Journal of
Mathematical
Inequalities
Volume 1, Number 1 (2007), 23–30

TOWARDS A WELL–DEFINED MEDIAN

AHMAD AL-SALMAN AND MOWAFFAQ HAJJA

(*communicated by J. Pečarić*)

Abstract. The diagonal Δ of $\mathbb{R}^n$ is Chebeshev with respect to the p-norm for every $p \in (1,\infty]$ but not for $p = 1$. As a result, the median is multi-valued, since the median of a data set $\{a_1, \cdots, a_n\}$ can be thought of as the number(s) μ for which the point $(\mu, \cdots, \mu)$ is a point on Δ that best approximates the point $(a_1, \cdots, a_n)$ with respect to the ℓ_1-norm. In this note, it is proved that if $(\mu_p, \cdots, \mu_p)$ is the unique point on Δ that best approximates a fixed point $(a_1, \cdots, a_n)$ with respect to the ℓ_p-norm for $p \in (1,\infty]$, then as p decreases to 1, μ_p converges, and its limit is proposed to be called *the* median of $\{a_1, \cdots, a_n\}$. Along the way, μ_p is shown to be continuous in p for all $p \in (1,\infty]$ in the sense that μ_p converges to μ_q as p goes to q for every $q \in (0,\infty]$.

1. Preliminaries

We fix $n \geqslant 1$, and we let Δ stand for the diagonal of $\mathbb{R}^n$. Thus

$$\Delta = \{\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n : x_1 = \cdots = x_n\}.$$

For $p \in [1,\infty]$, the ℓ_p-norm on $\mathbb{R}^n$ is defined by

$$\|\mathbf{x}\|_p = \begin{cases} \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}, & \text{if } p \in [1,\infty) \\ \max\{|x_j| : j = 1, \cdots, n\}, & \text{if } p = \infty, \end{cases}$$

where $\mathbf{x} = (x_1, \cdots, x_n)$. Note that $\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p$.

It is easy to see that Δ is Chebeshev in $\mathbb{R}^n$ with respect to the ℓ_p-norm for all $p \in (1,\infty]$. In other words, given any $\mathbf{a} = (a_1, \cdots, a_n) \in \mathbb{R}^n$ and any $p \in (1,\infty]$, there is a unique point $(\mu_p, \cdots, \mu_p)$ in Δ whose ℓ_p-distance from $\mathbf{a}$ is minimal. This must be well-known, but it also follows from the discussion below. We keep $\mathbf{a} = (a_1, \cdots, a_n)$ fixed throughout, and we let $(\mu_p, \cdots, \mu_p)$ be its best approximant in Δ with respect to the ℓ_p-norm.

It is clear that μ_∞ is the mid-point of the range of the data set $\{a_1, \cdots, a_n\}$, i.e.,

$$\mu_\infty = \frac{\max\{a_1, \cdots, a_n\} + \min\{a_1, \cdots, a_n\}}{2}. \tag{1}$$

Mathematics subject classification (2000): 26E60, 41A50, 41A52.

Key words and phrases: Median, mean, best approximation, convex, Chebeshev set, Fermat-Torricelli point.

Paper No. 01-03

Here, a data set is any set with possible repetitions. If $p \in (1, \infty)$, then μ_p is the unique point x for which the function $f_p : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by

$$f_p(x) = \|(x, \cdots, x) - (a_1, \cdots, a_n)\|_p^p = \sum_{j=1}^{n} |x - a_j|^p \tag{2}$$

attains its minimum. In other words, μ_p is the unique zero of

$$f_p'(x) = p \sum_{j=1}^{n} \operatorname{sign}(x - a_j)|x - a_j|^{p-1}, \tag{3}$$

where $\operatorname{sign}(t)$ is defined to be 1 if $t \geqslant 0$ and -1 otherwise. The existence and uniqueness of μ_p follow from the strict convexity of f_p which in turn follows from the fact that

$$f_p''(x) = p(p-1) \sum_{j=1}^{n} |x - a_j|^{p-2} > 0.$$

The number μ_p may be called the ℓ_p*-best-approximation mean* of $\mathbf{a}$, or rather of the correspomding data set $\{a_1, \cdots, a_n\}$. It is clear that μ_p is *strictly internal* in the sense that

$$\min\{a_1, \cdots, a_n\} \leqslant \mu_p \leqslant \max\{a_1, \cdots, a_n\},$$

with strict inequality except in the case when $a_1 = \cdots = a_n$. In fact, if $a_1 \leqslant \cdots \leqslant a_n$ and if they are not all equal, then (3) shows that $f_p'(a_1) < 0$ and $f_p'(a_n) > 0$.

