Bull. Aust. Math. Soc. **87** (2013), 177–194
doi:10.1017/S0004972712001098

SOME REVERSES OF THE JENSEN INEQUALITY WITH APPLICATIONS

S. S. DRAGOMIR

(Received 23 February 2012; accepted 21 November 2012; first published online 7 February 2013)

Abstract

Two new reverses of the celebrated Jensen's inequality for convex functions in the general setting of the Lebesgue integral, with applications to means, Hölder's inequality and f-divergence measures in information theory, are given.

2010 *Mathematics subject classification*: primary 26D15; secondary 26D20, 94A05.

Keywords and phrases: Jensen's inequality, Hölder's inequality, measurable functions, Lebesgue integral, divergence measures, f-divergence measures.

1. Introduction

Let $(\Omega, \mathcal{A}, \mu)$ be a measurable space consisting of a set Ω, a σ-algebra $\mathcal{A}$ of parts of Ω and a countably additive and positive measure μ on $\mathcal{A}$ with values in $\mathbb{R} \cup \{\infty\}$. For a μ-measurable function $w : \Omega \to \mathbb{R}$, with $w(x) \geq 0$ for μ-a.e. (almost every) $x \in \Omega$, consider the Lebesgue space

$$L_w(\Omega, \mu) := \left\{ f : \Omega \to \mathbb{R} \mid f \text{ is } \mu\text{-measurable and } \int_\Omega w(x)|f(x)|\, d\mu(x) < \infty \right\}.$$

For simplicity of notation, we write everywhere in the following $\int_\Omega w\, d\mu$ instead of $\int_\Omega w(x)\, d\mu(x)$.

If $f, g : \Omega \to \mathbb{R}$ are μ-measurable functions, $\int_\Omega w\, d\mu = 1$ and $f, g, fg \in L_w(\Omega, \mu)$, then we may consider the *Čebyšev functional*

$$T_w(f, g) := \int_\Omega wfg\, d\mu - \int_\Omega wf\, d\mu \int_\Omega wg\, d\mu.$$

The following result is known in the literature as the *Grüss inequality*:

$$|T_w(f, g)| \leq \tfrac{1}{4}(\Gamma - \gamma)(\Delta - \delta),$$

provided

$$-\infty < \gamma \le f(x) \le \Gamma < \infty, \quad -\infty < \delta \le g(x) \le \Delta < \infty$$

for μ-a.e. $x \in \Omega$.

The constant $\frac{1}{4}$ is sharp in the sense that it cannot be replaced by a smaller constant.

If we assume that $-\infty < \gamma \le f(x) \le \Gamma < \infty$ for μ-a.e. $x \in \Omega$, then, by the Grüss inequality for $g = f$ and by Schwarz's integral inequality,

$$\int_\Omega w\left|f - \int_\Omega wf\,d\mu\right| d\mu \le \left(\int_\Omega wf^2\,d\mu - \left(\int_\Omega wf\,d\mu\right)^2\right)^{1/2} \le \frac{1}{2}(\Gamma - \gamma). \tag{1.1}$$

To provide a reverse of the celebrated Jensen's integral inequality for convex functions, in 2002, the author [12] obtained the following result.

Theorem 1.1. *Let $\Phi : [m, M] \subset \mathbb{R} \to \mathbb{R}$ be a differentiable convex function on (m, M) and $f : \Omega \to [m, M]$ such that $\Phi \circ f, f, \Phi' \circ f, (\Phi' \circ f) \cdot f \in L_w(\Omega, \mu)$, where $w \ge 0$ μ-a.e. on Ω with $\int_\Omega w\,d\mu = 1$. Then we have the inequality*

$$\begin{aligned} 0 &\le \int_\Omega w(\Phi \circ f)\,d\mu - \Phi\left(\int_\Omega wf\,d\mu\right) \\ &\le \int_\Omega w(\Phi' \circ f)f\,d\mu - \int_\Omega w(\Phi' \circ f)\,d\mu \int_\Omega wf\,d\mu \\ &\le \frac{1}{2}(\Phi'(M) - \Phi'(m)) \int_\Omega w\left|f - \int_\Omega wf\,d\mu\right| d\mu. \end{aligned} \tag{1.2}$$

For a generalisation of the first inequality in (1.2) without the differentiability assumption and the derivative Φ' replaced with a selection φ from the subdifferential $\partial\Phi$, see Niculescu [27].

If $\mu(\Omega) < \infty$ and $\Phi \circ f, f, \Phi' \circ f, (\Phi' \circ f) \cdot f \in L(\Omega, \mu)$, then we have the inequality

$$\begin{aligned} 0 &\le \frac{1}{\mu(\Omega)} \int_\Omega (\Phi \circ f)\,d\mu - \Phi\left(\frac{1}{\mu(\Omega)} \int_\Omega f\,d\mu\right) \\ &\le \frac{1}{\mu(\Omega)} \int_\Omega (\Phi' \circ f)f\,d\mu - \frac{1}{\mu(\Omega)} \int_\Omega (\Phi' \circ f)\,d\mu \cdot \frac{1}{\mu(\Omega)} \int_\Omega f\,d\mu \\ &\le \frac{1}{2}(\Phi'(M) - \Phi'(m)) \frac{1}{\mu(\Omega)} \int_\Omega \left|f - \frac{1}{\mu(\Omega)} \int_\Omega f\,d\mu\right| d\mu. \end{aligned}$$

The following discrete inequality is of interest as well.

Corollary 1.2. *Let $\Phi : [m, M] \to \mathbb{R}$ be a differentiable convex function on (m, M). If $x_i \in [m, M]$ and $w_i \ge 0$ $(i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^n w_i = 1$, then we have the*

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

counterpart of Jensen's weighted discrete inequality:

$$\begin{aligned}0 &\le \sum_{i=1}^{n} w_i\Phi(x_i) - \Phi\Big(\sum_{i=1}^{n} w_i x_i\Big)\\ &\le \sum_{i=1}^{n} w_i\Phi'(x_i)x_i - \sum_{i=1}^{n} w_i\Phi'(x_i)\sum_{i=1}^{n} w_i x_i \\ &\le \frac{1}{2}(\Phi'(M)-\Phi'(m))\sum_{i=1}^{n} w_i\Big|x_i - \sum_{j=1}^{n} w_j x_j\Big|.\end{aligned} \tag{1.3}$$

REMARK 1.3. The inequality between the first and the second terms in (1.3) was proved in 1994 by Dragomir and Ionescu [15].

