

Bounds for the Normalised Jensen Functional

Sever S. Dragomir

ABSTRACT. New inequalities for the general case of convex functions defined on linear spaces which improve the famous Jensen's inequality are established. Particular instances in the case of normed spaces and for complex and real n -tuples are given. Refinements of Shannon's inequality and the positivity of Kullback-Leibler divergence are obtained.

1. Introduction

Jensen's inequality for convex function is one of the most known and extensively used inequality in various filed of Modern Mathematics. It is a source of many classical inequalities including the generalised triangle inequality, the arithmetic mean-geometric mean-harmonic mean inequality, the positivity of *relative entropy* in Information Theory, Schannon's inequality, Ky Fan's inequality, Levinson's inequality and other results. For classical and contemporary developments related to the Jensen inequality, see [6], [7] and [3] where further references are provided.

To be more specific, we recall that, if X is a linear space and $C \subseteq X$ a convex subset in X , then for any convex function $f : C \rightarrow \mathbb{R}$ and any $z_i \in C, r_i \geq 0$ for $i \in \{1, \dots, k\}, k \geq 2$ with $\sum_{i=1}^k r_i = R_k > 0$ one has the *weighted Jensen's inequality*:

$$(J) \quad \frac{1}{R_k} \sum_{i=1}^k r_i f(z_i) \geq f\left(\frac{1}{R_k} \sum_{i=1}^k r_i z_i\right).$$

If $f : C \rightarrow \mathbb{R}$ is strictly convex and $r_i > 0$ for $i \in \{1, \dots, k\}$ then the equality case holds in (J) if and only if $z_1 = \dots = z_n$.

The main aim of the present note is to provide an elementary refinement of this classical inequality and point out a few applications in relation with some fundamental inequalities in various fields of Mathematics.

1991 *Mathematics Subject Classification*. Primary 26D15; Secondary 26D10.
Key words and phrases. Jensen's Inequality, Convex Functions,

2. Bounds for the Normalised Jensen Functional

By \mathcal{P}_n we denote the set of all nonnegative n -tuples (p_1, \dots, p_n) with the property that $\sum_{i=1}^n p_i = 1$. Consider the *normalised Jensen functional*

$$\mathcal{J}_n(f, \mathbf{x}, \mathbf{p}) = \sum_{i=1}^n p_i f(x_i) - f\left(\sum_{i=1}^n p_i x_i\right) \geq 0$$

where $f : C \rightarrow \mathbb{R}$ be a convex function on the convex set C and $\mathbf{x} = (x_1, \dots, x_n) \in C^n$ and $\mathbf{p} \in \mathcal{P}_n$.

The following result holds.

THEOREM 1. *If $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$, $q_i > 0$ for each $i \in \{1, \dots, n\}$ then*

$$(2.1) \quad \max_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\} \mathcal{J}_n(f, \mathbf{x}, \mathbf{q}) \geq \mathcal{J}_n(f, \mathbf{x}, \mathbf{p}) \geq \min_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\} \mathcal{J}_n(f, \mathbf{x}, \mathbf{q}) \quad (\geq 0).$$

PROOF. We give here a direct proof based on the Jensen inequality for appropriate choices of the elements in (J). The reader is invited to try to find other proofs, eventually simpler than this one.

Denote $m := \min_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\}$ and observe that $0 \leq m \leq 1$.