2. Continuity of the ℓ_p-best-approximation mean for $p \in (1, \infty]$

In this section, we establish the continuity of μ_p for $p \in (1, \infty]$.

THEOREM 1. *For a fixed point* $\mathbf{a} = (a_1, \cdots, a_n)$, *let* μ_p, $1 < p < \infty$, *be the unique value that minimizes the function* f *defined in (2), and let* μ_∞ *be defined by (1). Then* μ_p *is continuous for* $p \in (1, \infty]$. *In other words,*

$$\lim_{p \to q} \mu_p = \mu_q \ \ \forall\, q \in (1, \infty].$$

Proof. Let $q \in (1, \infty]$ be given, and let p_k be any sequence that converges to q such that μ_{p_k} converges to t, necessarily finite since μ_p is internal. We are to show that $t = \mu_q$. For simplicity, set $t_k = \mu_{p_k}$. In view of (3), μ_p is defined by

$$\sum_{j=1}^{n} \operatorname{sign}(\mu_p - a_j)|\mu_p - a_j|^{p-1} = 0.$$

Then

$$\sum_{j=1}^{n} \operatorname{sign}(t_k - a_j)|t_k - a_j|^{p_k-1} = 0. \tag{4}$$

If $q < \infty$, we take limits as k goes to infinity and we obtain

$$\sum_{j=1}^{n} \operatorname{sign}(t - a_j)|t - a_j|^{q-1} \;=\; 0.$$

Therefore t satisfies the equation that defines μ_q, and therefore $t = \mu_q$, as desired.

If $q = \infty$, we shall show that t is equal to μ_∞. Thus we assume that $t \neq \mu_\infty$ and seek a contradiction. Without loss of generality, we assume that $a_1 \leqslant \cdots \leqslant a_n$ and that $t > \mu_\infty$. Then $\mu_\infty = (a_n + a_1)/2$, and our assumption takes the form $t > (a_1 + a_n)/2$. Dividing (4) by the mid-range $u = (a_n - a_1)/2$ and taking the limit, we see that

$$\lim_{k\to\infty} \left(\sum_{j=1}^{n} \operatorname{sign}(t_k - a_j) \left| \frac{t_k - a_j}{u} \right|^{p_k - 1} \right) \;=\; 0. \tag{5}$$

Since $t > (a_1 + a_n)/2$, then every a_j for which $t - a_j$ is negative has the property that $|t - a_j|$ is less than $|u|$. For such an a_j,

$$\lim_{k\to\infty} \left| \frac{t_k - a_j}{u} \right|^{p_k - 1} = 0.$$

Therefore we may discard the terms in (5) for which $t - a_j < 0$ or equivalently neglect their signs. We obtain

$$\lim_{n\to\infty} \left(\sum_{j=1}^{n} \left| \frac{t_n - a_j}{u} \right|^{p_n - 1} \right) \;=\; 0.$$

This contradicts the asumption that $|t - a_1| > |u|$ which implies that

$$\lim_{k\to\infty} \left| \frac{t_k - a_1}{u} \right|^{p_k - 1} = \infty.$$

Therefore $t = (a_1 + a_n)/2$, as desired. $\square$

3. Towards a unique median

We now turn to the case $p = 1$. Since the function f_1 defined in (2) attains its minimum but not necessarily at a unique point, it follows that Δ is proximinal but not Chebeshev with respect to the ℓ_1-norm. In fact, if $a_1 \leqslant \cdots \leqslant a_n$, then f_1 attains its minimum at the unique point $a_{(n+1)/2}$ if n is odd and at any point in the interval $[a_{n/2}, a_{1+n/2}]$ if n is even. This explains the non-uniqueness of the *median* as ordinarily defined in statistics.

