JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **139**, 537–551 (1989)

Entropic Means

AHARON BEN-TAL*

Faculty of Industrial Engineering and Management, Technion, Israel Institute of Technology, Haifa, Israel, and Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, Michigan 48109

ABRAHAM CHARNES

Center for Cybernetic Studies, The University of Texas, Austin, Texas 78712

AND

MARC TEBOULLE†

Department of Mathematics, University of Maryland Baltimore County, Catonsville, Maryland 21228

Submitted by J. L. Brenner

Received April 6, 1987

We show how to generate means as optimal solutions of a minimization problem whose objective function is an entropy-like functional. Accordingly, the resulting mean is called *entropic mean.* It is shown that the entropic mean satisfies the basic properties of a weighted homogeneous mean (in the sense of J. L. Brenner and B. C. Carlson (*J. Math. Anal. Appl.* **123** (1987), 265–280)) and that all the classical means as well as many others are special cases of entropic means. Comparison theorems are then proved and used to derive inequalities between various means. Generalizations to entropic means of random variables are considered. An extremal principle for generating the Hardy–Littlewood–Polya generalized mean is also derived. Finally, it is shown that an asymptotic formula originally discovered by L. Hoehn and I. Niven (*Math. Mag.* **58** (1958), 151–156) for power means, and later extended by J. L. Brenner and B. C. Carlson (*J. Math. Anal. Appl.* **123** (1987), 265–280) is valid for entropic means. © 1989 Academic Press, Inc.

* This work was supported in part by the NSF Grant ECS-860-4354.

† Part of this work was conducted while visiting the Center for Cybernetic Studies, The University of Texas, Austin. The author greatly appreciates the hospitality and support provided at this Institution.

0022-247X/89 $3.00

1. Introduction

Most of the classical results on means can still be found in the monographs of Hardy, Littlewood, and Polya [9] and Beckenbach and Bellman [1]. In a recent paper, Brenner and Carlson [5] summarized many important means that have been studied in recent years as well as the classical one that are discussed in [1, 9]. Means are also discussed extensively in the recent book of J. Borwein and P. Borwein [3].

In this paper we show how to generate means as optimal solutions of a minimization problem (E) with a "distance" function as objective. The special feature of this approach is the choice of the "distance" which is defined in terms of an *Entropy*-like function. Accordingly the resulting mean is called *entropic mean.* The motivation for the extremal problem (E) is given in Section 2. We show that the entropic mean satisfies the basic properties of a general mean; see Theorem 2.1. In Section 3 we present many examples demonstrating that all the classical means as well as many others, are special cases of entropic means. Comparison theorems are then proved in Section 4 and used to derive inequalities between various means. In Section 5 we extend our results to derive the entropic mean for random variables and show how classical "measures of centrality" used in statistics (Expectation, Quantiles, in particular Median) are also special cases of entropic means. In Section 6 we consider a new entropy-like distance function, which allows us to derive the generalized mean of Hardy–Littlewood and Polya (HLP). Finally, in the last section it is shown that entropic means are weighted homogeneous means and as such (as shown by Brenner and Carlson [5]) have an interesting asymptotic property, originally discovered by Hoehn and Niven [10] for some special cases of power means.

2. The Entropic Mean

Let $a=(a_1, ..., a_n)$ be given strictly positive numbers and let $w=(w_1, ..., w_n)$ be given weights, i.e., $\sum_{i=1}^n w_i = 1$, $w_i > 0$, $i=1, ..., n$.

We define the mean of $(a_1, ..., a_n)$ as the value x for which the sum of a distance from x to each a_i denoted "dist"(x, a_i) is minimal; i.e., the mean is the optimal solution of

$$\min\left\{\sum_{i=1}^n w_i \text{“dist”}(x, a_i) \colon x \in \mathbb{R}_+ := (0, +\infty)\right\}. \qquad (E)$$

Note that here, the weights w_i give the "relative importance" to the error "dist"(x, a_i). The notation "dist"(α, β) refers here to some measure of distance between α, $\beta > 0$, such a measure must satisfy

$$\text{"dist"}(\alpha, \beta) = \begin{cases} 0 & \text{if} \quad \alpha = \beta \\ > 0 & \text{if} \quad \alpha \neq \beta. \end{cases}$$

In this paper we investigate the concept of *ϕ-divergence* (or *ϕ-relative entropy* as a possible choise for "dist"$(\cdot, \cdot)$. This concept was introduced by Csiszar [7] to measure the dissimilarity, or divergence, between two probability measures. Let ϕ: $\mathbb{R}_+ \to \mathbb{R}$ be a strictly convex differentiable function with $(0, 1] \subset \operatorname{dom} \phi$ and such that $\phi(1) = 0$, $\phi'(1) = 0$. We denote the class of such functions by Φ. In [7] the "distance" between two discrete probability measures was defined as

$$I_\phi(p, q) := \sum_{j=1}^{n} q_j \phi\left(\frac{p_j}{q_j}\right); \qquad p, q \in D_n = \left\{x \in \mathbb{R}^n \colon \sum_{j=1}^{n} x_j = 1, x > 0\right\}.$$

Adopting this concept, we define the distance from x to a_i by "dist"$\{x, a_i\} := d_\phi(x, a_i) := a_i \phi(x/a_i)$ for each $i = 1, ..., n$. The optimization problem (E) is now

$$\min \left\{ \sum_{i=1}^{n} w_i a_i \phi\left(\frac{x}{a_i}\right) \colon x \in \mathbb{R}_+ \right\} \tag{E_{d_ϕ}}$$

and the resulting optimal solution denoted $\bar{x}_\phi(a) := \bar{x}_\phi(a_1, ..., a_n)$ will be called the *entropic mean* of $(a_1, ..., a_n)$. The choice of $d_\phi(x, a_i)$ as a "distance" is supported by the following results.