Using the results (1.2) and (1.1), we can state the following string of reverse inequalities:

$$\begin{aligned}0 &\le \int_\Omega w(\Phi\circ f)\,d\mu - \Phi\Big(\int_\Omega wf\,d\mu\Big)\\ &\le \int_\Omega w(\Phi'\circ f)f\,d\mu - \int_\Omega w(\Phi'\circ f)\,d\mu\int_\Omega wf\,d\mu\\ &\le \frac{1}{2}(\Phi'(M)-\Phi'(m))\int_\Omega w\Big|f - \int_\Omega wf\,d\mu\Big|\,d\mu\\ &\le \frac{1}{2}(\Phi'(M)-\Phi'(m))\Big(\int_\Omega wf^2\,d\mu - \Big(\int_\Omega wf\,d\mu\Big)^2\Big)^{1/2}\\ &\le \frac{1}{4}(\Phi'(M)-\Phi'(m))(M-m),\end{aligned} \tag{1.4}$$

provided that $\Phi : [m, M] \subset \mathbb{R} \to \mathbb{R}$ is a differentiable convex function on (m, M) and $f : \Omega \to [m, M]$ such that $\Phi\circ f, f, \Phi'\circ f, (\Phi'\circ f)\cdot f \in L_w(\Omega, \mu)$, where $w \ge 0$ μ-a.e. on Ω with $\int_\Omega w\,d\mu = 1$.

REMARK 1.4. The inequality between the first, second and last terms from (1.4) was proved in the general case of positive linear functionals in 2001 by the author [11].

Motivated by the above results, we establish in the current paper two new reverses of Jensen's integral inequality for a convex function. Some natural applications for inequalities between means, reverses of Hölder's inequality and for the f-divergence measure that play an important role in information theory are given as well.

2. Reverse inequalities

The following reverse of Jensen's inequality holds.

THEOREM 2.1. *Let $\Phi : I \to \mathbb{R}$ be a continuous convex function on the interval of real numbers I and let $m, M \in \mathbb{R}$, $m < M$, with $[m, M] \subset \mathring{I}$ (where $\mathring{I}$ is the interior of I).*

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

If $f : \Omega \to \mathbb{R}$ is μ-measurable, satisfies the bounds

$$-\infty < m \le f(x) \le M < \infty \quad \text{for } \mu\text{-a.e. } x \in \Omega$$

and is such that $f, \Phi \circ f \in L_w(\Omega, \mu)$, where $w \ge 0$ μ-a.e. on Ω with $\int_\Omega w \, d\mu = 1$, then

$$\begin{aligned} 0 &\le \int_\Omega w(\Phi \circ f) \, d\mu - \Phi(\bar{f}_{\Omega,w}) \\ &\le \frac{(M - \bar{f}_{\Omega,w})(\bar{f}_{\Omega,w} - m)}{M - m} \sup_{t \in (m,M)} \Psi_\Phi(t; m, M) \\ &\le (M - \bar{f}_{\Omega,w})(\bar{f}_{\Omega,w} - m) \frac{\Phi'_-(M) - \Phi'_+(m)}{M - m} \\ &\le \frac{1}{4}(M - m)(\Phi'_-(M) - \Phi'_+(m)), \end{aligned} \tag{2.1}$$

where $\bar{f}_{\Omega,w} := \int_\Omega w(x)f(x) \, d\mu(x) \in [m, M]$ and $\Psi_\Phi(\cdot; m, M) : (m, M) \to \mathbb{R}$ is defined by

$$\Psi_\Phi(t; m, M) = \frac{\Phi(M) - \Phi(t)}{M - t} - \frac{\Phi(t) - \Phi(m)}{t - m}.$$

We also have the inequality

$$\begin{aligned} 0 &\le \int_\Omega w(\Phi \circ f) \, d\mu - \Phi(\bar{f}_{\Omega,w}) \le \frac{1}{4}(M - m)\Psi_\Phi(\bar{f}_{\Omega,w}; m, M) \\ &\le \frac{1}{4}(M - m)(\Phi'_-(M) - \Phi'_+(m)), \end{aligned} \tag{2.2}$$

provided that $\bar{f}_{\Omega,w} \in (m, M)$.

Proof. By the convexity of Φ,

$$\begin{aligned} &\int_\Omega w(x)\Phi(f(x)) \, d\mu(x) - \Phi(\bar{f}_{\Omega,w}) \\ &\quad = \int_\Omega w(x)\Phi\Big(\frac{m(M - f(x)) + M(f(x) - m)}{M - m}\Big) \, d\mu(x) \\ &\qquad - \Phi\Big(\int_\Omega w(x)\Big(\frac{m(M - f(x)) + M(f(x) - m)}{M - m}\Big) \, d\mu(x)\Big) \\ &\quad \le \int_\Omega \frac{(M - f(x))\Phi(m) + (f(x) - m)\Phi(M)}{M - m} w(x) \, d\mu(x) \\ &\qquad - \Phi\Big(\frac{m(M - \bar{f}_{\Omega,w}) + M(\bar{f}_{\Omega,w} - m)}{M - m}\Big) \\ &\quad = \frac{(M - \bar{f}_{\Omega,w})\Phi(m) + (\bar{f}_{\Omega,w} - m)\Phi(M)}{M - m} \\ &\qquad - \Phi\Big(\frac{m(M - \bar{f}_{\Omega,w}) + M(\bar{f}_{\Omega,w} - m)}{M - m}\Big) := B. \end{aligned} \tag{2.3}$$

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

By denoting

$$\Delta_\Phi(t; m, M) := \frac{(t-m)\Phi(M) + (M-t)\Phi(m)}{M-m} - \Phi(t), \quad t \in [m, M],$$

we have

$$\begin{aligned}\Delta_\Phi(t; m, M) &= \frac{(t-m)\Phi(M) + (M-t)\Phi(m) - (M-m)\Phi(t)}{M-m} \\ &= \frac{(t-m)\Phi(M) + (M-t)\Phi(m) - (M-t+t-m)\Phi(t)}{M-m} \\ &= \frac{(t-m)(\Phi(M)-\Phi(t)) - (M-t)(\Phi(t)-\Phi(m))}{M-m} \\ &= \frac{(M-t)(t-m)}{M-m}\Psi_\Phi(t; m, M)\end{aligned}$$

for any $t \in (m, M)$.

Therefore we have the equality

$$B = \frac{(M-\bar{f}_{\Omega,w})(\bar{f}_{\Omega,w}-m)}{M-m}\Psi_\Phi(\bar{f}_{\Omega,w}; m, M), \tag{2.4}$$

provided that $\bar{f}_{\Omega,w} \in (m, M)$.

For $\bar{f}_{\Omega,w} = m$ or $\bar{f}_{\Omega,w} = M$ the inequality (2.1) is obvious. If $\bar{f}_{\Omega,w} \in (m, M)$, then

$$\begin{aligned}\Psi_\Phi(\bar{f}_{\Omega,w}; m, M) &\le \sup_{t\in(m,M)} \Psi_\Phi(t; m, M) \\ &= \sup_{t\in(m,M)}\left(\frac{\Phi(M)-\Phi(t)}{M-t} - \frac{\Phi(t)-\Phi(m)}{t-m}\right) \\ &\le \sup_{t\in(m,M)}\left(\frac{\Phi(M)-\Phi(t)}{M-t}\right) + \sup_{t\in(m,M)}\left(-\frac{\Phi(t)-\Phi(m)}{t-m}\right) \\ &= \sup_{t\in(m,M)}\left(\frac{\Phi(M)-\Phi(t)}{M-t}\right) - \inf_{t\in(m,M)}\left(\frac{\Phi(t)-\Phi(m)}{t-m}\right) \\ &= \Phi'_-(M) - \Phi'_+(m),\end{aligned}$$

which by (2.3) and (2.4) produces the desired result (2.1).