In this section, $\mathbf{a} = (a_1, \cdots, a_n)$ will stand for an arbitrary, but fixed, point in $\mathbb{R}^n$, and $\mu_p = \mu_p(\mathbf{a})$, $p \in (1, \infty]$, is defined, as above, to be the point at which the function f_p defined in (2) attains its minimum. We shall show that the limit of μ_p as p decreases to 1 exists, and we propose to call this limit, that we denote by μ^*, *the* median of $\mathbf{a}$, or rather of the data set $\{a_1, \cdots, a_n\}$. It is obvious that if $n = 2$, then $\mu_p = (a_1 + a_n)/2$ for all p. Also, if all the a_j are equal, then μ_p equals their common value for all p. Thus we exclude the trivial cases $n < 3$ and $a_1 = \cdots = a_n$. Also, there is no loss in generality in assuming that $a_1 \leqslant \cdots \leqslant a_n$.

It is expectedly impossible to have a closed formula for μ_p. However, given $w \in \mathbb{R}$, it is easy to decide whether $w = \mu_p$, $w < \mu_p$, or $w > \mu_p$. In fact, having the parabola-like graph of f_p in mind, it is obvious that

$$\begin{aligned} &w < \mu_p,\ w = \mu_p,\ \text{or}\ w > \mu_p\ \text{ according as} \\ &f_p'(w) < 0,\ f_p'(w) > 0,\ \text{or}\, f_p'(w) = 0,\ \text{respectively.} \end{aligned} \tag{6}$$

We shall freely use this simple, but quite useful, observation.

We start with a simple lemma that will be used in proving the key Lemma 3. This in turn will be used in proving our main Theorem 4.

LEMMA 2. *Let $\alpha_1, \cdots, \alpha_n$, $n \geqslant 1$, be non-zero real numbers, and let $r_1, \cdots, r_n$ be distinct real numbers. Then the function*

$$F(x) = \alpha_1 e^{r_1 x} + \cdots + \alpha_n e^{r_n x}$$

has finitely many zeros in $\mathbb{R}$. The same holds for the function

$$\alpha_1 c_1^x + \cdots + \alpha_n c_n^x,$$

where $c_1, \cdots, c_n$ are any distinct positive real numbers.

Proof. Without loss of generality, assume that $r_1 < \cdots < r_n$, and let

$$g(x) = e^{-r_1 x} f(x) \text{ and } h(x) = e^{-r_n x} f(x).$$

Then

$$\lim_{x \to -\infty} g(x) = \alpha_1 \text{ and } \lim_{x \to +\infty} h(x) = \alpha_n.$$

Therefore g has no zeros near $-\infty$ and h has no zeros near ∞. Since F, g, and h have the same zeros, it follows that the zeros of F all lie in some compact interval. If the set of zeros of F is infinite, it has an accumulation point and F would have to be identically zero. This follows because F, as a function of the complex variable x, is entire. Therefore F has finitely many zeros, as claimed.

The last statement follows by setting $c_j = e^{r_j}$. $\square$

LEMMA 3. *For $p > 1$, let $f_p : \mathbb{R} \to \mathbb{R}$ be defined as in (2) and let μ_p be the unique point at which $f_p(x)$ attains its minimum. Let w be any real number. Then there exists $\delta > 0$ such that either $\mu_p \geqslant w$ for all $p \in (1, 1+\delta)$, or $\mu_p \leqslant w$ for all $p \in (1, 1+\delta)$.*

Proof. We suppose, by way of contradiction, that there are sequences p_i and q_i decreasing to 1 such that $\mu_{q_i} < w < \mu_{q_i}$. It follows from (6) that

$$f_{q_i}'(w) > 0 > f_{p_i}'(w). \tag{7}$$

Now let

$$F(p) \;=\; \frac{f_p'(w)}{p} \;=\; \sum_{j=1}^{n} \operatorname{sign}(w-a_j)|w-a_j|^{p-1}.$$

