A New Refinement of Jensen's Inequality in Linear Spaces with Applications

A NEW REFINEMENT OF JENSEN'S INEQUALITY IN LINEAR SPACES WITH APPLICATIONS

S.S. DRAGOMIR

Abstract. A new refinement of Jensen's celebrated inequality for functions defined on convex sets in linear spaces is given. Applications for norms, mean f-deviation and f-divergences are provided as well.

1. Introduction

Let C be a convex subset of the linear space X and f a convex function on C. If $\mathbf{p} = (p_1, \ldots, p_n)$ is a probability sequence and $\mathbf{x} = (x_1, \ldots, x_n) \in C^n$, then

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i), \tag{1.1}$$

is well known in the literature as *Jensen's inequality.*

The Jensen inequality for convex functions plays a crucial role in the Theory of Inequalities due to the fact that other inequalities such as the arithmetic mean-geometric mean inequality, Hölder and Minkowski inequalities, Ky Fan's inequality etc. can be obtained as particular cases of it.

In 1989, J. Pečarić and the author obtained the following refinement of (1.1) (see [13]):

$$\begin{aligned} f\left(\sum_{i=1}^{n} p_i x_i\right) &\leq \sum_{i_1,\ldots,i_{k+1}=1}^{n} p_{i_1} \ldots p_{i_{k+1}} f\left(\frac{x_{i_1} + \cdots + x_{i_{k+1}}}{k+1}\right) \\ &\leq \sum_{i_1,\ldots,i_k=1}^{n} p_{i_1} \ldots p_{i_k} f\left(\frac{x_{i_1} + \cdots + x_{i_k}}{k}\right) \\ &\leq \cdots \leq \sum_{i=1}^{n} p_i f(x_i), \end{aligned} \tag{1.2}$$

for $k \geq 1$ and $\mathbf{p}$,$\mathbf{x}$ as above.

Date: January 07, 2009.

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

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

If $q_1, \dots, q_k \geq 0$ with $\sum_{j=1}^{k} q_j = 1$, then the following refinement obtained in 1994 by the author also holds (see [6]):

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i_1,\dots,i_k=1}^{n} p_{i_1} \dots p_{i_k} f\left(\frac{x_{i_1} + \dots + x_{i_k}}{k}\right) \leq \sum_{i_1,\dots,i_k=1}^{n} p_{i_1} \dots p_{i_k} f\left(q_1 x_{i_1} + \dots + q_k x_{i_k}\right) \leq \sum_{i=1}^{n} p_i f(x_i), \tag{1.3}$$

where $1 \leq k \leq n$ and $\mathbf{p}$, $\mathbf{x}$ are as above.

More recently the author obtained a different refinement of Jensen's inequality incorporated in (see [8]):

$$\begin{aligned} f\left(\sum_{j=1}^{n} p_j x_j\right) &\leq \min_{k\in\{1,\dots,n\}} \left[(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + p_k f(x_k)\right] \\ &\leq \frac{1}{n}\left[\sum_{k=1}^{n} (1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + \sum_{k=1}^{n} p_k f(x_k)\right] \\ &\leq \max_{k\in\{1,\dots,n\}} \left[(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + p_k f(x_k)\right] \\ &\leq \sum_{j=1}^{n} p_j f(x_j), \end{aligned} \tag{1.4}$$

where f, x_k and p_k are as above.

For other refinements and applications related to Ky Fan's inequality, the arithmetic mean-geometric mean inequality, the generalized triangle inequality, the f-Divergence measure etc., see [3]-[12].

In this paper, a new refinement of Jensen's celebrated inequality for functions defined on convex sets in linear spaces is given. Applications for norms, mean f-deviation and f-divergences are provided as well.