Since, obviously,

$$\frac{(M-\bar{f}_{\Omega,w})(\bar{f}_{\Omega,w}-m)}{M-m} \le \frac{1}{4}(M-m),$$

then by (2.3) and (2.4) we deduce the first inequality (2.2). The second part is clear. □

Corollary 2.2. *Let $\Phi : I \to \mathbb{R}$ be a continuous convex function on the interval of real numbers I and $m, M \in \mathbb{R}$, $m < M$, with $[m, M] \subset \mathring{I}$. If $x_i \in [m, M]$ and $p_i \ge 0$*

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

for $i \in \{1, \ldots, n\}$ *with* $\sum_{i=1}^{n} p_i = 1$, *then we have the inequalities*

$$\begin{aligned}0 &\le \sum_{i=1}^{n} p_i \Phi(x_i) - \Phi(\bar{x}_p)\\ &\le \frac{(M-\bar{x}_p)(\bar{x}_p - m)}{M-m} \sup_{t\in(m,M)} \Psi_\Phi(t; m, M)\\ &\le (M-\bar{x}_p)(\bar{x}_p - m)\frac{\Phi'_-(M) - \Phi'_+(m)}{M-m}\\ &\le \frac{1}{4}(M-m)(\Phi'_-(M) - \Phi'_+(m)),\end{aligned} \tag{2.5}$$

and

$$\begin{aligned}0 &\le \sum_{i=1}^{n} p_i \Phi(x_i) - \Phi(\bar{x}_p) \le \frac{1}{4}(M-m)\Psi_\Phi(\bar{x}_p; m, M)\\ &\le \frac{1}{4}(M-m)(\Phi'_-(M) - \Phi'_+(m)),\end{aligned} \tag{2.6}$$

where $\bar{x}_p := \sum_{i=1}^{n} p_i x_i \in (m, M)$.

REMARK 2.3. Define the weighted arithmetic mean of the positive n-tuple $x = (x_1, \ldots, x_n)$ with the nonnegative weights $w = (w_1, \ldots, w_n)$ by

$$A_n(w, x) := \frac{1}{W_n} \sum_{i=1}^{n} w_i x_i,$$

where $W_n := \sum_{i=1}^{n} w_i > 0$, and the weighted geometric mean of the same n-tuple by

$$G_n(w, x) := \left(\prod_{i=1}^{n} x_i^{w_i}\right)^{1/W_n}.$$

It is well known that the following arithmetic mean–geometric mean inequality holds true:

$$A_n(w, x) \ge G_n(w, x).$$

Applying the inequality between the first and third terms in (2.5) for the convex function $\Phi(t) = -\log t$, $t > 0$,

$$\begin{aligned}1 \le \frac{A_n(w, x)}{G_n(w, x)} &\le \exp\left(\frac{1}{Mm}(M - A_n(w, x))(A_n(w, x) - m)\right)\\ &\le \exp\left(\frac{1}{4}\frac{(M-m)^2}{mM}\right),\end{aligned}$$

provided that $0 < m \le x_i \le M < \infty$ for $i \in \{1, \ldots, n\}$.

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

Also, if we apply the inequality (2.6) for the same function Φ we obtain

$$\begin{aligned}
1 &\le \frac{A_n(w,x)}{G_n(w,x)} \\
&\le \bigg(\Big(\frac{M}{A_n(w,x)}\Big)^{M-A_n(w,x)}\Big(\frac{m}{A_n(w,x)}\Big)^{A_n(w,x)-m}\bigg)^{-(M-m)/4} \\
&\le \exp\Big(\frac{1}{4}\frac{(M-m)^2}{mM}\Big).
\end{aligned}$$

The following result also holds.

Theorem 2.4. *With the assumptions of Theorem 2.1, we have the inequalities*

$$\begin{aligned}
0 &\le \int_\Omega w(\Phi\circ f)\,d\mu(x) - \Phi(\bar{f}_{\Omega,w}) \\
&\le 2\max\Big\{\frac{M-\bar{f}_{\Omega,w}}{M-m}, \frac{\bar{f}_{\Omega,w}-m}{M-m}\Big\}\Big(\frac{\Phi(m)+\Phi(M)}{2} - \Phi\Big(\frac{m+M}{2}\Big)\Big) \\
&\le \frac{1}{2}\max\{M-\bar{f}_{\Omega,w}, \bar{f}_{\Omega,w}-m\}(\Phi'_-(M)-\Phi'_+(m)).
\end{aligned} \tag{2.7}$$

Proof. We first recall the following result obtained by the author in [14] that provides a refinement and a reverse for the weighted Jensen's discrete inequality:

$$\begin{aligned}
n\min_{i\in\{1,\ldots,n\}}\{p_i\}&\Big(\frac{1}{n}\sum_{i=1}^n \Phi(x_i) - \Phi\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)\Big) \\
&\le \frac{1}{P_n}\sum_{i=1}^n p_i\Phi(x_i) - \Phi\Big(\frac{1}{P_n}\sum_{i=1}^n p_i x_i\Big) \\
&\le n\max_{i\in\{1,\ldots,n\}}\{p_i\}\Big(\frac{1}{n}\sum_{i=1}^n \Phi(x_i) - \Phi\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)\Big),
\end{aligned} \tag{2.8}$$

where $\Phi : C \to \mathbb{R}$ is a convex function defined on the convex subset C of the linear space X, $\{x_i\}_{i\in\{1,\ldots,n\}} \subset C$ are vectors and $\{p_i\}_{i\in\{1,\ldots,n\}}$ are nonnegative numbers with $P_n := \sum_{i=1}^n p_i > 0$.

For $n = 2$ we deduce from (2.8) that

$$\begin{aligned}
2\min\{t, 1-t\}&\Big(\frac{\Phi(x)+\Phi(y)}{2} - \Phi\Big(\frac{x+y}{2}\Big)\Big) \\
&\le t\Phi(x) + (1-t)\Phi(y) - \Phi(tx+(1-t)y) \\
&\le 2\max\{t, 1-t\}\Big(\frac{\Phi(x)+\Phi(y)}{2} - \Phi\Big(\frac{x+y}{2}\Big)\Big)
\end{aligned} \tag{2.9}$$

for any $x, y \in C$ and $t \in [0, 1]$.