Then (7) says that $F(q_i) > 0 > F(p_i)$. Since F is continuous as a function of p, it follows from the intermediate value theorem that $F(\xi_i) = 0$ for some ξ_i between p_i and q_i. Also ξ_i decreases to 1 by the Sandwich theorem. Therefore 1 is an accumulation point of the set of zeros of F, contradicting Lemma 2 and completing the proof. $\square$

THEOREM 4. *For $p > 1$, let $f_p : \mathbb{R} \to \mathbb{R}$ be defined as in (2) and let μ_p be the unique point at which $f_p(x)$ attains its minimum. Then as p decreases to 1, μ_p converges, say to μ^*. Also, the convergence is ultimately one-sided in the sense that there exists $\delta > 0$ such that $\mu_p - \mu^*$ has the same sign for all $p \in (1, 1+\delta)$. If the number of indices i for which a_i is greater than μ^* is M and the number of indices i for which a_i is less than μ^* is m, then $M \leqslant n/2$ and $m \leqslant n/2$. Also, if $M > m$, then the convergence is ultimately right-sided, and if $M < m$, then the convergence is ultimately left-sided. Consequently μ^* lies in the traditional median set of* $\mathbf{a}$ *in the sense that μ^* equals $a_{(n+1)/2}$ if n is odd and lies in the interval $[a_{n/2}, a_{1+n/2}]$ if n is even.*

Proof. If $\limsup_{p\to1}\mu_p$ and $\liminf_{p\to1}\mu_p$ are not equal, then any real number w between them would contradict Lemma 3. Therefore $\lim_{p\to1}\mu_p$ exists. Let us denote it by μ^*. Taking $w = \mu^*$ in Lemma 3, we conclude that the convergence is ultimately one-sided.

Now let us treat the case when the convergence of μ_p is ultimately right-sided. Thus $\mu_p \geqslant \mu^*$ for all p close enough to 1. Let m, s, and M be defined by

$$a_1 \leqslant \cdots \leqslant a_m < a_{m+1} = \cdots = a_{m+s} = \mu^* < a_{m+s+1} \leqslant \cdots \leqslant a_{m+s+M},$$

where $m, M \geqslant 0$, $s \geqslant 1$, and $m+s+M = n$. Let a be the common value of $a_{m+1}, \cdots, a_{m+s}$. Then in a right neighborhood of μ^*, we have

$$f_p'(x) \;=\; p\sum_{j=1}^{m}(x-a_j)^{p-1} + ps(x-a)^{p-1} - p\sum_{j=m+s+1}^{m+s+M}(a_j-x)^{p-1}.$$

Therefore,

$$s(\mu_p-a)^{p-1} \;=\; \sum_{j=m+s+1}^{m+s+M}(a_j-\mu_p)^{p-1} - \sum_{j=1}^{m}(\mu_p-a_j)^{p-1}.$$

The right-hand side converges, as p decreases to 1 (and consequently μ_p decreases to μ^*) to $M-m$. Therefore (the indeterminate form) $(\mu_p-a)^{p-1}$ converges, as p decreases to 1, to some L with $sL = M-m$. Also, $L \in [0,1]$, because $(\mu_p-a)^{p-1} \in [0,1]$ for all p sufficiently close to 1. Therefore $0 \leqslant M-m \leqslant s$. Since $s = n-m-M$, it follows that $M-m \leqslant n-m-M$, and therefore $M \leqslant n/2$. Hence $m \leqslant M \leqslant n/2$, as claimed. This also shows that if $m > M$, then the convergence cannot be ultimately right-sided and therefore it must be left-sided.

The case when the convergence of μ_p is ultimately left-sided is similar, and the proof is complete. $\square$

The ultimately left-sidednes of the convergence of μ_p raises the question whether the convergence of μ_p is indeed ultimately momotone. Thus given $a_1, \cdots, a_n$ and letting μ_p be as defined earlier, does there exist $\delta > 0$ such that the convergence of μ_p as p decreases to 1 within the interval $(1, 1+\delta)$ is montone. The next theorem provides an affirmative answer when $n = 3$.