If we apply Jensen's inequality (J) for $k = n + 1$,

$$\begin{aligned} z_1 &= \sum_{i=1}^n q_i x_i, \quad z_{j+1} = x_j \in C, \quad j \in \{1, \dots, n\}; \\ r_1 &= m, \quad r_{j+1} = \left(\frac{p_j}{q_j} - m \right) q_j \geq 0, \quad j \in \{1, \dots, n\} \end{aligned}$$

for which

$$\sum_{j=1}^{n+1} r_j = m + \sum_{j=1}^n \left(\frac{p_j}{q_j} - m \right) q_j = 1$$

then we have

$$(2.2) \quad \begin{aligned} & m f\left(\sum_{i=1}^n q_i x_i\right) + \sum_{j=1}^n \left(\frac{p_j}{q_j} - m \right) q_j f(x_j) \\ & \geq f\left[m \left(\sum_{i=1}^n q_i x_i \right) + \sum_{j=1}^n \left(\frac{p_j}{q_j} - m \right) q_j x_j \right]. \end{aligned}$$

Since

$$\begin{aligned} & m f\left(\sum_{i=1}^n q_i x_i\right) + \sum_{j=1}^n \left(\frac{p_j}{q_j} - m \right) q_j f(x_j) \\ & = \sum_{j=1}^n p_j f(x_j) - m \left[\sum_{j=1}^n q_j f(x_j) - f\left(\sum_{j=1}^n q_j x_j\right) \right] \end{aligned}$$

and

$$f\left[m \left(\sum_{i=1}^n q_i x_i \right) + \sum_{j=1}^n \left(\frac{p_j}{q_j} - m \right) q_j x_j \right] = f\left(\sum_{j=1}^n p_j x_j\right)$$

hence by (2.2) we get the second inequality in (2.1).

Now, let $M := \max_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\}$ and observe that $M \geq 1$.

If we apply Jensen's inequality (J) for $k = n + 1$,

$$\begin{aligned} z_j &= x_j, \quad z_{n+1} = \sum_{i=1}^n p_i x_i, \quad j \in \{1, \dots, n\}, \\ r_j &= \frac{1}{M} \left(M - \frac{p_j}{q_j} \right) q_j, \quad r_{n+1} = \frac{1}{M}, \quad j \in \{1, \dots, n\} \end{aligned}$$

for which

$$\sum_{j=1}^{n+1} r_j = \frac{1}{M} \sum_{j=1}^n \left(M - \frac{p_j}{q_j} \right) q_j + \frac{1}{M} = 1,$$

then we have

$$\begin{aligned} (2.3) \quad & \frac{1}{M} \sum_{j=1}^n \left(M - \frac{p_j}{q_j} \right) q_j f(x_j) + \frac{1}{M} f \left(\sum_{i=1}^n p_i x_i \right) \\ & \geq f \left[\frac{1}{M} \sum_{j=1}^n \left(M - \frac{p_j}{q_j} \right) q_j x_j + \frac{1}{M} \sum_{i=1}^n p_i x_i \right]. \end{aligned}$$

Since

$$\begin{aligned} & \frac{1}{M} \sum_{j=1}^n \left(M - \frac{p_j}{q_j} \right) q_j f(x_j) + \frac{1}{M} f \left(\sum_{i=1}^n p_i x_i \right) \\ & = \sum_{j=1}^n q_j f(x_j) - \frac{1}{M} \left[\sum_{j=1}^n p_j f(x_j) - \left(\sum_{i=1}^n p_i x_i \right) \right] \end{aligned}$$

and

$$f \left[\frac{1}{M} \sum_{j=1}^n \left(M - \frac{p_j}{q_j} \right) q_j x_j + \frac{1}{M} \sum_{i=1}^n p_i x_i \right] = f \left(\sum_{j=1}^n p_j x_j \right),$$

hence by (2.3) we deduce the first part of (2.1) and the proof is complete. ■

If we consider for the uniform distribution $\mathbf{u} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)$ the unweighted Jensen functional

$$\mathcal{J}_n(f, \mathbf{x}) := \mathcal{J}_n(f, \mathbf{x}, \mathbf{u}) = \frac{1}{n} \sum_{i=1}^n f(x_i) - f \left(\frac{1}{n} \sum_{i=1}^n x_i \right)$$

then we can state the following particular case of interest as well:

