Publ. Math. Debrecen
73/3-4 (2008), 461–469

On a class of mean values including Jensen functionals

By SLAVKO SIMIC (Belgrade)

Abstract. We introduce a new class of multi-variable mean values with a quotient of two Jensen functionals as an argument. By some examples in information theory its fruitfulness is shown.

1. Introduction

At the beginning let us remind the reader about some well known mathematical entities we shall need in the sequel.

1.1. Denote by P_n the set of all positive n-tuples $(p_1, p_2, \dots p_n)$, $n \geq 2$, with the property $\sum_1^n p_i = 1$.

The Jensen functional $J_n(p, x; f)$ is defined by

$$J_n(p, x; f) := \sum_1^n p_i f(x_i) - f\left(\sum_1^n p_i x_i\right),$$

where $f : I \to \mathbb{R}$ be a (strictly) convex function on an interval I, $x = (x_1, x_2, \dots, x_n) \in I^n$ and $p \in P_n$.

The famous Jensen's inequality asserts that

$$J_n(p, x; f) \geq 0, \tag{1}$$

with the equality case if and only if $x_1 = x_2 = \cdots = x_n$.

Jensen's inequality is one of the most known and extensively used inequalities in various fields of mathematics. For classical and recent developments related to this inequality see ([2], [4]) where further references are given.

Mathematics Subject Classification: 39B22, 26D20.
Key words and phrases: Jensen functional, divided differences, Cauchy means, Csiszár's f-divergence.

1.2. To define the Cauchy mean value for several variables, we need a notion of divided differences ([1]).

Definition 1. For a function $f : I \to \mathbb{R}$, I being a real interval, the divided differences of f on distinct points $x_i \in I$ are defined inductively by

$$[x_1]_f := f(x_1),$$

$$[x_1, \dots, x_n]_f := \frac{[x_1, \dots, x_{n-1}]_f - [x_2, \dots, x_n]_f}{x_1 - x_n} \quad (n = 2, 3, \dots).$$

For instance,

$$[x_1, x_2]_f = \frac{f(x_1) - f(x_2)}{x_1 - x_2}; \quad [x_1, x_2, x_3]_f = \frac{(x_2 - x_3)f(x_1) + (x_3 - x_1)f(x_2) + (x_1 - x_2)f(x_3)}{(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)}, \tag{2}$$

etc.

The following *Cauchy mean value theorem for divided differences* is due to Leach and Sholander [3].

Theorem A. *Let $x_1 < \cdots < x_n$ and assume that $f^{(n-1)}, g^{(n-1)}$ exist with $g^{(n-1)}(u) \neq 0$ on $[x_1, x_n]$. Then there is a $t \in (x_1, x_n)$ such that*

$$\frac{f^{(n-1)}(t)}{g^{(n-1)}(t)} = \frac{[x_1, \dots, x_n]_f}{[x_1, \dots, x_n]_g}.$$

Supposing that the function $u \to \frac{f^{(n-1)}(u)}{g^{(n-1)}(u)}$ is invertible, we get that

$$\left(\frac{f^{(n-1)}}{g^{(n-1)}}\right)^{-1} \left(\frac{[x_1, \dots, x_n]_f}{[x_1, \dots, x_n]_g}\right)$$

is a mean value of numbers $x_1, \dots, x_n$.

The aim of this paper is to give a mean value theorem for Jensen functionals i.e., to bound a Jensen functional with another one. This is done by the following

Theorem B. *Let $f, g : I \to \mathbb{R}$ be twice continuously differentiable strictly convex functions and $\phi : I \to \mathbb{R}^+$ be a continuous and strictly monotonic function on I. Let $x \in I$, p be defined as above. Then the expression*

$$\phi^{-1}\left(\frac{J_n(p, x; f)}{J_n(p, x; g)}\right) \quad (n \geq 2),$$

represents a mean value of numbers $x_1, \dots, x_n$*, that is*

$$\min\{x_1, \dots, x_n\} \le \phi^{-1}\left(\frac{J_n(p,x;f)}{J_n(p,x;g)}\right) \le \max\{x_1, \dots, x_n\}, \tag{3}$$

if and only if the relation

$$f''(t) = \phi(t) g''(t) \tag{4}$$

holds for each $t \in I$.

Proof. We prove firstly the necessity of the condition (4).