LEMMA 2.1. *Let $\phi \in \Phi$. Then*

(a) *For any $\beta_2 > \beta_1 \geqslant \alpha > 0$ or $0 < \beta_2 < \beta_1 \leqslant \alpha$,*

$$d_\phi(\beta_2, \alpha) > d_\phi(\beta_1, \alpha). \tag{2.1}$$

(b) *For any $\alpha_2 \geqslant \alpha_1 > \beta > 0$ or $0 < \alpha_2 < \alpha_1 < \beta$,*

$$d_\phi(\beta, \alpha_2) > d_\phi(\beta, \alpha_1). \tag{2.2}$$

Proof. (a) Since ϕ is strictly convex, $d_\phi(\cdot, \alpha)$ is strictly convex for any $\alpha > 0$, thus by the gradient inequality for $d_\phi(\cdot, \alpha)$,

$$d_\phi(\beta_2, \alpha) = \alpha\phi\left(\frac{\beta_2}{\alpha}\right) > \alpha\phi\left(\frac{\beta_1}{\alpha}\right) + (\beta_2 - \beta_1)\phi'\left(\frac{\beta_1}{\alpha}\right). \tag{2.3}$$

From (a) and $\phi'(1)=0$, the second term in the right hand side of (2.1) is nonnegative, hence

$$d_\phi(\beta_2, \alpha) > \alpha\phi\left(\frac{\beta_1}{\alpha}\right) = d_\phi(\beta_1, \alpha).$$

(b) Since ϕ is strictly convex, using the gradient inequality it is easy to verify that $d_\phi(\beta, \cdot)$ is strictly convex for any $\beta > 0$. Then the proof of (2.2) follows using the same arguments as in (a). ∎

COROLLARY 2.1. *Let $\phi \in \Phi$ and α, $\beta > 0$. Then $d_\phi(\beta, \alpha) \geqslant 0$ with equality if and only if $\alpha = \beta$.*

Proof. Since $\phi(1)=0$, then $d_\phi(\alpha,\alpha)=0$. Set $\beta_2=\beta$, $\beta_1=\alpha$ in (2.1) of Lemma 2.1, then $d_\phi(\beta, \alpha) > d_\phi(\alpha, \alpha) = 0$. ∎

Remark 2.1. The differentiability assumption on ϕ can be relaxed. Since ϕ is convex its left and right derivative $\phi'_-(x)$ and $\phi'_+(x)$ exist and are finite and increasing. Moreover the subdifferential of ϕ is $\partial\phi(x) = [\phi'_-(x), \phi'_+(x)]$. Then Lemma 2.1 remains valid if we substitute $0 \in \partial\phi(1)$ for $\phi'(1)$ in (2.1). The existence of $0 \in \partial\phi(1)$ is guaranteed for all strictly convex function $\phi(x) > 0$, all $x \in \mathbb{R}_+ \setminus \{1\}$; see Example 5.1, Section 5.

Note that $d_\phi(\alpha, \beta)$ is not a distance in the usual sense; i.e., it is not symmetric and does not satisfy the triangle inequality. However, d_ϕ is homogeneous, i.e., $\forall \lambda > 0$, $d_\phi(\lambda\alpha, \lambda\beta) = \lambda d_\phi(\alpha, \beta)$. (Compare with (6.1).) The next result demonstrates that the entropic mean $\bar{x}_\phi$ defined as the optimal solution of problem (E_{d_ϕ}), posssesses all the essential properties of a general mean.

THEOREM 2.1. *Let $\phi \in \Phi$. Then*

(i) *There exists a unique continuous function $\bar{x}_\phi$ which solves (E_{d_ϕ}) such that*

$$\min_{1 \leqslant i \leqslant n} \{a_i\} \leqslant \bar{x}_\phi(a) \leqslant \max_{1 \leqslant i \leqslant n} \{a_i\}$$

for all $a_i > 0$. In particular $\bar{x}_\phi(\alpha, ..., \alpha) = \alpha$.

(ii) *The mean $\bar{x}_\phi$ is strict, i.e.,*

$$\min_{1 \leqslant i \leqslant n} \{a_i\} < \max_{1 \leqslant i \leqslant n} \{a_i\} \Rightarrow \min_{1 \leqslant i \leqslant n} \{a_i\} < x_\phi(a) < \max_{1 \leqslant i \leqslant n} \{a_i\}.$$

(iii) *$\bar{x}_\phi$ is homogeneous (scale invariant), i.e.,*

$$\bar{x}_\phi(\lambda a) = \lambda \bar{x}_\phi(a) \qquad \textit{for all} \quad \lambda > 0,\ a_i > 0.$$

(iv) *If* $w_i = w$ *for all* i *then* $\bar{x}_\phi$ *is symmetric; i.e.,* $\bar{x}_\phi(a_1, ..., a_n)$ *is invariant to permutations of the* a_i*'s* > 0.

