp]`.

[9] S.S. DRAGOMIR, Inequalities for superadditive functionals with applications, *Bull. Austral. Math Soc.* **77**(2008), 401-411.

[10] S.S. DRAGOMIR, J. PEČARIĆ and L.E. PERSSON, Properties of some functionals related to Jensen's inequality, *Acta Math. Hungarica,* **70** (1996), 129-143.

[11] H. JEFFREYS, An invariant form for the prior probability in estimation problems, *Proc. Roy. Soc. London, Ser. A*, **186** (1946), 453-461.

[12] J.N. KAPUR, A comparative assessment of various measures of directed divergence, *Advances in Management Studies,* **3** (1984), No. 1, 1-16.

[13] S. KULLBACK, *Information Theory and Statistics,* J. Wiley, New York, 1959.

[14] S. KULLBACK and R.A. LEIBLER, On information and sufficiency, *Annals Math. Statist.,* **22** (1951), 79-86.

[15] F. LIESE and I. VAJDA, *Convex Statistical Distances,* Teubner Verlag, Leipzig, 198

[16] A. RENYI, On measures of entropy and information, *Proc. Fourth Berkeley Symp. Math. Statist. Prob., Vol. 1, University of California Press, Berkeley, 1961.*

[17] F. TOPSOE, Some inequalities for information divergence and related measures of discrimination, *Res. Rep. Coll. RGMIA*, **2** (1) (1999), 85-98.7.

[18] I. VAJDA, *Theory of Statistical Inference and Information,* Kluwer, Boston, 1989.

Research Group in Mathematical Inequalities & Applications, School of Engineering & Science, Victoria University, PO Box 14428, Melbourne City, MC 8001, Australia.

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

URL: `http://www.staff.vu.edu.au/rgmia/dragomir/`