2. General Results

Let C be a convex subset in the real linear space X and assume that $f : C \to \mathbb{R}$ is a convex function on C. If $x_i \in C$ and $p_i > 0, i \in \{1, ..., n\}$ with $\sum_{i=1}^{n} p_i = 1$, then for any nonempty subset J of $\{1, ..., n\}$ we put $\bar{J} := \{1, ..., n\} \setminus J \, (\neq \emptyset)$ and define $P_J := \sum_{i\in J} p_i$ and $\bar{P}_J := P_{\bar{J}} = \sum_{j\in\bar{J}} p_j = 1 - \sum_{i\in J} p_i$. For the convex function f and the n-tuples $\mathbf{x} := (x_1, ..., x_n)$ and $\mathbf{p} := (p_1, ..., p_n)$ as above, we can define the following functional

$$D(f, \mathbf{p}, \mathbf{x}; J) := P_J f\left(\frac{1}{P_J}\sum_{i\in J} p_i x_i\right) + \bar{P}_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}} p_j x_j\right) \tag{2.1}$$

where here and everywhere below $J \subset \{1, ..., n\}$ with $J \neq \emptyset$ and $J \neq \{1, ..., n\}$.

It is worth to observe that for $J = \{k\}, k \in \{1, ..., n\}$ we have the functional

$$D_k(f, \mathbf{p}, \mathbf{x}) := D(f, \mathbf{p}, \mathbf{x}; \{k\}) = p_k f(x_k) + (1 - p_k) f\left(\frac{\sum_{i=1}^{n} p_i x_i - p_k x_k}{1 - p_k}\right) \tag{2.2}$$

that has been investigated in the earlier paper [8].

Theorem 1. *Let C be a convex subset in the real linear space X and assume that $f : C \to \mathbb{R}$ is a convex function on C. If $x_i \in C$ and $p_i > 0, i \in \{1, ..., n\}$ with $\sum_{i=1}^{n} p_i = 1$ then for any nonempty subset J of $\{1, ..., n\}$ we have*

$$\sum_{k=1}^{n} p_k f(x_k) \geq D(f, \mathbf{p}, \mathbf{x}; J) \geq f\left(\sum_{k=1}^{n} p_k x_k\right). \tag{2.3}$$

Proof. By the convexity of the function f we have

$$\begin{aligned} D(f, \mathbf{p}, \mathbf{x}; J) &= P_J f\left(\frac{1}{P_J} \sum_{i \in J} p_i x_i\right) + \bar{P}_J f\left(\frac{1}{\bar{P}_J} \sum_{j \in \bar{J}} p_j x_j\right) \\ &\geq f\left[P_J\left(\frac{1}{P_J} \sum_{i \in J} p_i x_i\right) + \bar{P}_J\left(\frac{1}{\bar{P}_J} \sum_{j \in \bar{J}} p_j x_j\right)\right] \\ &= f\left(\sum_{k=1}^{n} p_k x_k\right) \end{aligned}$$

for any J, which proves the second inequality in (2.3).

By the Jensen inequality we also have

$$\begin{aligned} \sum_{k=1}^{n} p_k f(x_k) &= \sum_{i \in J} p_i f(x_i) + \sum_{j \in \bar{J}} p_j f(x_j) \\ &\geq P_J f\left(\frac{1}{P_J} \sum_{i \in J} p_i x_i\right) + \bar{P}_J f\left(\frac{1}{\bar{P}_J} \sum_{j \in \bar{J}} p_j x_j\right) \\ &= D(f, \mathbf{p}, \mathbf{x}; J) \end{aligned}$$

for any J, which proves the first inequality in (2.3). □

Remark 1. *We observe that the inequality (2.3) can be written in an equivalent form as*

$$\sum_{k=1}^{n} p_k f(x_k) \geq \max_{\emptyset \neq J \subset \{1, ..., n\}} D(f, \mathbf{p}, \mathbf{x}; J) \tag{2.4}$$

and

$$\min_{\emptyset \neq J \subset \{1, ..., n\}} D(f, \mathbf{p}, \mathbf{x}; J) \geq f\left(\sum_{k=1}^{n} p_k x_k\right). \tag{2.5}$$

These inequalities imply the following results that have been obtained earlier by the author in [8] *utilising a different method of proof slightly more complicated:*

$$\sum_{k=1}^{n} p_k f(x_k) \geq \max_{k \in \{1, ..., n\}} D_k(f, \mathbf{p}, \mathbf{x}) \tag{2.6}$$

and

$$\min_{k\in\{1,...,n\}} D_k(f,\mathbf{p},\mathbf{x}) \geq f\left(\sum_{k=1}^{n} p_k x_k\right). \tag{2.7}$$