Applying (3) in case $n = 2$ for arbitrary $x_1, x_2 \in I;\ x_1 \ne x_2$, we get

$$\min\{x_1, x_2\} \le \phi^{-1}\left(\frac{J_2(p,x;f)}{J_2(p,x;g)}\right) \le \max\{x_1, x_2\}.$$

Hence

$$\lim_{x_1 \to x_2} \phi^{-1}\left(\frac{J_2(p,x;f)}{J_2(p,x;g)}\right) = x_2. \tag{5}$$

But, since ϕ (i.e. ϕ^{-1}) is continuous on I, due to l'Hospital's rule we obtain

$$\begin{aligned}\lim_{x_1 \to x_2} \phi^{-1}\left(\frac{J_2(p,x;f)}{J_2(p,x;g)}\right) &= \lim_{x_1 \to x_2} \phi^{-1}\left(\frac{p_1 f'(x_1) - p_1 f'(p_1x_1 + p_2x_2)}{p_1 g'(x_1) - p_1 g'(p_1x_1 + p_2x_2)}\right) \\ = \lim_{x_1 \to x_2} \phi^{-1}\left(\frac{f''(x_1) - p_1 f''(p_1x_1 + p_2x_2)}{g''(x_1) - p_1 g''(p_1x_1 + p_2x_2)}\right) &= \phi^{-1}(f''(x_2)/g''(x_2)). \end{aligned}\tag{6}$$

Comparing (5) and (6), the desired result follows.

Now, assuming the condition (4), we shall prove the inequalities (3) by induction on n.

Note that, for $x_1, x_2 \in I$, $x_1 \ne x_2$,

$$\frac{J_2(p,x;f)}{J_2(p,x;g)} = \frac{[x_1, x_2, p_1x_1 + p_2x_2]_f}{[x_1, x_2, p_1x_1 + p_2x_2]_g}. \tag{7}$$

Since $\phi(\cdot)$ is strictly monotone on I, it follows by (4) that f''/g'' is invertible.

Applying Theorem A for $n = 3$, $x_3 = p_1x_1 + p_2x_2$, we get

$$\begin{aligned}\min\{x_1, x_2\} = \min\{x_1, x_2, p_1x_1 + p_2x_2\} &\le \left(\frac{f''}{g''}\right)^{-1}\left(\frac{[x_1, x_2, p_1x_1 + p_2x_2]_f}{[x_1, x_2, p_1x_1 + p_2x_2]_g}\right) \\ &\le \max\{x_1, x_2, p_1x_1 + p_2x_2\} = \max\{x_1, x_2\},\end{aligned}$$

i.e., by (4) and (7),

$$\min\{x_1, x_2\} \le \phi^{-1}\left(\frac{J_2(p,x;f)}{J_2(p,x;g)}\right) \le \max\{x_1, x_2\}. \tag{8}$$

Therefore, the relation (3) is proved in case $n = 2, x \in I$, $p \in P_2$.

Next, it is not difficult to check the identity

$$J_n(p, x; f) := \sum_1^n p_i f(x_i) - f\left(\sum_1^n p_i x_i\right) = (1 - p_n)\left(\sum_1^{n-1} p_i' f(x_i) - f\left(\sum_1^{n-1} p_i' x_i\right)\right)$$
$$+ [(1 - p_n)f(T) + p_n f(x_n) - f((1 - p_n)T + p_n x_n)],$$

where

$$T := \sum_1^{n-1} p_i' x_i;\ p_i' := p_i/(1 - p_n), \quad i = 1, 2, \ldots, n - 1;\ \sum_1^{n-1} p_i' = 1.$$

Suppose firstly that $\phi(\cdot)$ is increasing on I.

In this case, by induction hypothesis and (8), we get

$$\begin{aligned} J_n(p, x; f) &\le \phi(\max\{x_1, x_2, \ldots x_{n-1}\})(1 - p_n)\left(\sum_1^{n-1} p_i' g(x_i) - g\left(\sum_1^{n-1} p_i' x_i\right)\right) \\ &\quad + \phi(\max\{T, x_n\})\big[(1 - p_n)g(T) + p_n g(x_n) - g((1 - p_n)T + p_n x_n)\big] \\ &\le \phi(\max\{x_1, x_2, \ldots, x_n\})((1 - p_n)\left(\sum_1^{n-1} p_i' g(x_i) - g\left(\sum_1^{n-1} p_i' x_i\right)\right) \\ &\quad + [(1 - p_n)g(T) + p_n g(x_n) - g((1 - p_n)T + p_n x_n)]) \\ &= \phi(\max\{x_1, x_2, \ldots, x_n\}) J_n(p, x; g). \end{aligned}$$

