ON MEASURES OF CENTRAL LOCATION

HENDRA GUNAWAN
DEPARTMENT OF MATHEMATICS, BANDUNG INSTITUTE OF TECHNOLOGY
BANDUNG 40132, INDONESIA

Abstract. Given some data consisting of a finite sequence of real numbers, we study some measures of central location of the data which we define to be the best constants that approximate the data in certain norms. We also discuss how one measure of central location is related to another.

1. Introduction and Main Results

Given a finite sequence of real numbers, $(x_1, \dots, x_N)$, say, one usually computes **the average** or **the (arithmetic) mean**, among others, as the measures of central location. Here the mean is given by $\mu = \frac{1}{N}\sum_{i=1}^N x_i$. One property of the mean is that the sum of **the deviations** is equal to 0, that is, $\sum_{i=1}^N (x_i - \mu) = 0$. Another nice property of the mean is that it minimizes the function $F(\xi) := \sum_{i=1}^N (x_i - \xi)^2$. If we denote by l_N^2 the space of real sequences $X = (x_i)_{i=1}^N$ of length N, equipped with the norm $\|X\|_2 := \left(\sum_{i=1}^N x_i^2\right)^{1/2}$, then the last property says that the average of a sequence X is the best constant that approximates X in l_N^2.

Now, for each $1 \le p \le \infty$, let l_N^p denote the space of real sequences $X = (x_i)_{i=1}^N$ of length N, equipped with the norm $\|X\|_p := \left(\sum_{i=1}^N |x_i|^p\right)^{1/p}$. (Note that, for $p = \infty$, $\|X\|_\infty$ takes the maximum value of x_i, $i = 1, \dots, N$.) Then, given a sequence $X = (x_i)_{i=1}^N$, one may view it as a member of l_N^p and ask: what is the best constant that approximates X in l_N^p, that is, what is the value of $\xi = \xi_p$ that will minimize the function $F_p(\xi) := \|X - \xi\|_p$, where $X - \xi = (x_i - \xi)_{i=1}^N$?

The following theorem tells us that such a value of ξ_p at which F_p is minimum always exists and, in most cases, is unique.

2000 Mathematical Subject Classification: 26A06, 41A44, 41A50.
Keywords and phrases: measures of central location, approximation by constants.

Theorem 1. *For each $1 \le p \le \infty$, there exists $\xi_p \in \big[\min_{i=1,\ldots,N} x_i, \max_{i=1,\ldots,N} x_i\big]$ at which F_p is minimum; and such a value of ξ_p is unique except for the case where $p = 1$ and N is even.*

Proof. First observe that $F_p(\xi) > F_p\big(\min_{i=1,\ldots,N} x_i\big)$ for $\xi < \min_{i=1,\ldots,N} x_i$ and that $F_p(\xi) > F_p\big(\max_{i=1,\ldots,N} x_i\big)$ for $\xi > \max_{i=1,\ldots,N} x_i$. Hence our search for ξ_p at which F_p is minimum can be restricted on the interval $I := \big[\min_{i=1,\ldots,N} x_i, \max_{i=1,\ldots,N} x_i\big]$. But F_p is continous in ξ, and so by the Maximum-Minimum Theorem such a value of ξ_p must exist.

To prove the uniqueness, we shall use the fact that minimizing F_p amounts to minimizing $G_p := F_p^p$, especially for $1 < p < \infty$. For these values of p's, G_p is differentiable and $G_p'(\xi) = -p\sum_{i=1}^N |x_i - \xi|^{p-1}\mathrm{sgn}(x_i - \xi)$ for every $\xi \in I$. Since $G_p'\big(\min_{i=1,\ldots,N} x_i\big) < 0$ and $G_p'\big(\max_{i=1,\ldots,N} x_i\big) > 0$, G_p must achieve its minimum value at an interior point of I at which G_p' is 0. But G_p is also twice differentiable and $G_p''(\xi) = p(p-1)\sum_{i=1}^N |x_i - \xi|^{p-1} > 0$ for every $\xi \in I$, and so we conclude that such a point must be unique (for otherwise we would have a point in I at which G_p'' is equal to 0). This proves the uniqueness of ξ_p at which G_p, as well as F_p, is minimum for $1 < p < \infty$. The proofs of the two remaining cases are implicitly contained in the proofs of Theorems 2 and 3 below. □