(v) $\bar{x}_\phi$ *is isotone; i.e., for all* i *and fixed* $\{a_j\}_{j=1}^n > 0$, $j \neq i$, $\bar{x}_\phi(a_1, ..., a_{j-1}, \cdot, a_{i+1}, ..., a_n)$ *is an increasing function.*

Proof. (i) Since $\phi \in \Phi$, as a positive combination of strictly convex functions, $h(x) := \sum w_i a_i \phi(x/a_i)$ is strictly convex and then if an optimal solution $\bar{x} := \bar{x}_\phi$ exists it is unique. The optimality condition for problem (E_{d_ϕ}) implies

$$g(\bar{x}) := \sum_{i=1}^n w_i \phi'\left(\frac{\bar{x}}{a_i}\right) = 0. \tag{2.4}$$

Without loss of generality, assume that $a_1 \leqslant \cdots \leqslant a_n$. We show that for any $x \notin [a_1, a_n]$: $g(x) \neq 0$. If $x < a_i$ then $(x/a_i) < 1$ for all i. But ϕ is strictly convex, hence ϕ' is strictly increasing and thus

$$g(x) = \sum_{i=1}^n w_i \phi'\left(\frac{x}{a_i}\right) < \sum_{i=1}^n w_i \phi'(1) = \sum_{i=1}^n w_i \cdot 0 = 0.$$

Similarly if $x > \max a_i$ we have $g(x) > 0$. Therefore since g is continuous this implies that there exists a continuous function $\bar{x} = \bar{x}_\phi(a_1, ..., a_n) \in [a_1, a_n]$ and such that $g(\bar{x}) = 0$.

(ii) Without loss of generality assume $\min_{1 \leqslant i \leqslant n}\{a_i\} = a_1$, and let $\bar{x} = a_1$. Then from (2.4) and ϕ' being strictly increasing with $\phi'(1) = 0$, we have

$$0 = \sum_{i=1}^n w_i \phi'\left(\frac{\bar{x}}{a_i}\right) = w_1 \phi'(1) + \sum_{i=2}^n w_i \phi'\left(\frac{a_1}{a_i}\right) < 0$$

and hence a contradiction. With a similar proof $\bar{x} < \max\{a_i\}$.

(iii) For any $\lambda > 0$, let $\bar{y}$ be the optimal solution of min $\{\sum_{i=1}^n \lambda w_i a_i \phi(x/\lambda a_i): x \in \mathbb{R}_+\}$. The optimality condition is $\sum_{i=1}^n w_i \phi'(\bar{y}/\lambda a_i) = 0$ which is exactly (2.4) with $\bar{x} = \bar{y}/\lambda$.

(iv) If $w_i = w > 0$ for all i then (2.4) is $\sum_{i=1}^n \phi'(\bar{x}/a_i) = 0$ which is invariant to the permutations of the a_i.

(v) Let x be the optimal solution corresponding to $(a_1, ..., a_i, ..., a_n)$ and $\hat{x}$ let the optimal solution corresponding to $(a_1, ..., \hat{a}_i, ..., a_n)$ with $\hat{a}_i > a_i$ and fixed $\{a_j\}_{j \neq i}$. Suppose $\hat{x} < x$, then from the optimality condition (2.4) for x and $\hat{x}$, and ϕ' being strictly increasing we have

$$0 = \sum_{j=1}^n w_j \phi'\left(\frac{x}{a_j}\right) > \sum_{j \neq i} w_j \phi'\left(\frac{\hat{x}}{a_j}\right) + w_i \phi'\left(\frac{\hat{x}}{\hat{a}_i}\right) = 0$$

and hence a contradiction. ∎

Remark 2.2. If ϕ is not assumed *strictly* convex, then $\bar{x}_\phi$ is not necessarily unique and not necessarily in $[\min a_i, \max a_i]$, see Section 3, Example 3.6. For such ϕ the entropic mean should be redefined as the optimal solution of

$$\min\left\{\sum w_i a_i \phi\left(\frac{x}{a_i}\right): x \in [\min a_i, \max a_i]\right\}.$$

3. Examples

In this section we present may examples demonstrating that classical means as well as many others are special cases of entropic means, for particular choice of the kernel function ϕ. In each of the examples below, we solve the optimality condition equation

$$\sum_{i=1}^{n} w_i \phi'\left(\frac{x}{a_i}\right) = 0. \tag{3.1}$$

Its solution is denoted by $\bar{x}_\phi(a)$. In some examples the function ϕ will be parametrized with one or more parameters α, β, ..., and is denoted $\phi_{\alpha,\beta,\ldots}$. Correspondingly we will denote the entropic mean $\bar{x}_{\alpha,\beta,\ldots}(a)$. We start with the four classical means.

3.1. *Arithmetic mean.* $\phi(t) = -\log t + t - 1$. Then (3.1) yields $-(\sum w_i a_i/\bar{x}) + 1 = 0$, hence $\bar{x}_\phi(a) = \sum_{i=1}^n w_i a_i := A(a)$.

3.2. *Harmonic mean.* $\phi(t) = (t-1)^2$. Then (3.1) yields $\sum_{i=1}^n w_i((x/a_i) - 1)) = 0$, hence $\bar{x}_\phi(a) = (\sum_{i=1}^n (w_i/a_i))^{-1} := H(a)$.