Analogously,

$$J_n(p, x; f) \ge \phi(\min\{x_1, x_2, \ldots, x_n\}) J_n(p, x; g).$$

Since $\phi^{-1}(\cdot)$ is also monotone increasing on I, we finally get

$$\min\{x_1, \ldots, x_n\} \le \phi^{-1}\left(\frac{J_n(p, x; f)}{J_n(p, x; g)}\right) \le \max\{x_1, \ldots, x_n\}.$$

If $\phi(\cdot)$ is monotone decreasing on I, write $\phi(\cdot) = 1/\tilde{\phi}(\cdot)$ where $\tilde{\phi}(t) = 1/\phi(t)$ is a monotone increasing function, and proceed as above.

Particularly interesting is further examination of the means of type (3). Here is an instructive example.

Consider the triple $f(t) = t^s/s(s-1)$, $g(t) = t^r/r(r-1)$, $\phi(t) = t^{s-r}$; $t \in \mathbb{R}^+$, $s, r \in \mathbb{R}/\{0, 1\}$, $s \ne r$.

According to Theorem B, the expression

$$W_{r,s}(p,x) := \left(\frac{r(r-1)(\sum_1^n p_i x_i^s - (\sum_1^n p_i x_i)^s)}{s(s-1)\big(\sum_1^n p_i x_i^r - (\sum_1^n p_i x_i)^r\big)}\right)^{\frac{1}{s-r}},$$
$$n \geq 2,\ rs(r-1)(s-1)(s-r) \neq 0,$$

represents a mean value of positive numbers $x_1, \ldots, x_n$ for any $p \in P_n$.

By an appropriate limit process, this class of means can be continuously extended to the whole r, s plane. In this fashion we get

Definition 2. The means $W_{r,s}(p,x)$ are defined by the following

(a) $$W_{r,s}(p,x) := \left(\frac{r(r-1)(\sum p_i x_i^s - (\sum p_i x_i)^s)}{s(s-1)(\sum p_i x_i^r - (\sum p_i x_i)^r)}\right)^{1/(s-r)},$$
where $rs(r-1)(s-1)(r-s) \neq 0$;

(b) $$W_{s,s}(p,x) = \exp\left(\frac{\sum p_i x_i^s \log x_i - (\sum p_i x_i)^s \log(\sum p_i x_i)}{\sum p_i x_i^s - (\sum p_i x_i)^s} - \frac{1}{s} - \frac{1}{s-1}\right),$$
where $s(s-1) \neq 0$;

(c) $$W_{0,s}(p,x) = \left(\frac{\sum p_i x_i^s - (\sum p_i x_i)^s}{s(s-1)(\log(\sum p_i x_i) - \sum p_i \log x_i)}\right)^{1/s},$$
where $s(s-1) \neq 0$;

(d) $$W_{1,s}(p,x) = \left(\frac{\sum p_i x_i^s - (\sum p_i x_i)^s}{s(s-1)(\sum p_i x_i \log x_i - (\sum p_i x_i)\log(\sum p_i x_i))}\right)^{1/(s-1)},$$
where $s(s-1) \neq 0$;

(e) $$W_{0,1}(p,x) = \frac{\sum p_i x_i \log x_i - (\sum p_i x_i)\log(\sum p_i x_i)}{\log(\sum p_i x_i) - \sum p_i \log x_i},$$

(f) $$W_{0,0}(p,x) = \exp\left(\frac{\log^2(\sum p_i x_i) - \sum p_i \log^2 x_i}{2(\log(\sum p_i x_i) - \sum p_i \log x_i)} + 1\right),$$

(g) $$W_{1,1}(p,x) = \exp\left(\frac{\sum p_i x_i \log^2 x_i - (\sum p_i x_i)\log^2(\sum p_i x_i)}{2(\sum p_i x_i \log x_i - (\sum p_i x_i)\log(\sum p_i x_i))} - 1\right).$$

Since $W_{s,r}(p,x) = W_{r,s}(p,x)$, by the above the class of means $W_{r,s}(p,x)$ is well defined for each $r, s \in \mathbb{R}$.