COROLLARY 1. *If $\mathbf{p} \in \mathcal{P}_n$, then*

$$(2.4) \quad n \max_{1 \leq i \leq n} \{p_i\} \mathcal{J}_n(f, \mathbf{x}) \geq \mathcal{J}_n(f, \mathbf{x}, \mathbf{p}) \geq n \min_{1 \leq i \leq n} \{p_i\} \mathcal{J}_n(f, \mathbf{x}) \quad (\geq 0).$$

3. Applications in Normed Spaces

Let $(X, \|\cdot\|)$ be a real or complex normed linear space. It is well known that the function $f_p : X \rightarrow \mathbb{R}$, $f_p(x) = \|x\|^p$, $p \geq 1$ is convex on X . Applying the results obtained above one may easily state the following inequalities:

$$\begin{aligned}
 (3.1) \quad & \max_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\} \left[\sum_{j=1}^n q_j \|x_j\|^p - \left\| \sum_{j=1}^n q_j x_j \right\|^p \right] \\
 & \geq \sum_{j=1}^n p_j \|x_j\|^p - \left\| \sum_{j=1}^n p_j x_j \right\|^p \\
 & \geq \min_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\} \left[\sum_{j=1}^n q_j \|x_j\|^p - \left\| \sum_{j=1}^n q_j x_j \right\|^p \right] \quad (\geq 0)
 \end{aligned}$$

and

$$\begin{aligned}
 (3.2) \quad & \max_{1 \leq i \leq n} \{p_i\} \left[\sum_{j=1}^n \|x_j\|^p - n^{1-p} \left\| \sum_{j=1}^n x_j \right\|^p \right] \geq \sum_{j=1}^n p_j \|x_j\|^p - \left\| \sum_{j=1}^n p_j x_j \right\|^p \\
 & \geq \min_{1 \leq i \leq n} \{p_i\} \left[\sum_{j=1}^n \|x_j\|^p - n^{1-p} \left\| \sum_{j=1}^n x_j \right\|^p \right] \quad (\geq 0)
 \end{aligned}$$

for all $p \geq 1$.

If in (3.2) we choose $p_j := \frac{1}{\|x_j\|}$, where $x_j \in X \setminus \{0\}$, $j \in \{1, \dots, n\}$, then we get

$$\begin{aligned}
 (3.3) \quad & \max_{1 \leq j \leq n} \{\|x_j\|\} \left[\sum_{j=1}^n \|x_j\|^{p-1} - \left\| \sum_{j=1}^n \frac{x_j}{\|x_j\|} \right\|^p \right] \\
 & \geq \sum_{j=1}^n \|x_j\|^p - n^{1-p} \left\| \sum_{j=1}^n x_j \right\|^p \\
 & \geq \min_{1 \leq j \leq n} \{\|x_j\|\} \left[\sum_{j=1}^n \|x_j\|^{p-1} - \left\| \sum_{j=1}^n \frac{x_j}{\|x_j\|} \right\|^p \right].
 \end{aligned}$$

We remark that, for $p = 1$ one may get out of the previous results the following inequalities that are intimately related with the generalised triangle inequality in normed spaces:

$$\begin{aligned}
 (3.4) \quad & \max_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\} \left[\sum_{j=1}^n q_j \|x_j\| - \left\| \sum_{j=1}^n q_j x_j \right\| \right] \geq \sum_{j=1}^n p_j \|x_j\| - \left\| \sum_{j=1}^n p_j x_j \right\| \\
 & \geq \min_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\} \left[\sum_{j=1}^n q_j \|x_j\| - \left\| \sum_{j=1}^n q_j x_j \right\| \right] \quad (\geq 0),
 \end{aligned}$$