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

If we use the second inequality in (2.9) for the convex function $\Phi : I \to \mathbb{R}$ and $m, M \in \mathbb{R}$, $m < M$, with $[m, M] \subset \mathring{I}$, we have for $t = (M - \bar{f}_{\Omega,w})/(M - m)$ that

$$\frac{(M - \bar{f}_{\Omega,w})\Phi(m) + (\bar{f}_{\Omega,w} - m)\Phi(M)}{M - m} - \Phi\Big(\frac{m(M - \bar{f}_{\Omega,w}) + M(\bar{f}_{\Omega,w} - m)}{M - m}\Big) \leq 2\max\Big\{\frac{M - \bar{f}_{\Omega,w}}{M - m}, \frac{\bar{f}_{\Omega,w} - m}{M - m}\Big\}\Big(\frac{\Phi(m) + \Phi(M)}{2} - \Phi\Big(\frac{m + M}{2}\Big)\Big). \tag{2.10}$$

Using (2.3) and (2.10) we deduce the first inequality in (2.7).

Since

$$\frac{\frac{\Phi(m)+\Phi(M)}{2} - \Phi(\frac{m+M}{2})}{M - m} = \frac{1}{4}\Big(\frac{\Phi(M) - \Phi(\frac{m+M}{2})}{M - \frac{m+M}{2}} - \frac{\Phi(\frac{m+M}{2}) - \Phi(m)}{\frac{m+M}{2} - m}\Big)$$

and, by the gradient inequality,

$$\frac{\Phi(M) - \Phi(\frac{m+M}{2})}{M - \frac{m+M}{2}} \leq \Phi'_-(M)$$

and

$$\frac{\Phi(\frac{m+M}{2}) - \Phi(m)}{\frac{m+M}{2} - m} \geq \Phi'_+(m),$$

then

$$\frac{\frac{\Phi(m)+\Phi(M)}{2} - \Phi(\frac{m+M}{2})}{M - m} \leq \frac{1}{4}(\Phi'_-(M) - \Phi'_+(m)). \tag{2.11}$$

Making use of (2.10) and (2.11), we deduce the last part of (2.7). □

COROLLARY 2.5. *With the assumptions in Corollary 2.2, we have the inequalities*

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} p_i\Phi(x_i) - \Phi(\bar{x}_p) \\ &\leq 2\max\Big\{\frac{M - \bar{x}_p}{M - m}, \frac{\bar{x}_p - m}{M - m}\Big\}\Big(\frac{\Phi(m) + \Phi(M)}{2} - \Phi\Big(\frac{m + M}{2}\Big)\Big) \\ &\leq \frac{1}{2}\max\{M - \bar{x}_p, \bar{x}_p - m\}(\Phi'_-(M) - \Phi'_+(m)). \end{aligned}$$

REMARK 2.6. Since, obviously,

$$\frac{M - \bar{f}_{\Omega,w}}{M - m}, \frac{\bar{f}_{\Omega,w} - m}{M - m} \leq 1,$$

we obtain from the first inequality in (2.7) the simpler but coarser inequality

$$0 \leq \int_\Omega w(\Phi \circ f)\, d\mu(x) - \Phi(\bar{f}_{\Omega,w}) \leq 2\Big(\frac{\Phi(m) + \Phi(M)}{2} - \Phi\Big(\frac{m + M}{2}\Big)\Big).$$

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

The discrete version of this result, namely

$$0 \le \sum_{i=1}^{n} p_i \Phi(x_i) - \Phi(\bar{x}_p) \le 2\Big(\frac{\Phi(m)+\Phi(M)}{2} - \Phi\Big(\frac{m+M}{2}\Big)\Big),$$

was obtained in 2008 by Simic [34].

Remark 2.7. With the assumptions in Remark 2.3 we have the following reverse of the arithmetic mean–geometric mean inequality

$$1 \le \frac{A_n(w,x)}{G_n(w,x)} \le \Big(\frac{A(m,M)}{G(m,M)}\Big)^{2\max\{(M-A_n(w,x))/(M-m),(A_n(w,x)-m)/(M-m)\}}, \tag{2.12}$$

where $A(m, M)$ is the arithmetic mean and $G(m, M)$ is the geometric mean of the positive numbers m and M.

3. Applications for the Hölder inequality

It is well known that if $f \in L_p(\Omega, \mu)$, $p > 1$, where the Lebesgue space $L_p(\Omega, \mu)$ is defined by

$$L_p(\Omega,\mu) := \Big\{f : \Omega \to \mathbb{R} \mid f \text{ is } \mu\text{-measurable and } \int_\Omega |f(x)|^p \, d\mu(x) < \infty\Big\},$$

and $g \in L_q(\Omega, \mu)$ with $1/p + 1/q = 1$ then $fg \in L(\Omega, \mu) = L_1(\Omega, \mu)$ and the *Hölder inequality* holds true:

$$\int_\Omega |fg| \, d\mu \le \Big(\int_\Omega |f|^p \, d\mu\Big)^{1/p} \Big(\int_\Omega |g|^p \, d\mu\Big)^{1/q}.$$

Assume that $p > 1$. If $h : \Omega \to \mathbb{R}$ is μ-measurable, satisfies the bounds

$$0 < m \le |h(x)| \le M < \infty \quad \text{for } \mu\text{-a.e. } x \in \Omega$$

and is such that $h, |h|^p \in L_w(\Omega, \mu)$, for a μ-measurable function $w : \Omega \to \mathbb{R}$, with $w(x) \ge 0$ for μ-a.e. $x \in \Omega$ and $\int_\Omega w \, d\mu > 0$, then, from (2.1),

$$\begin{aligned} 0 &\le \frac{\int_\Omega |h|^p w \, d\mu}{\int_\Omega w \, d\mu} - \Big(\frac{\int_\Omega |h| w \, d\mu}{\int_\Omega w \, d\mu}\Big)^p \\ &\le \frac{(M - \overline{|h|}_{\Omega,w})(\overline{|h|}_{\Omega,w} - m)}{M-m} B_p(m, M) \\ &\le p \frac{M^{p-1} - m^{p-1}}{M-m} (M - \overline{|h|}_{\Omega,w})(\overline{|h|}_{\Omega,w} - m) \\ &\le \frac{1}{4} p(M-m)(M^{p-1} - m^{p-1}), \end{aligned} \tag{3.1}$$

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

where $\overline{|h|}_{\Omega,w} := \int_\Omega |h| w \, d\mu / \int_\Omega w \, d\mu \in [m, M]$, $\Psi_p(\cdot; m, M) : (m, M) \to \mathbb{R}$ is defined by

$$\Psi_p(t; m, M) = \frac{M^p - t^p}{M - t} - \frac{t^p - m^p}{t - m},$$

and

$$B_p(m, M) := \sup_{t \in (m,M)} \Psi_p(t; m, M).$$

From (2.2) we also have the inequality

$$\begin{aligned} 0 &\le \frac{\int_\Omega |h|^p w \, d\mu}{\int_\Omega w \, d\mu} - \left(\frac{\int_\Omega |h| w \, d\mu}{\int_\Omega w \, d\mu} \right)^p \le \frac{1}{4}(M - m)\Psi_p(\overline{|h|}_{\Omega,w}; m, M) \\ &\le \frac{1}{4} p(M - m)(M^{p-1} - m^{p-1}). \end{aligned} \tag{3.2}$$