Also, we obtain a whole variety of multi-variable mean values deserving further examination (see [10]).

2. Applications

Bounding a Jensen functional with another one, which is the content of Theorem B, can be used in many parts of analysis, numerical analysis, probability, etc. We are restricted here to some examples from information theory.

Let

$$\Omega_n = \left\{ p = \{p_i\}_1^n \,\middle|\, p_i > 0,\ \sum_1^n p_i = 1 \right\}, n \geq 2,$$

be the set of finite discrete probability distributions.

2.1. One of the most general probability measures which is of importance in information theory is the famous Csiszár's f-divergence $C_f(p\|q)$ ([7]), defined by

Definition 3. For a convex function $f : (0, \infty) \to \mathbb{R}$, the f-divergence measure is given by

$$C_f(p\|q) := \sum q_i f(p_i/q_i),$$

where $p, q \in \Omega_n$.

By Jensen's inequality for convex functions it follows that

$$C_f(p\|q) \geq f(1),$$

with equality iff $p = q$ in the case of strict convexity of f.

The distribution p represents here data or observations, while q typically represents a model or an approximation of p.

Some important information measures are just particular cases of the Csiszár's f-divergence.

For example,

(a) taking $f(x) = -x^{1/2}$, we obtain the negative value of Bhattacharya constant([6]) defined by

$$B = B(p\|q) := \sum \sqrt{p_i q_i}\,;$$

(b) for $f(x) = x \log x$, we obtain the Kullback–Leibler divergence ([5]) defined by

$$K(p\|q) := \sum p_i \log(p_i/q_i);$$

(c) for $f(x) = (\sqrt{x} - 1)^2$, we obtain the Hellinger distance

$$H^2(p, q) := \sum (\sqrt{p}_i - \sqrt{q}_i)^2;$$

(d) if we choose $f(x) = (x-1)^2$, then we get the χ^2-distance

$$\chi^2(p,q) := \sum (p_i - q_i)^2/q_i.$$

We quote now inequalities between those measures which are already known in the literature.

$$\chi^2(p,q) \geq H^2(p,q); \quad K(p\|q) \geq H^2(p,q); \quad K(p\|q) \leq \log(1+\chi^2(p,q)). \tag{$*$}$$

In particular, $K(p\|q) \leq \chi^2(p,q)$.

2.2. The generalized measure $K_s(p\|q)$, known as *the relative divergence of type* s ([8], [9]), is defined by

$$K_s(p\|q) := \begin{cases} (\sum p_i^s q_i^{1-s} - 1)/s(s-1), & s \in \mathbb{R}/\{0,1\}; \\ K(q\|p), & s = 0; \\ K(p\|q), & s = 1. \end{cases}$$

It include the Hellinger and χ^2 distances as particular cases.
Indeed,

$$K_{1/2}(p\|q) = 4\Big(1 - \sum \sqrt{p_i q_i}\,\Big) = 2\sum (p_i + q_i - 2\sqrt{p_i q_i}\,) = 2H^2(p,q);$$
$$K_2(p\|q) = \frac{1}{2}\Big(\sum \frac{p_i^2}{q_i} - 1\Big) = \frac{1}{2}\sum \frac{(p_i - q_i)^2}{q_i} = \frac{1}{2}\chi^2(p,q).$$

Now, we are enabled to formulate two theorems related to the bounds of Csiszár's f-divergence.

Theorem C. *Let $f, g : (0,\infty) \to \mathbb{R}$ be twice continuously differentiable strictly convex functions and assume that f''/g'' is invertible. Then the inequality*

$$\min\left\{\frac{p}{q}\right\} \leq \left(\frac{f''}{g''}\right)^{-1}\left(\frac{C_f(p\|q) - f(1)}{C_g(p\|q) - g(1)}\right) \leq \max\left\{\frac{p}{q}\right\},$$

holds for each $p, q \in \Omega_n$.