$$\begin{aligned}
(3.5) \quad & \max_{1 \leq i \leq n} \{p_i\} \left[\sum_{j=1}^n \|x_j\| - \left\| \sum_{j=1}^n x_j \right\| \right] \geq \sum_{j=1}^n p_j \|x_j\| - \left\| \sum_{j=1}^n p_j x_j \right\| \\
& \geq \min_{1 \leq i \leq n} \{p_i\} \left[\sum_{j=1}^n \|x_j\| - \left\| \sum_{j=1}^n x_j \right\| \right] \quad (\geq 0)
\end{aligned}$$

and

$$\begin{aligned}
(3.6) \quad & \max_{1 \leq j \leq n} \{\|x_j\|\} \left[n - \left\| \sum_{j=1}^n \frac{x_j}{\|x_j\|} \right\| \right] \geq \sum_{j=1}^n \|x_j\| - \left\| \sum_{j=1}^n x_j \right\| \\
& \geq \min_{1 \leq j \leq n} \{\|x_j\|\} \left[n - \left\| \sum_{j=1}^n \frac{x_j}{\|x_j\|} \right\| \right].
\end{aligned}$$

We note that the inequality (3.6) has been obtained in a different way by M. Kato, K.-S. Saito & T. Tamura in [4] where an analysis of the equality case for strictly convex spaces has been performed as well.

REMARK 1. Let \mathbb{C} be the field of complex numbers. If $z = \operatorname{Re} z + i \operatorname{Im} z$, then by $|\cdot|_p : \mathbb{C} \rightarrow [0, \infty)$, $p \in [1, \infty]$ we define the p -modulus of z as

$$|z|_p := \begin{cases} \max \{|\operatorname{Re} z|, |\operatorname{Im} z|\} & \text{if } p = \infty, \\ (|\operatorname{Re} z|^p + |\operatorname{Im} z|^p)^{\frac{1}{p}} & \text{if } p \in [1, \infty), \end{cases}$$

where $|a|$, $a \in \mathbb{R}$ is the usual modulus of the real number a .

For $p = 2$, we recapture the usual modulus of a complex number, i.e.,

$$|z|_2 = \sqrt{|\operatorname{Re} z|^2 + |\operatorname{Im} z|^2} = |z|, \quad z \in \mathbb{C}.$$

It is well known that $(\mathbb{C}, |\cdot|_p)$, $p \in [1, \infty]$ is a Banach space over the real number field \mathbb{R} .

It is obvious that all above inequalities hold for $|\cdot|_p$, the nonnegative n -tuples \mathbf{p}, \mathbf{q} and the complex numbers z_1, \dots, z_n . The details are omitted.

4. The Geometric Mean-Arithmetic Mean Inequality

Although the inequality between the arithmetic and geometric means

$$\frac{a+b}{2} \geq \sqrt{ab}, \quad a, b \geq 0;$$

was probably known in antiquity, the general result for weighted means seems to have first appeared in print in the nineteenth century, in the notes of Cauchy's course given at the École Royale in 1821, page 315.

If $\mathbf{p} \in \mathcal{P}_n$, $\mathbf{a} \geq \mathbf{0}$ and

$$G_n(\mathbf{p}, \mathbf{a}) := \prod_{i=1}^n a_i^{p_i}, \quad A_n(\mathbf{p}, \mathbf{a}) := \sum_{i=1}^n p_i a_i$$

then

$$(4.1) \quad A_n(\mathbf{p}, \mathbf{a}) \geq G_n(\mathbf{p}, \mathbf{a}).$$

If all $p_i > 0$, then the equality case holds in (4.1) if and only if \mathbf{a} is constant, i.e., all the components are equal to a constant $k \in \mathbb{R}$.

For classical and recent results related to the GA-inequality, see Chapter II of the book [1].