Proposition 3.1. *If $f \in L_p(\Omega, \mu)$, $g \in L_q(\Omega, \mu)$ with $p > 1$, $1/p + 1/q = 1$, and there exist constants $\gamma, \Gamma > 0$ such that*

$$\gamma \le \frac{|f|}{|g|^{q-1}} \le \Gamma \ \mu\text{-a.e on } \Omega,$$

then

$$\begin{aligned} 0 &\le \frac{\int_\Omega |f|^p \, d\mu}{\int_\Omega |g|^q \, d\mu} - \left(\frac{\int_\Omega |fg| \, d\mu}{\int_\Omega |g|^q \, d\mu} \right)^p \\ &\le \frac{B_p(\gamma, \Gamma)}{\Gamma - \gamma} \left(\Gamma - \frac{\int_\Omega |fg| \, d\mu}{\int_\Omega |g|^q \, d\mu} \right) \left(\frac{\int_\Omega |fg| \, d\mu}{\int_\Omega |g|^q \, d\mu} - \gamma \right) \\ &\le p \frac{\Gamma^{p-1} - \gamma^{p-1}}{\Gamma - \gamma} \left(\Gamma - \frac{\int_\Omega |fg| \, d\mu}{\int_\Omega |g|^q \, d\mu} \right) \left(\frac{\int_\Omega |fg| \, d\mu}{\int_\Omega |g|^q \, d\mu} - \gamma \right) \\ &\le \frac{1}{4} p(\Gamma - \gamma)(\Gamma^{p-1} - \gamma^{p-1}), \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} 0 &\le \frac{\int_\Omega |f|^p \, d\mu}{\int_\Omega |g|^q \, d\mu} - \left(\frac{\int_\Omega |fg| \, d\mu}{\int_\Omega |g|^q \, d\mu} \right)^p \\ &\le \frac{1}{4}(\Gamma - \gamma)\Psi_p\left(\frac{\int_\Omega |fg| \, d\mu}{\int_\Omega |g|^q \, d\mu}; \gamma, \Gamma \right) \le \frac{1}{4} p(\Gamma - \gamma)(\Gamma^{p-1} - \gamma^{p-1}), \end{aligned} \tag{3.4}$$

where $B_p(\cdot, \cdot)$ and $\Psi_p(\cdot; \cdot, \cdot)$ are defined above.

Proof. The inequalities (3.3) and (3.4) follow from (3.1) and (3.2) by choosing

$$h = \frac{|f|}{|g|^{q-1}} \quad \text{and} \quad w = |g|^q.$$

The details are omitted. □

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

Remark 3.2. We observe that for $p = q = 2$ we have $\Psi_2(t; \gamma, \Gamma) = \Gamma - \gamma = B_2(\gamma, \Gamma)$ and then from the first inequality in (3.3) we get the following reverse of the Cauchy–Bunyakovsky–Schwarz inequality:

$$\begin{aligned}\int_\Omega |g|^2\,d\mu \int_\Omega |f|^2\,d\mu - \Big(\int_\Omega |fg|\,d\mu\Big)^2 \\ \le \Big(\Gamma - \frac{\int_\Omega |fg|\,d\mu}{\int_\Omega |g|^2\,d\mu}\Big)\Big(\frac{\int_\Omega |fg|\,d\mu}{\int_\Omega |g|^2\,d\mu} - \gamma\Big)\Big(\int_\Omega |g|^2\,d\mu\Big)^2,\end{aligned}$$

provided that $f, g \in L_2(\Omega, \mu)$, and there exist constants $\gamma, \Gamma > 0$ such that

$$\gamma \le \frac{|f|}{|g|} \le \Gamma\ \mu\text{-a.e on } \Omega.$$

Corollary 3.3. *With the assumptions of Proposition 3.1 we have the following additive reverses of the Hölder inequality:*

$$\begin{aligned}0 &\le \Big(\int_\Omega |f|^p\,d\mu\Big)^{1/p}\Big(\int_\Omega |g|^q\,d\mu\Big)^{1/q} - \int_\Omega |fg|\,d\mu \\ &\le \Big(\frac{B_p(\gamma,\Gamma)}{\Gamma-\gamma}\Big)^{1/p}\Big(\Gamma - \frac{\int_\Omega |fg|\,d\mu}{\int_\Omega |g|^q\,d\mu}\Big)^{1/p}\Big(\frac{\int_\Omega |fg|\,d\mu}{\int_\Omega |g|^q\,d\mu} - \gamma\Big)^{1/p}\int_\Omega |g|^q\,d\mu \\ &\le p^{1/p}\Big(\frac{\Gamma^{p-1}-\gamma^{p-1}}{\Gamma-\gamma}\Big)^{1/p}\Big(\Gamma - \frac{\int_\Omega |fg|\,d\mu}{\int_\Omega |g|^q\,d\mu}\Big)^{1/p}\Big(\frac{\int_\Omega |fg|\,d\mu}{\int_\Omega |g|^q\,d\mu} - \gamma\Big)^{1/p} \\ &\quad\times \int_\Omega |g|^q\,d\mu \\ &\le \frac{1}{4^{1/p}}p^{1/p}(\Gamma-\gamma)^{1/p}(\Gamma^{p-1}-\gamma^{p-1})^{1/p}\int_\Omega |g|^q\,d\mu\end{aligned} \tag{3.5}$$

and

$$\begin{aligned}0 &\le \Big(\int_\Omega |f|^p\,d\mu\Big)^{1/p}\Big(\int_\Omega |g|^q\,d\mu\Big)^{1/q} - \int_\Omega |fg|\,d\mu \\ &\le \frac{1}{4^{1/p}}(\Gamma-\gamma)^{1/p}\Psi_p^{1/p}\Big(\frac{\int_\Omega |fg|\,d\mu}{\int_\Omega |g|^q\,d\mu}; m, M\Big)\int_\Omega |g|^q\,d\mu \\ &\le \frac{1}{4^{1/p}}p^{1/p}(\Gamma-\gamma)^{1/p}(\Gamma^{p-1}-\gamma^{p-1})^{1/p}\int_\Omega |g|^q\,d\mu,\end{aligned} \tag{3.6}$$

where $p > 1$ and $1/p + 1/q = 1$.