Proof. Noting that for $p, q \in \Omega_n$,

$$J_n\left(q, \frac{p}{q}; f\right) = \sum_1^n q_i f\left(\frac{p_i}{q_i}\right) - f\left(\sum_1^n q_i \frac{p_i}{q_i}\right) = C_f(p\|q) - f(1),$$

the proof follows applying Theorem B. $\square$

A similar assertion related to the relative divergence of type s is the next

Theorem D. *For each $r, s \in \mathbb{R}$, $r \neq s$; $p, q \in \Omega_n$, we have*

$$\min\left\{\frac{p}{q}\right\} \leq \left(\frac{K_s(p\|q)}{K_r(p\|q)}\right)^{\frac{1}{s-r}} \leq \max\left\{\frac{p}{q}\right\}.$$

Proof. Consider the function $f_s(t)$ defined to be

$$f_s(t) = \begin{cases} (t^s - st + s - 1)/s(s-1), & s(s-1) \neq 0; \\ t - \log t - 1, & s = 0; \\ t \log t - t + 1, & s = 1. \end{cases}$$

Since

$$f'_s(t) = \begin{cases} \dfrac{t^{s-1} - 1}{s-1}, & s(s-1) \neq 0; \\ 1 - \dfrac{1}{t}, & s = 0; \\ \log t, & s = 1, \end{cases}$$

and

$$f''_s(t) = t^{s-2}, \quad s \in \mathbb{R},\ t > 0,$$

it follows that $f_s(t)$ is a twice continuously differentiable convex function for $s \in \mathbb{R}$, $t \in \mathbb{R}^+$.

Also, for $p, q \in \Omega_n$, we have

$$\frac{J_n(q, \frac{p}{q}; f_s)}{J_n(q, \frac{p}{q}; f_r)} = \frac{K_s(p\|q)}{K_r(p\|q)},$$

and $\phi(t) = f''_s(t)/f''_r(t) = t^{s-r}$, which is invertible for $r \neq s$.

Hence, applying Theorem B we obtain the desired result.

Theorems C and D give an opportunity for instant comparison of two probability measures connected with the Csiszár's f-divergence. □

For example,

Theorem E. *For $p, q \in \Omega_n$, we have*

$$\min_{1 \leq i \leq n} (q_i/p_i) \leq \frac{K(q\|p)}{K(p\|q)} \leq \max_{1 \leq i \leq n} (q_i/p_i);$$

$$\min_{1 \leq i \leq n} (q_i/p_i)^3 \leq \frac{\chi^2(q,p)}{\chi^2(p,q)} \leq \max_{1 \leq i \leq n} (q_i/p_i)^3;$$

$$2\Big(\min_{1\le i\le n}(p_i/q_i)^{1/2}\Big)\le\frac{K(p\|q)}{H^2(p,q)}\le 2\Big(\max_{1\le i\le n}(p_i/q_i)^{1/2}\Big);$$

$$3+\log\Big(\min_{1\le i\le n}p_i/q_i\Big)\le\frac{K(p\|q)-2B^2\log B}{1-B^2}\le 3+\log\Big(\max_{1\le i\le n}p_i/q_i\Big),$$

where the last assertion follows from Definition 2 part (b), *taking* $s=2$, $x_i=\sqrt{q_i/p_i}$, $i=1,\dots,n$.

Details are left to the reader.

References

[1] P. S. Bullen, Handbook of Means and their Inequalities, *Kluwer Academic Publishers*, *Dordrecht*, 2003.

[2] D. S. Mitrinovic, J. E. Pecaric and A. M. Fink, Classical and New Inequalities in Analysis, *Kluwer Academic Publishers*, *Dordrecht*, 1993.

[3] E. Leach and M. Sholander, Multi-variable extended mean values, *J. Math. Anal. Appl.* **104** (1984), 390–407.

[4] D. S. Mitrinovic, Analytic Inequalities, *Springer*, *New York*, 1970.

[5] S. Kullback and R. A. Leibler, On information and sufficiency, *Annals of Mathematical Statistics* **22** (1951), 79–86.

[6] A. Rényi, On measures of Entropy and Information, Proc. IV Berkeley Symp. Math. Statist. Prob., Vol. 1., *University of California Press*, *Berkeley*, 1961.

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

[8] I. J. Taneja, New developments in generalized information measures, *Advances in Imaging and Electron Physics* **91** (1995), 37–135.

[9] I. Vajda, Theory of Statistical Inference and Information, *London*, 1989.

[10] S. Simic, On logarithmic convexity for differences of power means, *J. Ineq. Appl.* (2007), Article ID 37359 ID 37359, 8 pp.

SLAVKO SIMIC
MATHEMATICAL INSTITUTE SANU
KNEZA MIHAILA 36
11000 BELGRADE
SERBIA

E-mail: ssimicturing.mi.sanu.ac.yu

(Received May 5, 2008; revised September 30, 2008)