Moreover, since

$$\max_{\emptyset\neq J\subset\{1,...,n\}} D(f,\mathbf{p},\mathbf{x};J) \geq \max_{k\in\{1,...,n\}} D_k(f,\mathbf{p},\mathbf{x})$$

and

$$\min_{k\in\{1,...,n\}} D_k(f,\mathbf{p},\mathbf{x}) \geq \min_{\emptyset\neq J\subset\{1,...,n\}} D(f,\mathbf{p},\mathbf{x};J),$$

then the new inequalities (2.4) and (2.4) are better than the earlier results developed in [8].

The case of uniform distribution, namely, when $p_i = \frac{1}{n}$ for all $\{1,...,n\}$ is of interest as well. If we consider a natural number m with $1 \leq m \leq n-1$ and if we define

$$D_m(f,\mathbf{x}) := \frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right) \tag{2.8}$$

then we can state the following result:

Corollary 1. *Let C be a convex subset in the real linear space X and assume that $f : C \to \mathbb{R}$ is a convex function on C. If $x_i \in C$, then for any $m \in \{1,...,n-1\}$ we have*

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) \geq D_m(f,\mathbf{x}) \geq f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right). \tag{2.9}$$

In particular, we have the bounds

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} f(x_k) \\ \geq \max_{m\in\{1,...,n-1\}} \left[\frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right)\right]\end{aligned} \tag{2.10}$$

and

$$\begin{aligned}\min_{m\in\{1,...,n-1\}} \left[\frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right)\right] \\ \geq f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right).\end{aligned} \tag{2.11}$$

The following version of the inequality (2.3) may be useful for symmetric convex functions:

Corollary 2. *Let C be a convex function with the property that $0 \in C$. If $y_j \in X$ such that for $p_i > 0, i \in \{1, ..., n\}$ with $\sum_{i=1}^n p_i = 1$ we have $y_j - \sum_{i=1}^n p_i y_i \in C$ for any $j \in \{1, ..., n\}$, then for any nonempty subset J of $\{1, ..., n\}$ we have*

$$\sum_{k=1}^n p_k f\left(y_k - \sum_{i=1}^n p_i y_i\right) \geq P_J f\left[\bar{P}_J\left(\frac{1}{P_J}\sum_{i\in J} p_i y_i - \frac{1}{\bar{P}_J}\sum_{j\in\bar{J}} p_j y_j\right)\right] + \bar{P}_J f\left[P_J\left(\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}} p_j y_j - \frac{1}{P_J}\sum_{i\in J} p_i y_i\right)\right] \geq f(0). \tag{2.12}$$

Remark 2. *If C is as in Corollary 2 and $y_j \in X$ such that $y_j - \frac{1}{n}\sum_{i=1}^n y_i \in C$ for any $j \in \{1, ..., n\}$ then for any $m \in \{1, ..., n-1\}$ we have*

$$\frac{1}{n}\sum_{k=1}^n f\left(y_k - \frac{1}{n}\sum_{i=1}^n y_i\right) \geq \frac{m}{n} f\left[\frac{n-m}{n}\left(\frac{1}{m}\sum_{i=1}^m y_i - \frac{1}{n-m}\sum_{j=m+1}^n y_j\right)\right] + \frac{n-m}{n} f\left[\frac{m}{n}\left(\frac{1}{n-m}\sum_{j=m+1}^n y_j - \frac{1}{m}\sum_{i=1}^m y_i\right)\right] \geq f(0). \tag{2.13}$$

Remark 3. *It is also useful to remark that if $J = \{k\}$ where $k \in \{1, ..., n\}$ then the particular form we can derive from (2.12) can be written as*

$$\begin{aligned} \sum_{\ell=1}^n p_\ell f\left(y_\ell - \sum_{i=1}^n p_i y_i\right) &\geq p_k f\left[(1-p_k)\left(y_k - \frac{1}{1-p_k}\left(\sum_{j=1}^n p_j y_j - p_k y_k\right)\right)\right] \\ &\quad + (1-p_k) f\left[p_k\left(\frac{1}{1-p_k}\left(\sum_{j=1}^n p_j y_j - p_k y_k\right) - y_k\right)\right] \geq f(0) \end{aligned} \tag{2.14}$$

which is equivalent with

$$\sum_{\ell=1}^n p_\ell f\left(y_\ell - \sum_{i=1}^n p_i y_i\right) \geq p_k f\left(y_k - \sum_{j=1}^n p_j y_j\right) + (1-p_k) f\left[\frac{p_k}{1-p_k}\left(\sum_{j=1}^n p_j y_j - y_k\right)\right] \geq f(0) \tag{2.15}$$

for any $k \in \{1, ..., n\}$.