THEOREM 5. *If $n = 3$, then μ_p converges monotonically to the median.*

Proof. Let $a \leqslant b \leqslant c$ be given. We are to show that $\mu_p = \mu_p(a,b,c)$ converges ultimately monotonically. If $b - a = c - b$, then it follows from symmetry that $\mu_p = b$ for all p. Therefore assume that $b - a > c - b$.

Recall that μ_p is the unique point at which $f_p(x) = |x-a|^p + |x-b|^p + |x-c|^p$ attains its minimum. Since

$$f_p'(b) = p(b-a)^{p-1} - p(c-b)^{p-1} > 0,$$

it follows from (6) that $\mu_p < b$, and therefore μ lies in the open interval (a,b). For $x \in (a,b)$,

$$f_p(x) = (x-a)^p + (b-x)^p + (c-x)^p,$$

and

$$f_p'(x) \;=\; p(x-a)^{p-1} - p(b-x)^{p-1} - p(c-x)^{p-1}.$$

Therefore

$$(\mu_p - a)^{p-1} - (b-\mu_p)^{p-1} - p(c-\mu_p)^{p-1} = 0.$$

Let $q > p$. To show that $\mu_q < \mu_p$, it is enough to show that $f_q'(\mu_p) < 0$. In other words, we are to show that $(\mu_p - a)^{q-1} - (b-\mu_p)^{q-1} - (c-\mu_p)^{q-1} < 0$.

Let

$$(\mu_p - a)^{p-1} = \alpha,\ (b-\mu_p)^{p-1} = \beta,\ (c-\mu_p)^{p-1} = \gamma,\ \frac{q-1}{p-1} = s > 1.$$

Then α, β, and γ are positive, and $\alpha = \beta + \gamma$. Therefore

$$\alpha^s = (\beta+\gamma)^s > \beta^s + \gamma^s,$$

or equivalently

$$(\mu_p - a)^{q-1} - (b-\mu_p)^{q-1} - (c-\mu_p)^{q-1} > 0.$$

Thus $f_q'(\mu_p) > 0$, and therefore $\mu_q < \mu_p$.

Therefore as p decreases to 1, μ increases, necessarily to b. $\square$

The question of montonicity of convergence when $n > 3$ remnains open.

4. Relation to distance means and genuine uniqueness of μ^*

It is useful to view the afore-mentioned best approximation means μ_p from the perspective of the *distance means* introduced in [3]. Thus we start with any *distance* d on a set X for which the function

$$f_d(x) = d(x, a_1) + \cdots + d(x, a_n)$$

attains its minimum at a unique point $\nu = \nu_d(\mathbf{a})$ for every data set $\mathbf{a} = \{a_1, \cdots, a_n\}$ in X, and we call the point $\nu_d(\mathbf{a})$ the *distance mean* of $\mathbf{a}$ *associated* with d. Obviously, the ℓ_p-best-approximation mean μ_p, $p \in (1, \infty)$, is nothing but the distance mean associated with the distance d_p defined on $\mathbb{R}$ by

$$d_p(x, y) = |x - y|^p. \tag{8}$$

Here, a *distance* on X is a real-valued symmetric positive-definite function on $X \times X$ that does not necessarily satisfy the triangle inequality; see [2]. Note that μ_∞ does not correspond, at least not obviously, to a distance.

This viewpoint has the advantage of allowing one to talk simultaneously about means of sets of different sizes. More importantly, it allows us to think of several triangle (and generally simplex) centers as distance means. For example, considering the Euclidean norm $\|\cdots\|$ on $\mathbb{R}^n$ and letting α be the distance defined on $\mathbb{R}^n$ by

$$\alpha(\mathbf{x}, \mathbf{y}) \;=\; \|\mathbf{x} - \mathbf{y}\|^2, \;\; \mathbf{x},\; \mathbf{y} \in \mathbb{R}^n,$$

we immediately see that $\nu_\alpha(\mathbf{x}_1, \cdots, \mathbf{x}_N)$ is nothing but the centroid of the points $\mathbf{x}_1, \cdots, \mathbf{x}_N$ for every $\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathbb{R}^n$. Letting β be defined by