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

PROOF. By multiplying in (3.3) with $(\int_\Omega |g|^q\, d\mu)^p$,

$$\begin{aligned}
\int_\Omega |f|^p\, d\mu\Big(\int_\Omega |g|^q\, d\mu\Big)^{p-1} - \Big(\int_\Omega |fg|\, d\mu\Big)^p
&\le \frac{B_p(\gamma,\Gamma)}{\Gamma-\gamma}\Big(\Gamma - \frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu}\Big)\Big(\frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu} - \gamma\Big)\Big(\int_\Omega |g|^q\, d\mu\Big)^p \\
&\le p\frac{\Gamma^{p-1}-\gamma^{p-1}}{\Gamma-\gamma}\Big(\Gamma - \frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu}\Big)\Big(\frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu} - \gamma\Big)\Big(\int_\Omega |g|^q\, d\mu\Big)^p \\
&\le \frac{1}{4}p(\Gamma-\gamma)(\Gamma^{p-1}-\gamma^{p-1})\Big(\int_\Omega |g|^q\, d\mu\Big)^p,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
\int_\Omega |f|^p\, d\mu\Big(\int_\Omega |g|^q\, d\mu\Big)^{p-1}
&\le \Big(\int_\Omega |fg|\, d\mu\Big)^p + \frac{B_p(\gamma,\Gamma)}{\Gamma-\gamma}\Big(\Gamma - \frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu}\Big)\Big(\frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu} - \gamma\Big) \\
&\quad \times \Big(\int_\Omega |g|^q\, d\mu\Big)^p \\
&\le \Big(\int_\Omega |fg|\, d\mu\Big)^p + p\Big(\Gamma - \frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu}\Big)\Big(\frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu} - \gamma\Big) \\
&\quad \times \Big(\int_\Omega |g|^q\, d\mu\Big)^p \frac{\Gamma^{p-1}-\gamma^{p-1}}{\Gamma-\gamma} \\
&\le \Big(\int_\Omega |fg|\, d\mu\Big)^p + \frac{1}{4}p(\Gamma-\gamma)(\Gamma^{p-1}-\gamma^{p-1})\Big(\int_\Omega |g|^q\, d\mu\Big)^p.
\end{aligned} \tag{3.7}$$

Raising to the power $1/p$ with $p > 1$ and employing the elementary inequality that for $p > 1$ and $\alpha, \beta > 0$,

$$(\alpha+\beta)^{1/p} \le \alpha^{1/p} + \beta^{1/p},$$

we have from the first part of (3.7) that

$$\begin{aligned}
\Big(\int_\Omega |f|^p\, d\mu\Big)^{1/p}\Big(\int_\Omega |g|^q\, d\mu\Big)^{1-1/p}
&\le \int_\Omega |fg|\, d\mu + \Big(\frac{B_p(\gamma,\Gamma)}{\Gamma-\gamma}\Big)^{1/p}\Big(\Gamma - \frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu}\Big)^{1/p}\Big(\frac{\int_\Omega |fg|\, d\mu}{\int_\Omega |g|^q\, d\mu} - \gamma\Big)^{1/p} \\
&\quad \times \int_\Omega |g|^q\, d\mu,
\end{aligned} \tag{3.8}$$

and since $1 - 1/p = 1/q$ we get from (3.8) the first inequality in (3.5). The rest is obvious.

The inequality (3.6) can be proved in a similar manner; the details are omitted. □

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

If $h:\Omega\to\mathbb{R}$ is μ-measurable, satisfies the bounds

$$0<m\le|h(x)|\le M<\infty\quad\text{for }\mu\text{-a.e. } x\in\Omega$$

and is such that $h, |h|^p\in L_w(\Omega,\mu)$, for a μ-measurable function $w:\Omega\to\mathbb{R}$, with $w(x)\ge 0$ for μ-a.e. $x\in\Omega$ and $\int_\Omega w\,d\mu>0$, then from (2.7) we also have the inequality

$$\begin{aligned}0&\le\frac{\int_\Omega|h|^p w\,d\mu}{\int_\Omega w\,d\mu}-\Big(\frac{\int_\Omega|h|w\,d\mu}{\int_\Omega w\,d\mu}\Big)^p\\&\le 2\Big(\frac{m^p+M^p}{2}-\Big(\frac{m+M}{2}\Big)^p\Big)\max\Big\{\frac{M-\overline{|h|}_{\Omega,w}}{M-m},\frac{\overline{|h|}_{\Omega,w}-m}{M-m}\Big\}\\&\le\frac{1}{2}p(M^{p-1}-m^{p-1})\max\{M-\overline{|h|}_{\Omega,w},\overline{|h|}_{\Omega,w}-m\},\end{aligned}\tag{3.9}$$

where, as above, $\overline{|h|}_{\Omega,w}:=\int_\Omega|h|w\,d\mu/\int_\Omega w\,d\mu\in[m,M]$.

From (3.9) we can state the following result.

Proposition 3.4. *With the assumptions of Proposition 3.1 we have*

$$\begin{aligned}0&\le\frac{\int_\Omega|f|^p\,d\mu}{\int_\Omega|g|^q\,d\mu}-\Big(\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu}\Big)^p\\&\le 2\cdot\frac{\frac{\gamma^p+\Gamma^p}{2}-(\frac{\gamma+\Gamma}{2})^p}{\Gamma-\gamma}\max\Big\{\Gamma-\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu},\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu}-\gamma\Big\}\\&\le\frac{1}{2}p(\Gamma^{p-1}-\gamma^{p-1})\max\Big\{\Gamma-\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu},\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu}-\gamma\Big\}.\end{aligned}$$

Finally, the following additive reverse of the Hölder inequality can also be stated.

Corollary 3.5. *With the assumptions of Proposition 3.1,*

$$\begin{aligned}0&\le\Big(\int_\Omega|f|^p\,d\mu\Big)^{1/p}\Big(\int_\Omega|g|^q\,d\mu\Big)^{1/q}-\int_\Omega|fg|\,d\mu\\&\le 2^{1/p}\cdot\Big(\frac{\frac{\gamma^p+\Gamma^p}{2}-(\frac{\gamma+\Gamma}{2})^p}{\Gamma-\gamma}\Big)^{1/p}\\&\qquad\times\max\Big\{\Big(\Gamma-\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu}\Big)^{1/p},\Big(\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu}-\gamma\Big)^{1/p}\Big\}\int_\Omega|g|^q\,d\mu\\&\le\frac{1}{2^{1/p}}p^{1/p}\max\Big\{\Big(\Gamma-\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu}\Big)^{1/p},\Big(\frac{\int_\Omega|fg|\,d\mu}{\int_\Omega|g|^q\,d\mu}-\gamma\Big)^{1/p}\Big\}\\&\qquad\times(\Gamma^{p-1}-\gamma^{p-1})^{1/p}\int_\Omega|g|^q\,d\mu.\end{aligned}$$

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

REMARK 3.6. As a simpler but coarser inequality we have the following result:

$$0 \le \left(\int_\Omega |f|^p \, d\mu\right)^{1/p} \left(\int_\Omega |g|^q \, d\mu\right)^{1/q} - \int_\Omega |fg| \, d\mu$$
$$\le 2^{1/p} \cdot \left(\frac{\gamma^p + \Gamma^p}{2} - \left(\frac{\gamma + \Gamma}{2}\right)^p\right)^{1/p} \int_\Omega |g|^q \, d\mu,$$

where f and g are as above.