Remark 4. *Continuous versions for the Lebesgue integral are considered in* [9].

3. A Lower Bound for Mean f-Deviation

Let X be a real linear space. For a convex function $f : X \to \mathbb{R}$ with the properties that $f(0) = 0$, define the *mean f-deviation* of an n-tuple of vectors

$\mathbf{x}=(x_1,...,x_n)\in X^n$ with the probability distribution $\mathbf{p}=(p_1,...,p_n)$ by the non-negative quantity

$$K_f(\mathbf{p},\mathbf{x}):=\sum_{i=1}^{n}p_i f\left(x_i-\sum_{k=1}^{n}p_k x_k\right). \tag{3.1}$$

The fact that $K_f(\mathbf{p},\mathbf{x})$ is non-negative follows by Jensen's inequality, namely

$$K_f(\mathbf{p},\mathbf{x})\geq f\left(\sum_{i=1}^{n}p_i\left(x_i-\sum_{k=1}^{n}p_k x_k\right)\right)=f(0)=0.$$

A natural example of such deviations can be provided by the convex function $f(x):=\|x\|^r$ with $r\geq 1$ defined on a normed linear space $(X,\|\cdot\|)$. We denote this by

$$K_r(\mathbf{p},\mathbf{x}):=\sum_{i=1}^{n}p_i\left\|x_i-\sum_{k=1}^{n}p_k x_k\right\|^r \tag{3.2}$$

and call it the *mean r-absolute deviation* of the n-tuple of vectors $\mathbf{x}=(x_1,...,x_n)\in X^n$ with the probability distribution $\mathbf{p}=(p_1,...,p_n)$.

The following result that provides a lower bound for the mean f-deviation holds:

Theorem 2. *Let $f:X\to[0,\infty)$ be a convex function with $f(0)=0$. If $\mathbf{x}=(x_1,...,x_n)\in X^n$ and $\mathbf{p}=(p_1,...,p_n)$ is a probability distribution with all p_i nonzero, then*

$$\begin{aligned}K_f(\mathbf{p},\mathbf{x})\geq \max_{\emptyset\neq J\subset\{1,...,n\}}\Bigg\{&P_J f\left[\bar{P}_J\left(\frac{1}{P_J}\sum_{i\in J}p_i x_i-\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_j x_j\right)\right]\\ &+P_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_j y_j-\frac{1}{P_J}\sum_{i\in J}p_i y_i\right)\Bigg\}(\geq 0).\end{aligned} \tag{3.3}$$

In particular, we have

$$\begin{aligned}&K_f(\mathbf{p},\mathbf{x})\\ &\geq \max_{k\in\{1,...,n\}}\left\{(1-p_k)f\left[\frac{p_k}{1-p_k}\left(\sum_{l=1}^{n}p_l x_l-x_k\right)\right]+p_k f\left(x_k-\sum_{l=1}^{n}p_l x_l\right)\right\}(\geq 0).\end{aligned} \tag{3.4}$$

The proof follows from Corollary 2 and Remark 3.