3.3. *Root mean square.* $\phi(t) = 1 - 2\sqrt{t} + t$. Then (3.1) yields $-\sum_{i=1}^n w_i(a_i/x)^{1/2} + 1 = 0$, hence $\bar{x}_\phi(a) = (\sum_{i=1}^n w_i a_i^{1/2})^2 := R(a)$.

3.4. *Geometric mean.* $\phi(t) = t \log t - t + 1$. Then (3.1) yields $\sum w_i \log(x/a_i) = 0$, hence $\bar{x}_\phi(a) = \prod_{i=1}^n a_i^{w_i} := G(a)$.

The four previous examples are particular cases of the mean of order p which can be obtained as follows:

3.5. *Mean of order p.* $\phi_p(t) = (1/(p-1))(t^{1-p} - p) + t$, $p \neq 1$, $p > 0$. Then (3.1) yields

$$-x^{-p} \sum w_i a_i^p + 1 = 0, \qquad \text{hence} \quad \bar{x}_p(a) = \left(\sum_{i=1}^n w_i a_i^p\right)^{1/p}.$$

To extend $\bar{x}_p(a)$ for negative order, one may choose $\phi_q(t) = (t^q - tq)/(q-1) + 1$, $q > 0$, $q \neq 1$, which yields $\bar{x}_q(q) = (\sum_{i=1}^n w_i a_i^{1-q})^{1/(1-q)}$.

Note that $\bar{\phi}_q(t) = t\phi_q(1/t)$ (hence strictly convex $\forall t > 0$) and $\bar{\phi}_q(t) = \phi_p(t)$ for $p = q = \frac{1}{2}$ yielding the root mean square R, while $q = 2$ gives the harmonic mean H. A simple application of L'Hospital's rule shows that the arithmetic and geometric mean are re-obtained respectively by choosing for ϕ, the limiting cases

$$\lim_{p \to 1} \phi_p(t) = -\log t + t - 1, \qquad \lim_{q \to 1} \bar{\phi}_q(t) = t \log t - t + 1.$$

3.6. *Extreme means.* Let $\phi(t) = \max\{0, (1-t)\}^2$. Then $\phi'(t) = -2\max\{0, 1-t\}$, $\sum_{i=1}^n w_i \max(0, 1-(x/a_i)) = 0$, hence $\bar{x}_\phi(a) = \max_{1 \leqslant i \leqslant n} a_i$ implies $1 - (x/a_i) \leqslant 0$ for all i showing that x_ϕ is optimal. Note that here ϕ is not strictly convex. Indeed any $x \geqslant \max_{1 \leqslant i \leqslant n} a_i$ is an optimal solution of problem (E_{d_ϕ}), but following Remark 2.1, in the interval $[\min_{1 \leqslant i \leqslant n} a_i, \max_{1 \leqslant i \leqslant n} a_i]$, $\bar{x}_\phi$ is uniquely optimal.

3.7. *Lehmer mean* [11]. For $0 < p < 1$, let $\phi_p(t) = ((t^{2-p})/(2-p)) - ((t^{1-p})/(1-p)) + (1/(2-p)(1-p))$. Then (3.1) yields $x^{1-p} \sum w_i a_i^{p-1} - x^{-p} \sum w_i a_i^p = 0$, hence

$$\bar{x}_p(a) = \frac{\sum_{i=1}^n w_i a_i^p}{\sum_{i=1}^n w_i a_i^{p-1}}.$$

3.8. *Gini mean* [8]. For any $r, s \in \mathbb{R}$ such that $s \geqslant 0 > r$ or $s > 0 \geqslant r$ let $\phi_{r,s}(t) = ((t^{1-r} - 1)/(1-r)) - ((t^{1-s} - 1)/(1-s))$. Then (3.1) yields $x^{-r} \sum_{i=1}^n w_i a_i^r - x^{-s} \sum_{i=1}^n w_i a_i^s = 0$, hence

$$\bar{x}_{r,s}(a) = \left(\frac{\sum_{i=1}^n w_i a_i^s}{\sum_{i=1}^n w_i a_i^r} \right)^{1/(s-r)}.$$

The Lehmer mean and the mean order p are special cases $(r, s) = (p-1, p)$, $(0, p)$, respectively. With $\bar{\phi}_{r,s}(t) := t\phi_{r,s}(1/t)$ we obtain the mean of order $1 - q (q > 0)$ for $(r, s) = (0, q)$.

3.9. *Composition of means.* Let $\phi(t) = -\frac{2}{3}\log t + (t^2/3) - \frac{1}{3}$. Then (3.1) yields

$$\sum_{i=1}^n w_i \left(\frac{x}{a_i} - \frac{a_i}{x} \right) = 0, \qquad \text{hence} \quad \bar{x}_\phi(a) = \left(\frac{\sum_{i=1}^n w_i a_i}{\sum_{i=1}^n w_i / a_i} \right)^{1/2} = \sqrt{A(a)\, H(a)},$$

i.e., the geometric mean of A and H.

Note that in two dimensions, $n = 2$ with $w_1 = w_2 = \frac{1}{2}$, $\bar{x}_\phi(a_1, a_2) = \sqrt{a_1 a_2} = G(a_1, a_2)$. Hence different choice of ϕ (compare with Example 3.4) may induce the same mean.

