PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 138, Number 7, July 2010, Pages 2457–2462
S 0002-9939(10)10353-0
Article electronically published on March 10, 2010

BEST POSSIBLE GLOBAL BOUNDS FOR JENSEN FUNCTIONAL

SLAVKO SIMIC

(Communicated by Edward C. Waymire)

ABSTRACT. We determine the form of the best possible global bounds for the Jensen functional on the real line. Thereby, previous results on this topic are essentially improved. Some applications in Analysis and Information Theory are also given.

1. INTRODUCTION

The form of the Jensen functional $J_f(\mathbf{p},\mathbf{x})$ is given by (cf. [3]),

$$J_f(\mathbf{p},\mathbf{x}) := \sum p_i f(x_i) - f(\sum p_i x_i),$$

where $f(\cdot)$ stands for a convex function defined on the domain D_f, $D_f \subseteq \mathbb{R}$, $\mathbf{x} := \{x_i\}$ is a finite sequence of numbers from D_f and $\mathbf{p} := \{p_i\}$, $\sum p_i = 1$ denotes a positive weight sequence associated with $\mathbf{x}$.

Throughout the paper we assume that all terms of the sequence $\mathbf{x}$ belong to some closed interval I, i.e., that for some fixed a, b: $x_i \in [a,b] := I \subseteq D_f,\ i = 1, 2, \ldots$.

By definition (cf. [9]), the global bounds for $J_f(\mathbf{p},\mathbf{x})$ will depend only on f and I . For instance, the famous Jensen's inequality asserts that for $\mathbf{x} \in I$ and an arbitrary $\mathbf{p}$,

$$0 \le J_f(\mathbf{p},\mathbf{x}).$$

Jensen's inequality is one of the most known and extensively used inequalities in various fields of mathematics. Some important inequalities are just particular cases of this inequality, such as the weighted $A-G-H$ inequality, the Cauchy inequality, the Ky Fan and Hölder inequalities, etc. For classical and recent developments related to Jensen's inequality, see ([2], [7]), where further references are given.

By closer examination, one can see that the lower bound zero is of global nature since it does not depend on $\mathbf{p}$ or $\mathbf{x}$ but only on f and the interval I whereupon f is convex. Also, it is not difficult to show that zero is the best possible global lower bound for the Jensen functional.

In the same sense, an upper global bound for differentiable convex mappings has been given in [4] and for arbitrary convex mappings in [9], [8]. There are a number of papers where the global bounds are utilized in applications concerning

Received by the editors September 27, 2009.
2010 *Mathematics Subject Classification.* Primary 26B25, 26D20.
Key words and phrases. Jensen's discrete inequality, global bounds, Lagrange means, generalized $A-G-H$ inequality.

some parts of Analysis, Numerical Analysis, Information Theory, etc. (cf. [1], [2], [4], [5]).

The problem of best possible upper bounds is solved in some particular cases. For example, it is proved in [10] that the Specht ratio $S(t)$ is the best possible upper bound for the $A-G$ inequality with uniform weights, that is,

$$S(t)(x_1x_2\cdots x_n)^{1/n} \geq \frac{x_1+x_2+\cdots+x_n}{n} \; (\geq (x_1x_2\cdots x_n)^{1/n}),$$

where $S(t) := \frac{t^{1/(t-1)}}{e\log t^{1/(t-1)}}$ with $t=b/a$.

In this article we shall give an explicit form of the best possible global bound $T_f(a,b)$ for the Jensen functional. As an application, we determine $T_f(a,b)$ in the case of the generalized $A-G-H$ inequality as a combination of some well-known classical means. Similarly, new bounds are found for some probability measures which are of importance in Information Theory, thereby improving the results from [8].

2. Results and proofs

Our main result is contained in the following.

Theorem A. *Let f, $\mathbf{p},\mathbf{x}$ be defined as above and $p,q\geq 0$, $p+q=1$. Then*

(1)
$$J_f(\mathbf{p},\mathbf{x}) := \sum p_i f(x_i) - f(\sum p_i x_i) \leq \max_p [pf(a)+qf(b)-f(pa+qb)] := T_f(a,b).$$

This upper bound is very precise. For example,

$$T_{x^2}(a,b) = \max_p (pa^2+qb^2-(pa+qb)^2) = \max_p (pq(b-a)^2) = \frac{1}{4}(b-a)^2,$$

and we obtain at once the well-known pre-Grüss inequality [7],

$$\sum p_i x_i^2 - (\sum p_i x_i)^2 \leq \frac{1}{4}(b-a)^2,$$

with the constant $1/4$ as the best possible.