As a particular case of interest, we have the following:

Corollary 3. *Let $(X,\|\cdot\|)$ be a normed linear space. If $\mathbf{x}=(x_1,...,x_n)\in X^n$ and $\mathbf{p}=(p_1,...,p_n)$ is a probability distribution with all p_i nonzero, then for $r\geq 1$ we have*

$$\begin{aligned}&K_r(\mathbf{p},\mathbf{x})\\ &\geq \max_{\emptyset\neq J\subset\{1,...,n\}}\left\{P_J\bar{P}_J\left(\bar{P}_J^{r-1}+P_J^{r-1}\right)\left\|\frac{1}{P_J}\sum_{i\in J}p_i x_i-\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_j x_j\right\|^r\right\}(\geq 0).\end{aligned} \tag{3.5}$$

Remark 5. *By the convexity of the power function $f(t) = t^r, r \geq 1$ we have*

$$\begin{aligned} P_J \bar{P}_J \left(\bar{P}_J^{r-1} + P_J^{r-1} \right) &= P_J \bar{P}_J^r + \bar{P}_J P_J^r \\ &\geq \left(P_J \bar{P}_J + \bar{P}_J P_J \right)^r = 2^r P_J^r \bar{P}_J^r \end{aligned}$$

therefore

$$\begin{aligned} (3.6) \quad & P_J \bar{P}_J \left(\bar{P}_J^{r-1} + P_J^{r-1} \right) \left\| \frac{1}{P_J} \sum_{i \in J} p_i x_i - \frac{1}{\bar{P}_J} \sum_{j \in \bar{J}} p_j x_j \right\|^r \\ & \geq 2^r P_J^r \bar{P}_J^r \left\| \frac{1}{P_J} \sum_{i \in J} p_i x_i - \frac{1}{\bar{P}_J} \sum_{j \in \bar{J}} p_j x_j \right\|^r = 2^r \left\| \bar{P}_J \sum_{i \in J} p_i x_i - P_J \sum_{j \in \bar{J}} p_j x_j \right\|^r . \end{aligned}$$

Since

$$\begin{aligned} (3.7) \quad \bar{P}_J \sum_{i \in J} p_i x_i - P_J \sum_{j \in \bar{J}} p_j x_j &= (1 - P_J) \sum_{i \in J} p_i x_i - P_J \left(\sum_{k=1}^n p_k x_k - \sum_{i \in J} p_i x_i \right) \\ &= \sum_{i \in J} p_i x_i - P_J \sum_{k=1}^n p_k x_k, \end{aligned}$$

then by (3.5)-(3.7) we deduce the coarser but perhaps more useful lower bound

$$(3.8) \qquad K_r(\mathbf{p}, \mathbf{x}) \geq 2^r \max_{\emptyset \neq J \subset \{1,\ldots,n\}} \left\{ \left\| \sum_{i \in J} p_i x_i - P_J \sum_{k=1}^n p_k x_k \right\|^r \right\} (\geq 0) .$$

The case for mean r-absolute deviation is incorporated in:

Corollary 4. *Let $(X, \|\cdot\|)$ be a normed linear space. If $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and $\mathbf{p} = (p_1, ..., p_n)$ is a probability distribution with all p_i nonzero, then for $r \geq 1$ we have*

$$(3.9) \qquad K_r(\mathbf{p}, \mathbf{x}) \geq \max_{k \in \{1,\ldots,n\}} \left\{ \left[(1 - p_k)^{1-r} p_k^r + p_k \right] \left\| x_k - \sum_{l=1}^n p_l x_l \right\|^r \right\} .$$

Remark 6. *Since the function $h_r(t) := (1-t)^{1-r} t^r + t$, $r \geq 1$, $t \in [0, 1)$ is strictly increasing on $[0, 1)$, then*

$$\min_{k \in \{1,\ldots,n\}} \left\{ (1 - p_k)^{1-r} p_k^r + p_k \right\} = p_m + (1 - p_m)^{1-r} p_m^r,$$

where $p_m := \min_{k \in \{1,\ldots,n\}} p_k$. By (3.9), we then obtain the following simpler inequality:

$$(3.10) \qquad K_r(\mathbf{p}, \mathbf{x}) \geq \left[p_m + (1 - p_m)^{1-r} \cdot p_m^r \right] \max_{k \in \{1,\ldots,n\}} \left\| x_k - \sum_{l=1}^n p_l x_l \right\|^p ,$$

which is perhaps more useful for applications(see also [8]).