4. Comparison of Means

Given two functions ϕ, ψ in the class Φ, can we compare the corresponding entropic means $\bar{x}_\phi$ and $\bar{x}_\psi$?

THEOREM 4.1. *Let $\phi, \psi \in \Phi$ and denote by $\bar{x}_\phi$, $\bar{x}_\psi$ respectively the corresponding entropic means. If there exists a constant $K \neq 0$ such that*

$$K\phi'(t) \leqslant \psi'(t) \qquad \forall t \in \mathbb{R}_+ \setminus \{1\} \tag{4.1}$$

then,

$$\bar{x}_\phi(a) \geqslant \bar{x}_\psi(a).$$

Proof. From the optimality conditions for problems (E_{d_ϕ}), (E_{d_ψ}) it follows respectively that

$$\sum_{i=1}^{n} w_i \phi'\left(\frac{\bar{x}_\phi}{a_i}\right) = 0 \tag{4.2}$$

and

$$\sum_{i=1}^{n} w_i \psi'\left(\frac{\bar{x}_\psi}{a_i}\right) = 0. \tag{4.3}$$

Suppose that $\bar{x}_\phi < \bar{x}_\psi$. Since ψ is strictly convex ψ' is strictly increasing and thus using (4.1) we have

$$K\phi'\left(\frac{\bar{x}_\phi}{a_i}\right) \leqslant \psi'\left(\frac{\bar{x}_\phi}{a_i}\right) < \psi'\left(\frac{\bar{x}_\psi}{a_i}\right) \qquad \text{for all} \quad i = 1, ..., n.$$

Multiplying by $w_i > 0$ and summing the above inequalities imply

$$K \sum_{i=1}^{n} w_i \phi'\left(\frac{\bar{x}_\phi}{a_i}\right) < \sum_{i=1}^{n} w_i \psi'\left(\frac{\bar{x}_\psi}{a_i}\right).$$

Hence by (4.2) and (4.3), $0 < 0$ a contradiction. ∎

Theorem 4.1 may be useful to derive inequalities.

EXAMPLE 4.1. We show that the classical inequalities $A(a) \geqslant G(a) \geqslant H(a)$ are an easy consequence of Theorem 4.1. Using Examples 3.1 and 3.4 with $\phi(t) := -\log t + t - 1$ and $\psi(t) := t \log t - t + 1$ we have $\bar{x}_\phi = A$ and $\bar{x}_\psi = G$. The condition (4.1) of Theorem 4.2 is satisfied with $K = 1$. Indeed, by the convexity of $t \log t$ it follows that $\forall t > 0$, $\log t \geqslant 1 - 1/t$, i.e., $\psi'(t) \geqslant \phi'(t)$ and hence $A \geqslant G$. Using Examples 3.4 and 3.2 with $\phi(t) :=$

$t \log t - t + 1$ and $\psi(t) = (t-1)^2$ we have $\bar{x}_\phi = G$ and $\bar{x}_\psi = H$. From the concavity of $\log t$ it follows that $\forall t > 0$, $t - 1 \geqslant \log t$; i.e., with $K = 2$ the condition (4.1) is satisfied, i.e., $\psi'(t) \geqslant 2\phi'(t)$ and hence $G \geqslant H$.

THEOREM 4.2. *Let* $\phi_1, \phi_2 \in \Phi$ *and* $\phi_\lambda(t) := \lambda\phi_1(t) + (1-\lambda)\,\phi_2(t)$. *Then for all* $0 \leqslant \lambda \leqslant 1$

$$\min\{\bar{x}_{\phi_1}(a), \bar{x}_{\phi_2}(a)\} \leqslant \bar{x}_{\phi_\lambda}(a) \leqslant \max\{\bar{x}_{\phi_1}(a), \bar{x}_{\phi_2}(a)\}.$$

Proof. First note that for all $0 \leqslant \lambda \leqslant 1$, $\phi_\lambda \in \Phi$. Now $\bar{x}_{\phi_\lambda}$ is obtained from

$$\sum_{i=1}^{n} w_i \left\{ \lambda\phi_1'\left(\frac{\bar{x}_{\phi_\lambda}}{a_i}\right) + (1-\lambda)\phi_2'\left(\frac{\bar{x}_{\phi_\lambda}}{a_i}\right)\right\} = 0. \tag{4.4}$$

Assume $\bar{x}_{\phi_\lambda} < \min(\bar{x}_{\phi_1}, \bar{x}_{\phi_2})$ then since ϕ_1', ϕ_2' are strictly increasing we have with (4.4)

$$\lambda \sum_{i=1}^{n} w_i\phi_1'\left(\frac{\bar{x}_{\phi_1}}{a_i}\right) + (1-\lambda) \sum_{i=1}^{n} w_i\phi_2'\left(\frac{\bar{x}_{\phi_2}}{a_i}\right) > 0. \tag{4.5}$$

But from the optimality conditions for $\bar{x}_{\phi_1}$ and $\bar{x}_{\phi_2}$, the left hand of (4.5) is equal to zero, hence the contradiction. Similarly for $\bar{x}_{\phi_\lambda} > \max(\bar{x}_{\phi_1}(a), \bar{x}_{\phi_2}(a))$. ∎

EXAMPLE 4.2. Let $\phi_1(t) = -\log t + t - 1$ and $\phi_2(t) = (t-1)^2$. Then $\bar{x}_{\phi_1} = A(a)$ and $\bar{x}_{\phi_2} = H(A)$. Consider for $\lambda = \frac{2}{3}$, $\phi_\lambda(t) = \lambda\phi_1(t) + (1-\lambda)\,\phi_2(t)$ which is the function used in Example 3.9 and so $\bar{x}_{\phi_\lambda} = \sqrt{A(a)\,H(a)}$. Indeed as predicted by Theorem 4.2 (since $H(a) \leqslant A(a)$), $H(a) \leqslant \sqrt{A(a)\,H(a)} \leqslant A(a)$.