Proof. Since $x_i \in [a,b]$, there is a sequence $\{\lambda_i\}$, $\lambda_i\in[0,1]$, such that $x_i=\lambda_i a+(1-\lambda_i)b$, $i=1,2,\cdots$.

Hence,

$$\begin{aligned}\sum p_i f(x_i) - f(\sum p_i x_i) &= \sum p_i f(\lambda_i a+(1-\lambda_i)b) - f(\sum p_i(\lambda_i a+(1-\lambda_i)b)\\ &\leq \sum p_i(\lambda_i f(a)+(1-\lambda_i)f(b)) - f(a\sum p_i\lambda_i + b\sum p_i(1-\lambda_i))\\ &= f(a)(\sum p_i\lambda_i) + f(b)(1-\sum p_i\lambda_i) - f(a(\sum p_i\lambda_i)+b(1-\sum p_i\lambda_i)).\end{aligned}$$

Denoting $\sum p_i\lambda_i := p$, $1-\sum p_i\lambda_i := q$, we have that $0\leq p,q\leq 1$, $p+q=1$.

Consequently,

$$\begin{aligned}J_f(\mathbf{p},\mathbf{x}) := \sum p_i f(x_i) - f(\sum p_i x_i) &\leq pf(a)+qf(b)-f(pa+qb)\\ &\leq \max_p [pf(a)+qf(b)-f(pa+qb)] := T_f(a,b),\end{aligned}$$

and the proof of Theorem A is done. □

Theorem B. *For any $a,b\in D_f$, $a\neq b$, the bound $T_f(a,b)$ exists and is unique.*

Proof. Denote $g(p) := pf(a) + (1-p)f(b) - f(pa + (1-p)b)$ and, for fixed $p_1, p_2 \in (0,1)$, $p_1 a + (1-p_1)b := c_1$, $p_2 a + (1-p_2)b := c_2$. Hence $c_1, c_2 \in D_f$, and from the identity

$$\begin{aligned} &tg(p_1) + (1-t)g(p_2) - g(tp_1 + (1-t)p_2) \\ &\quad = f(tc_1 + (1-t)c_2) - tf(c_1) - (1-t)f(c_2),\ t \in (0,1), \end{aligned}$$

we conclude that $g(p)$ is concave for $0 \le p \le 1$ with $g(0) = g(1) = 0$. Therefore there exists a unique positive $g(p_0) = \max_p g(p) = T_f(a,b)$ attained at the unique $p_0 = p_0(a,b) \in (0,1)$. □

A way of determining the number p_0 for a differentiable convex mapping is given by the following.

Theorem C. *For fixed $a, b \in D_f$, $a \neq b$, denote by $\theta_f(a,b)$ the well known Lagrange mean value*

$$\theta_f(a,b) := (f')^{-1}\Big(\frac{f(b) - f(a)}{b-a}\Big).$$

The number p_0 should be taken such that the relation

$$p_0 a + (1-p_0)b = \theta_f(a,b)$$

is fulfilled.

Proof. Since $T_f(a,b) = \max_p g(p)$, by a standard argument we have that $g'(p_0) = 0$, i.e.

$$f(b) - f(a) = (b-a)f'(p_0 a + (1-p_0)b).$$

Since f is convex, the function f' is invertible and the result follows. □

We also have

Theorem D. *The bound $T_f(a,b)$ is the best possible global upper bound for the Jensen functional.*

Proof. Indeed, let $T'(a,b)$ be an arbitrary upper global bound. According to the definition, we have that the inequality

$$\sum_{i=1}^{n} p_i f(x_i) - f(\sum_{i=1}^{n} p_i x_i) \le T'(a,b)$$

holds for any weight sequence $\mathbf{p}$, $x_i \in [a,b] \subseteq D_f$ and $n \in \mathbb{N}$. □

In particular, for $n = 2$, $x_1 = a, x_2 = b, p_1 = p_0$, we get

$$T(a,b) \le T'(a,b).$$

Therefore the bound $T_f(a,b)$ is better than any other global bound, hence is the best possible.

3. Applications

Finally, we give some applications of Theorems A-C in Analysis and Information Theory.