4. Applications for f-Divergence Measures

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

$$I_f(\mathbf{p},\mathbf{q}):=\sum_{i=1}^{n}q_i f\left(\frac{p_i}{q_i}\right), \tag{4.1}$$

where $\mathbf{p}=(p_1,\ldots,p_n)$, $\mathbf{q}=(q_1,\ldots,q_n)$ are positive sequences, was introduced by Csiszár in [1], as a generalized measure of information, a "distance function" on the set of probability distributions $\mathbb{P}^n$. As in [1], we interpret undefined expressions by

$$f(0)=\lim_{t\to 0+}f(t),\qquad 0f\left(\frac{0}{0}\right)=0,$$
$$0f\left(\frac{a}{0}\right)=\lim_{q\to 0+}qf\left(\frac{a}{q}\right)=a\lim_{t\to\infty}\frac{f(t)}{t},\quad a>0.$$

The following results were essentially given by Csiszár and Körner [2]:

(i) *If f is convex, then $I_f(\mathbf{p},\mathbf{q})$ is jointly convex in $\mathbf{p}$ and $\mathbf{q}$;*

(ii) *For every $\mathbf{p},\mathbf{q}\in R_+^n$, we have*

$$I_f(\mathbf{p},\mathbf{q})\geq\sum_{j=1}^{n}q_j f\left(\frac{\sum_{j=1}^{n}p_j}{\sum_{j=1}^{n}q_j}\right). \tag{4.2}$$

If f is strictly convex, equality holds in (4.2) iff

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

If f is normalized, i.e., $f(1)=0$, then for every $\mathbf{p},\mathbf{q}\in R_+^n$ with $\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i$, we have the inequality

$$I_f(\mathbf{p},\mathbf{q})\geq 0. \tag{4.3}$$

In particular, if $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$, then (4.3) holds. This is the well-known positivity property of the f-divergence.

We endeavour to extend this concept for functions defined on a cone in a linear space as follows.

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 $\mathbf{q}\in\mathbb{P}^n$ 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 $\mathbf{q}$*

$$I_f(\mathbf{z},\mathbf{q}):=\sum_{i=1}^{n}q_i f\left(\frac{z_i}{q_i}\right). \tag{4.4}$$

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 $f:[0,\infty)\to\mathbb{R}$.

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

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

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

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

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

Theorem 3. *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*

$$\begin{aligned} I_f\left(\mathbf{x}, \mathbf{q}\right) &\geq \max_{\emptyset\neq J\subset\{1,\ldots,n\}} I_f\left(\mathbf{x}_J, \mathbf{q}_J\right) \geq I_f\left(\mathbf{x}_J, \mathbf{q}_J\right) \\ &\geq \min_{\emptyset\neq J\subset\{1,\ldots,n\}} I_f\left(\mathbf{x}_J, \mathbf{q}_J\right) \geq f\left(X_n\right) \end{aligned} \tag{4.5}$$

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

The proof follows by Theorem 1 and the details are omitted.

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:

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

$$I_f\left(\mathbf{p}, \mathbf{q}\right) \geq \max_{\emptyset\neq J\subset\{1,\ldots,n\}} \left[Q_J f\left(\frac{P_J}{Q_J}\right) + (1 - Q_J) f\left(\frac{1 - P_J}{1 - Q_J}\right)\right] (\geq 0). \tag{4.6}$$

In what follows we provide some lower bounds for a number of f-divergences that are used in various fields of Information Theory, Probability Theory and Statistics.

The *total variation distance* is defined by the convex function $f(t) = |t - 1|$, $t \in \mathbb{R}$ and given in:

$$V(p, q) := \sum_{j=1}^n q_j \left|\frac{p_j}{q_j} - 1\right| = \sum_{j=1}^n |p_j - q_j|. \tag{4.7}$$

The following improvement of the positivity inequality for the total variation distance can be stated as follows.

Proposition 1. *For any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, we have the inequality:*

$$V(p, q) \geq 2 \max_{\emptyset\neq J\subset\{1,\ldots,n\}} |P_J - Q_J| \quad (\geq 0). \tag{4.8}$$

The proof follows by the inequality (4.6) for $f(t) = |t-1|$, $t \in \mathbb{R}$.