Applying the inequality (2.1) for the choices $f : (0, \infty) \rightarrow \mathbb{R}$, $f(x) = -\ln x$ one can state the following inequality

$$(4.2) \quad \left(\frac{A_n(\mathbf{q}, \mathbf{a})}{G_n(\mathbf{q}, \mathbf{a})} \right)^{\max_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\}} \geq \frac{A_n(\mathbf{p}, \mathbf{a})}{G_n(\mathbf{p}, \mathbf{a})} \geq \left(\frac{A_n(\mathbf{q}, \mathbf{a})}{G_n(\mathbf{q}, \mathbf{a})} \right)^{\min_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\}}.$$

If, for the uniform distribution $\mathbf{u} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)$, we consider

$$A_n(\mathbf{a}) = A_n(\mathbf{u}, \mathbf{a}), G_n(\mathbf{a}) = G_n(\mathbf{u}, \mathbf{a}),$$

then we have

$$(4.3) \quad \left(\frac{A_n(\mathbf{a})}{G_n(\mathbf{a})} \right)^{n \max_{1 \leq i \leq n} \{p_i\}} \geq \frac{A_n(\mathbf{p}, \mathbf{a})}{G_n(\mathbf{p}, \mathbf{a})} \geq \left(\frac{A_n(\mathbf{a})}{G_n(\mathbf{a})} \right)^{n \min_{1 \leq i \leq n} \{p_i\}}.$$

Since there is the obvious relation between the arithmetic mean and the harmonic mean

$$H_n(\mathbf{p}, \mathbf{a}) := \frac{1}{\sum_{i=1}^n \frac{p_i}{a_i}} = A_n \left(\mathbf{p}, \frac{1}{\mathbf{a}} \right)^{-1}$$

where $\frac{1}{\mathbf{a}} := \left(\frac{1}{a_1}, \dots, \frac{1}{a_n} \right)$ and a_i are all positive, one can derive a similar inequality between the harmonic and geometric means. The details are omitted.

5. Inequalities for Shannon's Entropy

Let X be a random variable with the range $R = \{x_1, \dots, x_n\}$ and the probability distribution p_1, \dots, p_n ($p_i > 0$, $i = 1, \dots, n$). Define the *Shannon entropy* by

$$H(X) := - \sum_{i=1}^n p_i \ln p_i.$$

The following theorem is well known in the literature and concerns the maximum possible value of $H(X)$ in terms of the size of R [5, p. 27].

THEOREM 2. *Let X be defined as above. Then*

$$(5.1) \quad 0 \leq H(X) \leq \ln n.$$

Furthermore, $H(X) = 0$ iff $p_i = 1$ for some i and $H(X) = \ln n$ iff $p_i = \frac{1}{n}$ for all $i \in \{1, \dots, n\}$.

This fundamental result may be improved as follows:

PROPOSITION 1. *Let X be defined as above. If $p_i > 0$ for each $i \in \{1, \dots, n\}$, then*

$$(5.2) \quad \begin{aligned} 0 &\leq n \min_{1 \leq i \leq n} \{p_i\} \ln \left[\frac{A_n \left(\frac{1}{\mathbf{p}} \right)}{G_n \left(\frac{1}{\mathbf{p}} \right)} \right] \\ &\leq \ln n - H(X) \leq n \max_{1 \leq i \leq n} \{p_i\} \ln \left[\frac{A_n \left(\frac{1}{\mathbf{p}} \right)}{G_n \left(\frac{1}{\mathbf{p}} \right)} \right]. \end{aligned}$$

PROOF. It follows from Corollary 1 for the choices $f(t) = -\ln t$, $x_i = 1/p_i$ and appropriate elementary calculations. The details are omitted. ■

6. Inequalities for the Relative Entropy

The *relative entropy* is a measure of the distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy $D(\mathbf{p}||\mathbf{q})$ is a measure of the inefficiency of assuming that the distribution is \mathbf{q} when the true distribution is \mathbf{p} . For example, if we knew the true distribution of the random variable, then we could construct a code with average description length $H(\mathbf{p})$. If, instead, we used the code for a distribution \mathbf{q} , we would need $H(\mathbf{p}) + D(\mathbf{p}||\mathbf{q})$ bits on the average to describe the random variable [2, p. 18].