5. ENTROPIC MEAN FOR RANDOM VARIABLES

Let A be a nonnegative random variable with distribution F and support $\operatorname{supp} A := [\alpha, \beta]$, $0 \leqslant \alpha \leqslant \beta \leqslant +\infty$. A natural generalization of problem (E_{d_ϕ}) is

$$\min\left\{ EA\phi\left(\frac{x}{A}\right) := \int_\alpha^\beta t\phi\left(\frac{x}{t}\right) dF(t) \colon x \in \mathbb{R}_+ \right\},$$

where $E(\cdot)$ denotes the mathematical expectation with respect to the random variable A with distribution function $F(\cdot)$. Clearly, the discrete case defined in Section 2 corresponds to the discrete random variable A with $\Pr\{A = a_i\} := w_i$.

In the sequel we assume that all the expectations expressions exist and are finite. With a similar proof as given in Theorem 2.1 we have the following result.

THEOREM 5.1. *Then for any positive random variable A*:

(i) *There exist a unique $\bar{x}_\phi$ which solves (E_{d_ϕ}) such that*

$$\bar{x}_\phi \in \operatorname{supp} A.$$

(ii) *If A is a degenerate random variable, i.e., $A = C$ where C is a positive finite constant, $\bar{x}_\phi = C$.*

(iii) *For all $\lambda > 0$, $\bar{x}_\phi(\lambda A) = \lambda \bar{x}_\phi(A)$.*

Following Examples of Section 3, we can derive the associated *integral means* by solving the optimality condition equation

$$\int_\alpha^\beta \phi'\left(\frac{x}{t}\right) dF(t) = 0.$$

For example, choosing ϕ as defined in Examples 3.1, 3.2, 3.4, 3.5 one obtains respectively:

(1) The Expectation $\bar{x}_\phi = E(A) = \int t \, dF(t)$.

(2) The Harmonic Expectation $\bar{x}_\phi = 1/E(1/A) = 1/\int dF(t)/t$.

(3) The Geometric integral mean $\bar{x}_\phi = e^{E \log A} = e^{\int \log t \, dF(t)}$.

(4) The Integral mean of order p: $\bar{x}_p = \{\int t^p \, dF(t)\}^{1/p}$, $p > 0$.

In Section 2, Remark 2.1, we mention that the differentiability assumption on ϕ can be relaxed. The next example illustrates the derivation of an important "average" concept arising in statistics which is obtained by choosing a nondifferentiable kernel ϕ.

EXAMPLE 5.1. (θth quantile). Let

$$\phi(\xi) = \begin{cases} (1-\theta)(\xi - 1) & \text{if } \xi > 1 \\ \theta(1-\xi) & \text{if } 0 < \xi \leqslant 1 \end{cases} \qquad (0 < \theta < 1).$$

Clearly $\phi(t)$ is not differentiable at $\xi = 1$. Following Remark 2.1, $\partial\phi(1) = [-\theta, 1-\theta]$, hence $0 \in \partial\phi(1)$ and thus the corresponding $d_\phi(\cdot,\cdot)$ satisfies Lemma 2.1 and its Corollary. The objective function of problem (E_{d_ϕ}) is

$$\begin{aligned} h(x) &:= Et\phi\left(\frac{x}{A}\right) = \int_0^\infty t\phi\left(\frac{x}{t}\right) dF(t) \\ &= (1-\theta) \int_0^x (x-t) \, dF(t) + \theta \int_x^\infty (t-x) \, dF(t). \end{aligned}$$

Since $F(\cdot)$ is the distribution function of the positive random variable A this can be simplified to

$$h(x) = xF(x) - \theta x - \int_0^x t\, dF(t) + \theta E(A).$$

It follows that $h'(\bar{x}_\phi) = 0$ is simply $F(\bar{x}_\phi) = \theta$, i.e., $\bar{x}_\phi$ is the θth *quantile* of the continuous random variable A. In particular for $\theta = \frac{1}{2}$, $\bar{x}_\phi$ is the *median.*

6. An Extremal Principle for the (HLP) Generalized Mean

Means not possessing the properties listed in Theorem 2.1 cannot be derived from the solution of problem (E_{d_ϕ}) with the entropy type distance $d_\phi(x, a_i) = a_i \phi(x/a_i)$. In particular the generalized man of Hardy, Littlewood, and Polya $M_\psi(a, w) = \psi^{-1}\{\sum_{i=1}^n w_i \psi(a_i)\}$, where ψ is a strictly monotone function with inverse ψ^{-1}. Indeed this mean is not in general homogeneous (scale invariant). As an important example of the mean M_ψ which attracted much attention in the literature, we mention the *Logarithmic mean* or Stolarsky mean [6, 13, 14] defined in two variables $a, b > 0$, $a \neq b$, by

$$L(a, b) := \left(\frac{\log b - \log a}{b - a}\right)^{-1} = \psi^{-1}\left(\frac{\int_a^b \psi(x)\, dx}{b - a}\right) \tag{6.1}$$

with $\psi(x) := 1/x$. A multidimensional version was recently developed in [12], see also [3, pp. 269–271].