The K. Pearson χ^2-*divergence* is obtained for the convex function $f(t) = (1-t)^2$, $t \in \mathbb{R}$ and given by

$$\chi^2(p,q) := \sum_{j=1}^{n} q_j \left(\frac{p_j}{q_j} - 1\right)^2 = \sum_{j=1}^{n} \frac{(p_j - q_j)^2}{q_j}. \tag{4.9}$$

Proposition 2. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$,

$$\chi^2(p,q) \geq \max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \frac{(P_J - Q_J)^2}{Q_J(1-Q_J)} \right\} \tag{4.10}$$

$$\geq 4 \max_{\emptyset \neq J \subset \{1,\dots,n\}} (P_J - Q_J)^2 \quad (\geq 0).$$

Proof. On applying the inequality (4.6) for the function $f(t) = (1-t)^2$, $t \in \mathbb{R}$, we get

$$\chi^2(p,q) \geq \max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ (1-Q_J)\left(\frac{1-P_J}{1-Q_J} - 1\right)^2 + Q_J\left(\frac{P_J}{Q_J} - 1\right)^2 \right\}$$
$$= \max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \frac{(P_J - Q_J)^2}{Q_J(1-Q_J)} \right\}.$$

Since

$$Q_J(1-Q_J) \leq \frac{1}{4}\left[Q_J + (1-Q_J)\right]^2 = \frac{1}{4},$$

then

$$\frac{(P_J - Q_J)^2}{Q_J(1-Q_J)} \geq 4(P_J - Q_J)^2$$

for each $J \subset \{1,\dots,n\}$, which proves the last part of (4.10). □

The *Kullback-Leibler divergence* can be obtained for the convex function $f : (0,\infty) \to \mathbb{R}$, $f(t) = t \ln t$ and is defined by

$$KL(p,q) := \sum_{j=1}^{n} q_j \cdot \frac{p_j}{q_j} \ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n} p_j \ln\left(\frac{p_j}{q_j}\right). \tag{4.11}$$

Proposition 3. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, *we have:*

$$KL(p,q) \geq \ln\left[\max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \left(\frac{1-P_J}{1-Q_J}\right)^{1-P_J} \cdot \left(\frac{P_J}{Q_J}\right)^{P_J} \right\}\right] \geq 0. \tag{4.12}$$

Proof. The first inequality is obvious by Corollary 5. Utilising the inequality between the *geometric mean and the harmonic mean*,

$$x^\alpha y^{1-\alpha} \geq \frac{1}{\frac{\alpha}{x} + \frac{1-\alpha}{y}}, \qquad x, y > 0,\ \alpha \in [0,1]$$

we have for $x = \frac{P_J}{Q_J}$, $y = \frac{1-P_J}{1-Q_J}$ and $\alpha = P_J$ that

$$\left(\frac{1-P_J}{1-Q_J}\right)^{1-P_J} \cdot \left(\frac{P_J}{Q_J}\right)^{P_J} \geq 1,$$

for any $J \subset \{1,\dots,n\}$, which implies the second part of (4.12). □

Another divergence measure that is of importance in Information Theory is the *Jeffreys divergence*

$$J(p,q) := \sum_{j=1}^{n} q_j \cdot \left(\frac{p_j}{q_j} - 1\right) \ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n} (p_j - q_j) \ln\left(\frac{p_j}{q_j}\right), \tag{4.13}$$

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

Proposition 4. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, *we have:*

$$\begin{aligned} J(p,q) &\geq \ln\left(\max_{k\in\{1,\dots,n\}} \left\{\left[\frac{(1-P_J)\,Q_J}{(1-Q_J)\,P_J}\right]^{(Q_J - P_J)}\right\}\right) \\ &\geq \max_{k\in\{1,\dots,n\}} \left[\frac{(Q_J - P_J)^2}{P_J + Q_J - 2P_J Q_J}\right] \geq 0. \end{aligned} \tag{4.14}$$

Proof. On making use of the inequality (4.6) for $f(t) = (t-1)\ln t$, we have

$$\begin{aligned} &J(p,q) \\ &\geq \max_{k\in\{1,\dots,n\}} \left\{(1-Q_J)\left[\left(\frac{1-P_J}{1-Q_J} - 1\right)\ln\left(\frac{1-P_J}{1-Q_J}\right)\right] + Q_J\left(\frac{P_J}{Q_J} - 1\right)\ln\left(\frac{P_J}{Q_J}\right)\right\} \\ &= \max_{k\in\{1,\dots,n\}} \left\{(Q_J - P_J)\ln\left(\frac{1-P_J}{1-Q_J}\right) - (Q_J - P_J)\ln\left(\frac{P_J}{Q_J}\right)\right\} \\ &= \max_{k\in\{1,\dots,n\}} \left\{(Q_J - P_J)\ln\left[\frac{(1-P_J)\,Q_J}{(1-Q_J)\,P_J}\right]\right\}, \end{aligned}$$

proving the first inequality in (4.14).