4. Applications for f-divergence

One of the important issues in many applications of probability theory is finding an appropriate measure of *distance* (or *difference* or *discrimination*) between two probability distributions. A number of divergence measures for this purpose have been proposed and extensively studied by Jeffreys [19], Kullback and Leibler [24], Rényi [30], Havrda and Charvat [17], Kapur [22], Sharma and Mittal [32], Burbea and Rao [4], Rao [29], Lin [25], Csiszár [7], Ali and Silvey [1], Vajda [39], Shioya and Da-Te [33] and others (see, for example, [26], and the references therein).

These measures have been applied in a variety of fields such as: anthropology [29], genetics [26], finance, economics and political science [31, 36, 37], biology [28], the analysis of contingency tables [16], approximation of probability distributions [6, 23], signal processing [20, 21] and pattern recognition [2, 5]. A number of these measures of distance are specific cases of Csiszár f-divergence and so further exploration of this concept will have a flow-on effect to other measures of distance and to areas in which they are applied.

Assume that a set Ω and the σ-finite measure μ are given. Consider the set of all probability densities on μ to be $\mathcal{P} := \{p \mid p : \Omega \to \mathbb{R},\ p(x) \ge 0,\ \int_\Omega p(x) \, d\mu(x) = 1\}$. The Kullback–Leibler divergence [24] is well known among the information divergences. It is defined as

$$D_{KL}(p, q) := \int_\Omega p(x) \log\left(\frac{p(x)}{q(x)}\right) d\mu(x), \qquad p, q \in \mathcal{P}, \tag{4.1}$$

where log is to base e.

In information theory and statistics, various divergences are applied in addition to the Kullback–Leibler divergence. These are, for example, the *variation distance* D_v, *Hellinger distance* D_H [18], χ^2*-divergence* D_{χ^2}, α*-divergence* D_α, *Bhattacharyya distance* D_B [3], *harmonic distance* D_{Ha}, *Jeffreys distance* D_J [19], *triangular discrimination* D_Δ [38]. They are defined as follows:

$$D_v(p, q) := \int_\Omega |p(x) - q(x)| \, d\mu(x), \qquad p, q \in \mathcal{P}; \tag{4.2}$$

$$D_H(p, q) := \int_\Omega |\sqrt{p(x)} - \sqrt{q(x)}| \, d\mu(x), \qquad p, q \in \mathcal{P}; \tag{4.3}$$

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

$$D_{\chi^2}(p,q) := \int_\Omega p(x)\Big(\Big(\frac{q(x)}{p(x)}\Big)^2 - 1\Big)\,d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.4}$$

$$D_\alpha(p,q) := \frac{4}{1-\alpha^2}\Big(1 - \int_\Omega (p(x))^{(1-\alpha)/2}(q(x))^{(1+\alpha)/2}\,d\mu(x)\Big), \quad p,q \in \mathcal{P}; \tag{4.5}$$

$$D_B(p,q) := \int_\Omega \sqrt{p(x)q(x)}\,d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.6}$$

$$D_{Ha}(p,q) := \int_\Omega \frac{2p(x)q(x)}{p(x)+q(x)}\,d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.7}$$

$$D_J(p,q) := \int_\Omega (p(x)-q(x))\log\Big(\frac{p(x)}{q(x)}\Big)\,d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.8}$$

$$D_\Delta(p,q) := \int_\Omega \frac{(p(x)-q(x))^2}{p(x)+q(x)}\,d\mu(x), \quad p,q \in \mathcal{P}. \tag{4.9}$$

For other divergence measures, see Kapur [22] or the book online by Taneja [35].

Csiszár f-divergence is defined as follows [8]:

$$I_f(p,q) := \int_\Omega p(x) f\Big(\frac{q(x)}{p(x)}\Big)\,d\mu(x), \quad p,q \in \mathcal{P},$$

where f is convex on $(0,\infty)$. It is assumed that f is strictly convex and satisfies the condition that $f(1) = 0$. By appropriately defining this convex function, various divergences are derived. Most of the above distances (4.1)–(4.9) are particular instances of Csiszár f-divergence. There are also many others which are not in this class (see, for example, [35]). For the basic properties of Csiszár f-divergence, see [8, 9] and [39].

The following result holds.

Proposition 4.1. *Suppose that $f : (0,\infty) \to \mathbb{R}$ be a convex function with the property that $f(1) = 0$. Assume that $p, q \in \mathcal{P}$ and there exist constants $0 < r < 1 < R < \infty$ such that*

$$r \le \frac{q(x)}{p(x)} \le R \quad \text{for } \mu\text{-a.e. } x \in \Omega.$$

Then we have the inequalities

$$\begin{aligned} I_f(p,q) &\le \frac{(R-1)(1-r)}{R-r} \sup_{t\in(r,R)} \Psi_f(t;r,R) \\ &\le (R-1)(1-r)\frac{f'_-(R) - f'_+(r)}{R-r} \\ &\le \frac{1}{4}(R-r)(f'_-(R) - f'_+(r)), \end{aligned} \tag{4.10}$$

where $\Psi_f(\cdot;r,R) : (r,R) \to \mathbb{R}$ is defined by

$$\Psi_f(t;r,R) = \frac{f(R)-f(t)}{R-t} - \frac{f(t)-f(r)}{t-r}.$$

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

We also have the inequality

$$I_f(p,q) \le \frac{1}{4}(R-r)\frac{f(R)(1-r)+f(r)(R-1)}{(R-1)(1-r)} \le \frac{1}{4}(R-r)(f'_-(R)-f'_+(r)). \tag{4.11}$$

The proof follows by Theorem 2.1 by choosing $w(x)=p(x)$, $f(x)=q(x)/p(x)$, $m=r$ and $M=R$ and performing the required calculations. The details are omitted.

Using the same approach and Theorem 2.4 we can also state the following result.

PROPOSITION 4.2. *With the assumptions of Proposition 4.1,*

$$I_f(p,q) \le 2\max\left\{\frac{R-1}{R-r},\frac{1-r}{R-r}\right\}\left(\frac{f(r)+f(R)}{2}-f\left(\frac{r+R}{2}\right)\right) \le \frac{1}{2}\max\{R-1,1-r\}(f'_-(R)-f'_+(r)). \tag{4.12}$$

The above results can be used to obtain various inequalities for divergence measures in information theory that are particular instances of f-divergence.

Consider the Kullback–Leibler divergence

$$D_{KL}(p,q) := \int_\Omega p(x)\log\left(\frac{p(x)}{q(x)}\right)d\mu(x), \quad p,q\in\mathcal{P},$$

which is an f-divergence for the convex function $f:(0,\infty)\to\mathbb{R}$, $f(t)=-\log t$.