In this section we show that the generalized M_ψ can be also characterized as the solution of problem (E_{D_h}) but with a different type of an entropy like "distance" function $D_h(\cdot, \cdot)$.

Let h: $\mathbb{R}_+ \to \mathbb{R}$ be a strictly convex differentiable function, and let α, $\beta \in \mathbb{R}_+$. We define

$$D_h(\alpha, \beta) := h(\alpha) - h(\beta) - (\alpha - \beta)\, h'(\beta). \tag{6.2}$$

Then from the gradient inequality for $h(\cdot)$ it follows immediately that: $D_h(\alpha, \beta) = 0$ if $\alpha = \beta$ and $D_h(\alpha, \beta) > 0$ if $\alpha \neq \beta$. We note that when $h(t) = (t-1)^2$ then $D_h(\alpha, \beta) = (\alpha - \beta)^2$. Also when $h(t) = \phi(t) = t \log t$ then $D_h(\alpha, \beta) = d_\phi(\alpha, \beta) = \beta \log \alpha/\beta$. The function $D_h(\alpha, \beta)$ is not symmetric (except for $h(t)$ quadratic), does not satisfy the triangle inequality, and is not homogeneous (as $d_\phi(\cdot, \cdot)$ was). Adopting the "distance" D_h, problem (E_{D_h}) is now

$$\min\left\{\sum_{i=1}^n w_i D_h(x, a_i) \colon x \in \mathbb{R}_+\right\}. \tag{E_{D_h}}$$

Substituting $D_h(x, a_i) = h(x) - h(a_i) - (x - a_i)\, h'(a_i)$ in the objective function of (E_{D_h}) we have to solve (since the a_i are given numbers) the *convex* minimization problem

$$\min \left\{ h(x) - x \sum_{i=1}^{n} w_i h'(a_i) \colon x \in \mathbb{R}_+ \right\}.$$

$\bar{x}_h := \bar{x}_h(a)$ solves (E_{D_h}) if and only if

$$h'(\bar{x}_h) = \sum_{i=1}^{n} w_i h'(a_i). \tag{6.3}$$

Since h' is continuous and strictly increasing in $\mathbb{R}_+$ (h being strictly convex) it follows from Theorem 82 [9, p. 65], that there is a unique $\bar{x}_h$ solving (6.3) and such that $\min a_i \leqslant \bar{x}_h \leqslant \max a_i$, unless the a_i's are all equal. Thus

$$\bar{x}_h(a) = (h')^{-1} \left\{ \sum w_i h'(a_i) \right\},$$

where $(h')^{-1}$ denotes the inverse function of h'. Hence with $\psi := h'$, ψ is strictly monotone and we have characterized the generalized mean (HLP): $\bar{x}_\psi(a) = \psi^{-1}\{\sum \omega_i \psi(a_i)\}$.

It is important to note that the mean $\bar{x}_\psi$ is not necessarily homogeneous (while $\bar{x}_\phi$ was). In fact the only homogeneous means $\bar{x}_\psi$ are the means $\bar{x}_p$ obtained with $\psi(t) = t^p$ (see Theorem 84 [9, p. 68]) and these have been shown to be a special case of entropic means.

EXAMPLE 6.1. Let $h(t) = t \log t - (1 + t) \log(1 + t)$. Then h is strictly convex for all $t > 0$ and $h'(t) = \log(t/(1 + t))$, $(h')^{-1}(t) = e^t/(1 + e^t)$. Hence

$$\bar{x}_h(a) = \frac{\prod_{i=1}^{n} a_i^{w_i}}{\prod_{i=1}^{n} (1 + a_i)^{w_i} - \prod_{i=1}^{n} a_i^{w_i}} = \frac{G(a)}{G(1 + a) - G(a)}.$$

Generalization to the case of random variables can be obtained by solving

$$\min \left\{ \int_\alpha^\beta (h(x) - h(t) - (x - t)\, h'(t))\, dF)(t)) \colon x \in \mathbb{R}_+ \right\}$$

or equivalently

$$\min \left\{ h(x) - x \int_\alpha^\beta h'(t)\, dF(t) \colon x \in \mathbb{R}_+ \right\}.$$

The solution is $\bar{x}_h(A) = (h')^{-1}\{Eh'(A)\}$, i.e., with the notation $\psi := h'$

$$\bar{x}_\psi(A) = \psi^{-1} E\psi(A) = \psi^{-1}\left\{\int_\alpha^\beta \psi(t)\, dF(t)\right\}.$$

An interesting example is when A is a random variable uniformly distributed in the interval $[a, b]$, i.e., with density

$$f(t) = \begin{cases} 1/(b-a) & \text{if } \ a \leqslant t \leqslant b \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\bar{x}_\psi(A) = \bar{x}_\psi(a, b) = \psi^{-1}\left\{\int_a^b \frac{\psi(t)}{b-a}\, dt\right\}.$$