Utilising the elementary inequality for positive numbers,

$$\frac{\ln b - \ln a}{b - a} \geq \frac{2}{a+b}, \qquad a, b > 0$$

we have

$$\begin{aligned} &(Q_J - P_J)\left[\ln\left(\frac{1-P_J}{1-Q_J}\right) - \ln\left(\frac{P_J}{Q_J}\right)\right] \\ &= (Q_J - P_J) \cdot \frac{\ln\left(\frac{1-P_J}{1-Q_J}\right) - \ln\left(\frac{P_J}{Q_J}\right)}{\frac{1-P_J}{1-Q_J} - \frac{P_J}{Q_J}} \cdot \left[\frac{1-P_J}{1-Q_J} - \frac{P_J}{Q_J}\right] \\ &= \frac{(Q_J - P_J)^2}{Q_J(1-Q_J)} \cdot \frac{\ln\left(\frac{1-P_J}{1-Q_J}\right) - \ln\left(\frac{P_J}{Q_J}\right)}{\frac{1-P_J}{1-Q_J} - \frac{P_J}{Q_J}} \\ &\geq \frac{(Q_J - P_J)^2}{Q_J(1-Q_J)} \cdot \frac{2}{\frac{1-P_J}{1-Q_J} + \frac{P_J}{Q_J}} = \frac{2(Q_J - P_J)^2}{P_J + Q_J - 2P_J Q_J} \geq 0, \end{aligned}$$

for each $J \subset \{1, \dots, n\}$, giving the second inequality in (4.14). □

References

[1] I. Csiszár, Information-type measures of differences of probability distributions and indirect observations, *Studia Sci. Math. Hung.*, **2** (1967), 299-318.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[3] S.S. Dragomir, An improvement of Jensen's inequality, *Bull. Math. Soc. Sci. Math. Roumanie*, **34(82)** (1990), No. 4, 291-296.

[4] S.S. Dragomir, Some refinements of Ky Fan's inequality, *J. Math. Anal. Appl.*, **163**(2) (1992), 317-321.
[5] S.S. Dragomir, Some refinements of Jensen's inequality, *J. Math. Anal. Appl.*, **168**(2) (1992), 518-522.
[6] S.S. Dragomir, A further improvement of Jensen's inequality, *Tamkang J. Math.*, **25**(1) (1994), 29-36.
[7] S.S. Dragomir, A new improvement of Jensen's inequality, *Indian J. Pure and Appl. Math.*, **26**(10) (1995), 959-968.
[8] S.S. Dragomir, A refinement of Jensen's inequality with applications for f-divergence measures, Preprint, *Res. Rep. Coll.* **10**(2007), Supp., Article 15.
[9] S.S. Dragomir, Refinements of the Jensen integral inequality, *in preparation.*
[10] S.S. Dragomir and N.M. Ionescu, Some converse of Jensen's inequality and applications. *Rev. Anal. Numér. Théor. Approx.* **23** (1994), no. 1, 71–78. MR1325895 (96c:26012).
[11] S.S. Dragomir and C.J. Goh, A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory, *Mathl. Comput. Modelling*, **24**(1996), No. 2, 1-11.
[12] S.S. Dragomir, J. Pečarić and L.E. Persson, Properties of some functionals related to Jensen's inequality, *Acta Math. Hung.*, **70**(1-2) (1996), 129-143.
[13] J. Pečarić and S.S. Dragomir, A refinements of Jensen inequality and applications, *Studia Univ. Babeş-Bolyai, Mathematica,* **24**(1) (1989), 15-19.

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/