3.1. The following well known $A-G-H$ inequality [7] asserts that

$$A(\mathbf{p},\mathbf{x}) \ge G(\mathbf{p},\mathbf{x}) \ge H(\mathbf{p},\mathbf{x}),$$

where

$$A(\mathbf{p},\mathbf{x}) := \sum p_i x_i;\ G(\mathbf{p},\mathbf{x}) := \prod x_i^{p_i},\ H(\mathbf{p},\mathbf{x}) := (\sum p_i/x_i)^{-1},\ \mathbf{x} \in \mathbb{R}^+,$$

are the generalized arithmetic, geometric and harmonic means, respectively.

As an illustration of the above results, we determine some converses of the $A-G-H$ inequality.

Theorem E. *For $x_i \in [a,b],\ 0<a<b, i=1,2,\cdots,$ we have*

(*i*) $0 \le A(\mathbf{p},\mathbf{x}) - H(\mathbf{p},\mathbf{x}) \le 2(A(a,b)-G(a,b))$*;*

(*ii*) $0 \le A(\mathbf{p},\mathbf{x}) - G(\mathbf{p},\mathbf{x}) \le 2(A(a,b)-L(a,b)) - L(a,b)\log\frac{I(a,b)}{L(a,b)}$*;*

(*iii*) $1 \le \frac{A(\mathbf{p},\mathbf{x})}{G(\mathbf{p},\mathbf{x})} \le L(a,b)I(a,b)/G^2(a,b) := \Lambda(a,b)$, *where*

$$A(a,b) := \frac{a+b}{2};$$
$$G(a,b) := \sqrt{ab};$$
$$L(a,b) := \frac{b-a}{\log b - \log a};$$
$$I(a,b) := (b^b/a^a)^{1/(b-a)}/e,$$

are the arithmetic, geometric, logarithmic and identric means, respectively.

Note that

$$0<a<G(a,b)<L(a,b)<I(a,b)<A(a,b)<b.$$

As a consequence, we also get a converse of the $G-H$ inequality:

(*iv*) $1 \le \frac{G(\mathbf{p},\mathbf{x})}{H(\mathbf{p},\mathbf{x})} \le \Lambda(a,b)$.

3.2. Define the probability distributions P and Q of a discrete random variable X by

$$P(X=i)=p_i>0,\ Q(X=i)=q_i>0,\ i=1,2,\cdots,r;\ \sum p_i = \sum q_i = 1.$$

Among the other quantities, of utmost importance in Information Theory are the Kullback-Leibler divergence $D_{KL}(P||Q)$ and Shannon's entropy $H(X)$, defined as

$$D_{KL}(P||Q) := \sum p_i \log\frac{p_i}{q_i};$$
$$H(X) := \sum_1^r p_i \log\frac{1}{p_i}.$$

The distribution P represents here data and observations, while Q typically represents a theoretical model or an approximation of P. Both divergences are always non-negative.

Applying the above results we obtain the following estimates.

Theorem F. (i) *Denoting $m := \min(q_i/p_i),\ M := \max(q_i/p_i),\ i=1,2,\cdots,$ we have*

$$0 \le D_{KL}(P||Q) \le \log\Lambda(m,M).$$

(ii) *Denoting* $\mu := \min\{p_i\},\ \nu := \max\{p_i\}, i = 1, 2, \cdots,$ *we have*

$$0 \le \log r - H(X) \le \log \Lambda(\mu, \nu).$$

It is interesting to compare those results with the results from [8].

Proof of Theorem E. Since $\Theta_{1/x}(a, b) = G(a, b)$, applying Theorems C and A and changing the variable $1/x_i \to x_i$, we obtain the proof of part (i).

Analogously, considering the function $f(x) = e^x$, the proof of (ii) follows.

By Theorem A applied with $f(x) = -\log x$, we obtain

$$0 \le \log \frac{A(\tilde{p}, \tilde{x})}{G(\tilde{p}, \tilde{x})} \le T_{-\log x}(a, b) = \max_p [\log(pa + qb) - p \log a - q \log b].$$

Since $\theta_{-\log x}(a, b) = L(a, b)$, where $L(\cdot, \cdot)$ denotes the logarithmic mean, applying Theorem C we get $p_0 a + (1 - p_0)b = L(a, b)$, i.e. $p_0 = \frac{b - L(a,b)}{b-a}$.