$$\beta(\mathbf{x}, \mathbf{y}) \;=\; \|\mathbf{x} - \mathbf{y}\|, \;\; \mathbf{x},\; \mathbf{y} \in \mathbb{R}^n,$$

we see that $\nu_\beta(\mathbf{x}_1, \cdots, \mathbf{x}_N)$ exists and is unique only if the points $\mathbf{x}_1, \cdots, \mathbf{x}_N$ are non-collinear, and that in this case it is nothing but the Fermat-Torricelli point of these points; see [4]. To fit these considerations in the context of best approximation means, one would have to work with unusual spaces of the form $(\mathbb{R}^N)^n$ with

$$\|(\mathbf{x}_1, \cdots, \mathbf{x}_n)\|_p = \left(\sum_{j=1}^{n} \|\mathbf{x}_j\|_p^p \right)^{1/p}, \;\; \mathbf{x}_i \in \mathbb{R}^N.$$

The distance-mean approach also raises a question regarding the uniqueness of the *path* that we have taken in defining μ^*. More precisely, it is conceivable that there exist, beside the distances d_p defined in (8), other sequences δ_p of distances on $\mathbb{R}$ that converge, as p decreases to 1, to the ordinary Euclidean distance and such that the limit $\mu^\dagger$ of their associated distance means is not the same as our μ^*. In this case, $\mu^\dagger$ would have as valid a claim as μ^* to being *the* median. However, the fact that the distances d_p that we have chosen are highly natural may help us accept μ^* as a highly legitimate median.

Another approach for defining the median of a data set $\mathbf{a} = \{a_1, \cdots, a_n\}$ can be based on the Fermat-Torricelli point described above. As noted then, the function defined on $\mathbb{R}^N$ by

$$F(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_1\| + \cdots + \|\mathbf{x} - \mathbf{x}_n\| \tag{9}$$

attains its minimum at a unique point, the so-called Fermat-Torricelli point, if and only if $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are non-collinear. This raises the possibility of defining the median of $a_1, \cdots, a_n$ by embedding $\mathbb{R}$ in $\mathbb{R}^N$ for $N \geqslant 2$, perturbating the points $a_i = (a_i, 0, \cdots, 0)$ in such a way that they stop being collinear, finding the Fermat-Torricelli point of the resulting set of points, and then taking the limit as the a_i move back to their initial collinear position. One wonders whether there are natural ways for performing the perturbation, and how the resulting limit is related to μ^*.

It is of course legitimate to question the usefulness of the proposed μ^* and other possible definitions of the median, and that can be left to statisticians and workers in relevant disciplines to decide.

It would also be interesting to investigate whether μ_p, already proved to be continuous, is differentiable in p, and whether each μ_p is continuous or even differentiable in $\mathbf{a}$. One also raises the same question for μ^*.

Finally, one may consider continuous versions of the above investigation by looking at spaces of integrable functions on measure spaces. In this case, one would restrict attention to spaces of finite measure to ensure that diagonal, i.e., constant, functions are included.

Acknowledgments. The authors are supported by research grants from Yarmouk University, and are grateful for this support. The second-named author is also grateful to Mr. Zuhair Abu-Abbas for fruitful discussions, several years ago, that resulted in another proof that $\mu_p \to \mu_\infty$ as $p \to \infty$ [1].

REFERENCES

[1] Z. ABU-ABBAS, Private Communications, 1995.
[2] F. GALVIN AND S. D. SHORE, *Distance functions and topology*, Amer. Math. Monthly, **98** (1991), 620–623.
[3] M. HAJJA, *Distance means and best approximation means: Some elementary questions*, preprint.
[4] Y. S. KUPITZ AND H. MARTINI, *Geometric aspects of the generalized Fermat-Torricelli problem, Intuitive Geometry*, Bolyai Society Mathematical Studies, Vol. 6, pp. 55–127, 1997.

(Received August 27, 2006)

Ahmad Al-Salman
Mathematics Department
Yarmouk University
Irbid, Jordan
e-mail: alsalman@yu.edu.jo

Mowaffaq Hajja
Mathematics Department
Yarmouk University
Irbid, Jordan
e-mail: mhajja@yu.edu.jo
e-mail: mowhajja@yahoo.com