If $p,q\in\mathcal{P}$ such that there exist constants $0<r<1<R<\infty$ with

$$r \le \frac{q(x)}{p(x)} \le R \quad \text{for } \mu\text{-a.e. } x\in\Omega,$$

then we get from (4.10) that

$$D_{KL}(p,q) \le \frac{(R-1)(1-r)}{rR},$$

from (4.11) that

$$D_{KL}(p,q) \le \tfrac{1}{4}(R-r)\log(R^{-1/(R-1)}r^{-1/(1-r)})$$

and from (4.12) that

$$D_{KL}(p,q) \le 2\max\left\{\frac{R-1}{R-r},\frac{1-r}{R-r}\right\}\log\left(\frac{A(r,R)}{G(r,R)}\right) \le \frac{1}{2}\max\{R-1,1-r\}\left(\frac{R-r}{rR}\right),$$

where $A(r,R)$ is the arithmetic mean and $G(r,R)$ is the geometric mean of the positive numbers r and R.

https://doi.org/10.1017/S0004972712001098 Published online by Cambridge University Press

Acknowledgement

The author would like to thank the anonymous referee for valuable remarks that have been implemented in the final version of the paper.

References

[1] S. M. Ali and S. D. Silvey, 'A general class of coefficients of divergence of one distribution from another', *J. R. Stat. Soc. Ser. B* **28** (1966), 131–142.
[2] M. Ben-Bassat, 'f-entropies, probability of error and feature selection', *Inform. Control* **39** (1978), 227–242.
[3] A. Bhattacharyya, 'On a measure of divergence between two statistical populations defined by their probability distributions', *Bull. Calcutta Math. Soc.* **35** (1943), 99–109.
[4] I. Burbea and C. R. Rao, 'On the convexity of some divergence measures based on entropy function', *IEEE Trans. Inform. Theory* **28**(3) (1982), 489–495.
[5] C. H. Chen, *Statistical Pattern Recognition* (Hoyderc Book Co., Rocelle Park, New York, 1973).
[6] C. K. Chow and C. N. Lin, 'Approximating discrete probability distributions with dependence trees', *IEEE Trans. Inform. Theory* **14**(3) (1968), 462–467.
[7] I. Csiszár, 'Information-type measures of difference of probability distributions and indirect observations', *Studia Math. Hungarica* **2** (1967), 299–318.
[8] I. Csiszár, 'On topological properties of f-divergences', *Studia Math. Hungarica* **2** (1967), 329–339.
[9] I. Csiszár and J. Körner, *Information Theory: Coding Theorem for Discrete Memoryless Systems* (Academic Press, New York, 1981).
[10] S. S. Dragomir, 'A converse result for Jensen's discrete inequality via Gruss' inequality and applications in information theory', *An. Univ. Oradea Fasc. Mat.* **7** (1999/2000), 178–189.
[11] S. S. Dragomir, 'On a reverse of Jensen's inequality for isotonic linear functionals', *J. Inequal. Pure Appl. Math.* **2**(3) (2001), (Article 36).
[12] S. S. Dragomir, 'A Grüss type inequality for isotonic linear functionals and applications', *Demonstratio Math.* **36**(3) (2003), 551–562.
[13] S. S. Dragomir, 'Bounds for the deviation of a function from the chord generated by its extremities', *Bull. Aust. Math. Soc.* **78**(2) (2008), 225–248.
[14] S. S. Dragomir, 'Bounds for the normalized Jensen functional', *Bull. Aust. Math. Soc.* **74**(3) (2006), 471–476.
[15] S. S. Dragomir and N. M. Ionescu, 'Some converse of Jensen's inequality and applications', *Rev. Anal. Numér. Théor. Approx.* **23**(1) (1994), 71–78.
[16] D. V. Gokhale and S. Kullback, *Information in Contingency Tables* (Marcel Decker, New York, 1978).
[17] J. H. Havrda and F. Charvat, 'Quantification method classification process: concept of structural α-entropy', *Kybernetika* **3** (1967), 30–35.
[18] E. Hellinger, 'Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen', *J. reine angew. Math.* **36** (1909), 210–271.
[19] H. Jeffreys, 'An invariant form for the prior probability in estimating problems', *Proc. R. Soc. Lond. A* **186** (1946), 453–461.
[20] T. T. Kadota and L. A. Shepp, 'On the best finite set of linear observables for discriminating two Gaussian signals', *IEEE Trans. Inform. Theory* **13** (1967), 288–294.
[21] T. Kailath, 'The divergence and Bhattacharyya distance measures in signal selection', *IEEE Trans. Comm. Technology.* (1967), 52–60; (Vol COM-15).
[22] J. N. Kapur, 'A comparative assessment of various measures of directed divergence', *Adv. Management Stud.* **3** (1984), 1–16.
[23] D. Kazakos and T. Cotsidas, 'A decision theory approach to the approximation of discrete probability densities', *IEEE Trans. Perform. Anal. Machine Intell.* **1** (1980), 61–67.

[24] S. Kullback and R. A. Leibler, 'On information and sufficiency', *Annals Math. Statist.* **22** (1951), 79–86.

[25] J. Lin, 'Divergence measures based on the Shannon entropy', *IEEE Trans. Inform. Theory* **37**(1) (1991), 145–151.

[26] M. Mei, 'The theory of genetic distance and evaluation of human races', *Japan J. Human Genetics* **23** (1978), 341–369.

[27] C. P. Niculescu, 'An extension of Chebyshev's inequality and its connection with Jensen's inequality', *J. Inequal. Appl.* **6**(4) (2001), 451–462.

[28] E. C. Pielou, *Ecological Diversity* (Wiley, New York, 1975).

[29] C. R. Rao, 'Diversity and dissimilarity coefficients: a unified approach', *Theoretic Population Biol.* **21** (1982), 24–43.

[30] A. Rényi, 'On measures of entropy and information', *Proc. Fourth Berkeley Symp. Math. Stat. Prob.* **1** (1961), 547–561, (University of California Press).

[31] A. Sen, *On Economic Inequality* (Oxford University Press, London, 1973).

[32] B. D. Sharma and D. P. Mittal, 'New nonadditive measures of relative information', *J. Comb. Inf. Syst. Sci.* **2**(4) (1977), 122–132.

[33] H. Shioya and T. Da-Te, 'A generalisation of Lin divergence and the derivative of a new information divergence', *Elec. and Comm. in Japan* **78**(7) (1995), 37–40.

[34] S. Simic, 'On a global upper bound for Jensen's inequality', *J. Math. Anal. Appl.* **343** (2008), 414–419.

[35] I. J. Taneja, *Generalised Information Measures and their Applications*, http://www.mtm.ufsc.br/~taneja/bhtml/bhtml.html.

[36] H. Theil, *Economics and Information Theory* (North-Holland, Amsterdam, 1967).

[37] H. Theil, *Statistical Decomposition Analysis* (North-Holland, Amsterdam, 1972).

[38] F. Topsoe, 'Some inequalities for information divergence and related measures of discrimination', *Res. Rep. Coll., RGMIA* **2**(1) (1999), 85–98.

[39] I. Vajda, *Theory of Statistical Inference and Information* (Kluwer Academic Publishers, Dordrecht-Boston, 1989).

S. S. DRAGOMIR, Mathematics, School of Engineering & Science,
Victoria University, PO Box 14428, Melbourne City,
MC 8001, Australia
and
School of Computational & Applied Mathematics,
University of the Witwatersrand, Private Bag 3,
Johannesburg 2050, South Africa
e-mail: sever.dragomir@vu.edu.au