Because $0 < a < b$, we get $0 < p_0 < 1$ and, after some calculation, it follows that

$$\begin{aligned} 0 &\le \log \frac{A(\tilde{p}, \tilde{x})}{G(\tilde{p}, \tilde{x})} \le \log L(a, b) - p_0 \log a - (1 - p_0) \log b \\ &= \log(\frac{b - a}{\log b - \log a}) - \log(ab) + \frac{b \log b - a \log a}{b - a} - 1. \end{aligned}$$

Exponentiating, we obtain the third assertion from Theorem E. □

Remark 1. Note that $\Lambda(a, b)$ is exactly the Specht ratio $S(t)$. Therefore part (iii) of Theorem E shows that Specht's ratio is the best upper bound for the generalized $A - G$ inequality also (cf. [11]).

By the change of variables $x_i \to 1/x_i,\ i = 1, 2, \cdots,$ the proof of the fourth proposition easily follows from the previous one since then

$$A(\mathbf{p}, \frac{1}{\mathbf{x}}) = \frac{1}{H(\mathbf{p}, \mathbf{x})}, \quad G(\mathbf{p}, \frac{1}{\mathbf{x}}) = \frac{1}{G(\mathbf{p}, \mathbf{x})},$$

and

$$\Lambda(1/b, 1/a) = \Lambda(a, b).$$

Proof of Theorem F. A variant of the inequality (iii) from Theorem E asserts that

$$0 \le \log(\sum p_i x_i) - \sum p_i \log x_i \le \log \Lambda(a, b).$$

Putting there $x_i = q_i/p_i,\ i = 1, 2, \cdots$ with $a = m := \min\{x_i\}, b = M := \max\{x_i\},\ i = 1, 2, \cdots,$ and taking into account that $\sum q_i = 1$, the assertion from part (i) follows.

Taking the distribution Q to be uniform, the last result follows from the previous one since

$$\Lambda(\frac{1}{r\nu}, \frac{1}{r\mu}) = \Lambda(\frac{1}{\nu}, \frac{1}{\mu}) = \Lambda(\mu, \nu). \qquad \square$$

References

[1] Beesack P. R., Pecaric J.: "On Jensen's inequality for convex functions", *J. Math. Anal. Appl.*, 110(1985), pp. 536-552. MR805275 (87a:26019)

[2] Budimir I., Dragomir S. S., Pecaric J.: "Further reverse results for Jensen's discrete inequality and applications in Information Theory", *J. Inequal. Pure Appl. Math.*, 2(1)(2001), Art. 5. MR1822298 (2002g:26025)

[3] Dragomir S. S.: "Bounds for the normalized Jensen functional", *Bulletin of the Aust. Math. Soc.*, 74, no. 3(2006), pp. 471-476. MR2273755 (2007f:26021)

[4] Dragomir S. S: "A converse result for Jensen's discrete inequality via Grüss inequality and applications in Information Theory", *Analele Univ. Oradea. Fasc. Math.*, 7(1999-2000), pp. 178-189. MR1774897 (2001f:26028)

[5] Dragomir S. S., Goh C. J.: "Some bounds on entropy measures in Information Theory", *Appl. Math. Letters*, 10, no. 3(1997), pp. 23-28. MR1457634 (98h:94007)

[6] Hardy G. H., Littlewood J. E., Pólya G.: *Inequalities*, Cambridge Univ. Press, Cambridge, 1988. MR0944909 (89d:26016)

[7] Mitrinovic D. S.: *Analytic Inequalities*, Springer, New York, 1970. MR0274686 (43:448)

[8] Simic S.: "Jensen's inequality and new entropy bounds", *Appl. Math. Letters*, 22(2009), pp. 1262–1265. MR2532551

[9] Simic S. "On a global upper bound for Jensen's inequality", *J. Math. Anal. Appl.*, 343 (1) (2008), pp. 414-419. MR2412138 (2009c:26047)

[10] Specht W.: "Zur Theorie der elementaren Mittel", *Math. Z.*, 74(1960), pp. 91–98. MR0117309 (22:8090)

[11] Tominaga M.: "Specht's ratio in the Young inequality", *Sci. Math. Japon.*, 55(2002), pp. 583–588. MR1901045 (2003c:47035)

Mathematical Institute Sanu, Kneza Mihaila 36, 11000 Belgrade, Serbia
E-mail address: ssimic@turing.mi.sanu.ac.rs