DEFINITION 1. *The relative entropy or Kullback-Leibler distance between two probability distributions \mathbf{p} and \mathbf{q} is defined by*

$$D(\mathbf{p}||\mathbf{q}) := \sum_{i=1}^n p_i \ln \left(\frac{p_i}{q_i} \right) = E_{\mathbf{p}} \ln \left(\frac{\mathbf{p}(X)}{\mathbf{q}(X)} \right).$$

In the above definition, we use the convention (based on continuity arguments) that $0 \ln \left(\frac{0}{q} \right) = 0$ and $p \ln \left(\frac{p}{0} \right) = \infty$.

It is well-known that relative entropy is always non-negative and is zero if and only if $\mathbf{p} = \mathbf{q}$. However, it is not a true distance between distributions since it is not symmetric and does not satisfy the triangle inequality.

The following theorem is of fundamental importance [2, p. 26].

THEOREM 3. (*Information Inequality*) *Let \mathbf{p}, \mathbf{q} , be two probability distributions. Then*

$$(6.1) \quad D(\mathbf{p}||\mathbf{q}) \geq 0$$

with equality if and only if

$$(6.2) \quad p_i = q_i \text{ for all } x \in \{1, \dots, n\}.$$

This fundamental result may be improved as follows.

PROPOSITION 2. *Let \mathbf{p}, \mathbf{q} , be two probability distributions. Define the χ^2 divergence of \mathbf{p}, \mathbf{q} by the quantity*

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

If $p_i, q_i > 0$ for all $x \in \{1, \dots, n\}$, then

$$\begin{aligned} & \max_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\} [D(\mathbf{q} \parallel \mathbf{p}) - \ln(\chi^2(\mathbf{p}, \mathbf{q}) + 1)] \\ & \geq D(\mathbf{p} \parallel \mathbf{q}) \\ & \geq \min_{1 \leq i \leq n} \left\{ \frac{p_i}{q_i} \right\} [D(\mathbf{q} \parallel \mathbf{p}) - \ln(\chi^2(\mathbf{p}, \mathbf{q}) + 1)] (\geq 0). \end{aligned}$$

PROOF. The proof follows by Theorem 1 on choosing $f(t) = -\ln t$, $x_i = \frac{q_i}{p_i}$ and performing appropriate elementary calculations. The details are omitted. ■

References

- [1] P.S. BULLEN, *Handbook of Mean and Their Inequalities*, Kluwer Academic Publishers, Dordrecht/Boston/London, 2003.
- [2] T.M. COVER and J.A. THOMAS, *Elements of Information Theory*, John Wiley & Sons Inc., 1991.
- [3] S.S. DRAGOMIR and C.E.M. PEARCE, *Selected Topics on Hermite-Hadamard Inequalities and Applications*, RGMIA Monographs, Victoria University, 2000. (ONLINE: <http://rgmia.vu.edu.au/monographs/>).
- [4] M. KATO, K.-S. SAITO and T. TAMURA, Sharp triangle inequality and its reverse in Banach spaces, *submitted*.
- [5] R.J. McELIECE, *The Theory of Information and Coding*, Addison Wesley Publishing Company, Reading, 1977.
- [6] D.S. MITRINOVIĆ, J.E. PEČARIĆ and A.M. FINK, *Classical and New Inequalities in Analysis*, Kluwer Academic Publishers, Dordrecht, 1993.
- [7] J. E. PEČARIĆ, F. PROSCHAN and Y.L. TONG, *Convex Functions, Partial Orderings and Statistical Applications*, Academic Press, 1992.

SCHOOL OF COMPUTER SCIENCE & MATHEMATICS, VICTORIA UNIVERSITY, MELBOURNE, VICTORIA, AUSTRALIA

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

URL: <http://rgmia.vu.edu.au/dragomir/>