The choice $\psi(t) = 1/t$ gives the logarithm mean $\bar{x}_\psi(A) = L(a, b)$. More generally, with $\psi(t) = t^{p-1}$, $(p \neq 1)$, the Stolarsky's power mean [14] is recovered

$$S_p(a, b) = \left\{\frac{b^p - a^p}{p(b-a)}\right\}^{1/(p-1)}.$$

7. Asymptotic Behavior of Entropic Means

In a recent paper, Hoehn and Niven [10] discovered that the mean of order p

$$x_p(a_1, ..., a_n) = \left(\sum w_i a_i^p\right)^{1/p}$$

has the asymptotic property

$$x_p(a_1 + \xi, ..., a_n + \xi) - \xi \to \sum_{i=1}^{n} w_i a_i, \qquad \xi \to +\infty. \tag{7.1}$$

They proved (7.1) for some special values of p. Their results were extended to unweighted power means of all orders by Brenner [4], and for non-homogeneous means by Boas and Brenner [2]. A further extension to a great variety of homogeneous means, including weighted means is given in Brenner and Carlson [5].

In this section we show that the entropic means have the Hoehn–Niven property. Our result is a direct application of [5, Theorem 1].

THEOREM 7.1. *Let $\phi \in \Phi$ and assume ϕ three times continuously differentiable in the neighborhood of $t=1$. If $a_1, ..., a_n$ are fixed and $\xi \to +\infty$ then*

$$x_\phi(a_1+\xi, ..., a_n+\xi) = \xi + \sum_{i=1}^{n} w_i a_i + \mathcal{O}\left(\frac{1}{\xi}\right). \tag{7.2}$$

Proof. From Theorem 2.1, $x_\phi(a)$ is a homogeneous mean satisfying $x_\phi(1) := x_\phi(1, ..., 1) = 1$. We next show that x_ϕ is a weighted mean in the sense of Brenner and Carlson [5], i.e.,

$$\frac{\partial x_\phi}{\partial a_j}(1, ..., 1) = w_j, \qquad j = 1, ..., n. \tag{7.3}$$

Indeed the entropic mean $x_\phi(a)$ is the optimal solution of problem (E_{d_ϕ}) and thus satisfies the optimality condition

$$\sum_{i=1}^{n} w_i \phi'\left(\frac{x_\phi(a)}{a_i}\right) = 0. \tag{7.4}$$

Differentiating the identity (in terms of a) (7.4) with respect to a_j we obtain

$$\frac{\partial x_\phi(a)}{\partial a_j} \sum_{i=1}^{n} \frac{w_i}{a_i} \phi''\left(\frac{x_\phi(a)}{a_i}\right) = \frac{w_j}{a_j^2} \phi''\left(\frac{x_\phi(a)}{a_j}\right) x_\phi(a), \qquad j = 1, ..., n. \tag{7.5}$$

Setting $a_i = 1$ for all $i = 1, ..., n$ and using $x_\phi(1) = 1$, $\phi''(1) > 0$, and $\sum_{i=1}^{n} w_i = 1$, it follows from (7.5) that $(\partial x_\phi / \partial a_j)(1, ..., 1) = w_j$, $j = 1, ..., n$.

Further, the differentiability assumption of ϕ implies that $x_\phi(\cdot)$ is twice continuously differentiable in the neighborhood of $(1, 1, ..., 1)$. Thus, invoking [5, Theorem 1], the asymptotic result (7.2) follows. ∎

ACKNOWLEDGMENT

We are grateful to the Editors for referring us to [5] which provided the motivation for the last section.

REFERENCES

1. E. F. BECKENBACH AND R. BELLMAN, "Inequalities," Springer, Berlin, 1961.
2. R. P. BOAS AND J. L. BRENNER, Asymptotic behavior of inhomogeneous means, *J. Math. Anal. Appl.* **123** (1987), 262–264.
3. J. M. BORWEIN AND P. B. BORWEIN, Pi and the AGM, a study in analytic number theory and computational complexity, "Canadian Mathematical Society Series of Monographs and Advanced Texts," Wiley–Interscience, New York, 1987.

4. J. L. Brenner, Limits of means for large values of the variables, *Pi Mu Epsilon J.* **8** (1985), 160–163.
5. J. L. Brenner and B. C. Carlson, Homogeneous mean values: Weights and asymptotics, *J. Math. Anal. Appl.* **123** (1987), 265–280.
6. B. C. Carlson, The logarithm mean, *MAA Monthly* **79** (1972), 615–618.
7. I. Csiszar, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.* **2** (1967), 299–318.
8. C. Gini, Di una Formula Compressiva delle Medie, *Metron* **13** (1938), 3–22.
9. G. H. Hardy, J. E. Littlewood, and G. Polya, "Inequalities," 2nd ed., Cambridge Univ. Press, London, 1959.
10. L. Hoehn and I. Niven, Averages on the move, *Math. Mag.* **58** (1985), 151–156.
11. D. H. Lehmer, On the compounding of certain means, *J. Math. Anal. Appl.* **36** (1971), 183–200.
12. A. O. Pittenger, The logarithm mean in n-variables, *MAA Monthly* (1985), 99–104.
13. K. B. Stolarsky, Generalizations of the logarithm mean, *Math. Mag.* **48** (1975), 87–92.
14. K. B. Stolarsky, The power and generalized logarithmic means, *MAA Monthly* **87** (1980), 545–548.