For $p = 2$, we know that F_2 achieves its minimum at $\xi_2 = \mu$, the mean. For $1 \le p \le \infty$ in general, we may interpret the value of ξ_p at which F_p is minimum as *the measure of central location of X in l_N^p*. In addition to the classical result for $p = 2$, we also have interesting results for $p = 1$ and for $p = \infty$. See [**?**] for analogous results in the continuous case. See also [**?**] for related work.

Theorem 2. *If X is a monotone sequence, then F_1 is minimized at* **the median**.

`Note`. If N is odd, the median of $X = (x_i)_{i=1}^N$ is the middle term $x_{\frac{N+1}{2}}$. If N is even, the median of X is $\frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1})$.

Proof. Let us first consider the case where N is odd. Without loss of generality, we may assume that X is increasing. Our task is to show that $F_1(\xi) = \sum_{i=1}^N |x_i - \xi|$ is minimized at $\xi_1 = x_{\frac{N+1}{2}}$, that is,

$$\sum_{i=1}^{N} |x_i - x_{\frac{N+1}{2}}| \leq \sum_{i=1}^{N} |x_i - \xi|$$

for every $\xi \in [x_1, x_N]$. To do so, observe that if $\xi = x_{\frac{N+1}{2}} + \delta$ for some $\delta > 0$, then

$$\begin{aligned}
\sum_{i=1}^{N} |x_i - \xi| &= (x_{\frac{N+1}{2}} + \delta - x_1) + \cdots + (x_{\frac{N+1}{2}} + \delta - x_{\frac{N+1}{2}}) \\
&\quad + |x_{\frac{N+3}{2}} - x_{\frac{N+1}{2}} - \delta| + \cdots + |x_N - x_{\frac{N+1}{2}} - \delta| \\
&\geq \sum_{i=1}^{\frac{N+1}{2}} |x_i - x_{\frac{N+1}{2}}| + \frac{N+1}{2}\delta + \sum_{i=\frac{N+3}{2}}^{N} |x_i - x_{\frac{N+1}{2}}| - \frac{N-1}{2}\delta \\
&= \sum_{i=1}^{N} |x_i - x_{\frac{N+1}{2}}| + \delta \\
&> \sum_{i=1}^{N} |x_i - x_{\frac{N+1}{2}}|.
\end{aligned}$$

Similarly, if $\xi = x_{\frac{N+1}{2}} - \delta$ for some $\delta > 0$, then

$$\sum_{i=1}^{N} |x_i - \xi| > \sum_{i=1}^{N} |x_i - x_{\frac{N+1}{2}}|.$$

This proves the case where N is odd.

The proof of the case where N is even is similar, except that now F_1 is minimized at $\xi_1 = \frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1})$, that is,

$$\sum_{i=1}^{N} \Big|x_i - \frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1})\Big| \leq \sum_{i=1}^{N} |x_i - \xi|$$

for every $\xi \in [x_1, x_N]$, and that the equality holds for every $\xi \in [x_{\frac{N}{2}}, x_{\frac{N}{2}+1}]$ (meaning that F_1 actually achieves its minimum at every point in $[x_{\frac{N}{2}}, x_{\frac{N}{2}+1}]$). □

Theorem 3. *For any sequence X, the function F_∞ achieves its minimum value at $\xi_\infty = \frac{1}{2}\big(\min_{i=1,\ldots,N} x_i + \max_{i=1,\ldots,N} x_i\big)$.*

Proof. Obvious. □

It may be suitable for us to call $\xi_\infty = \frac{1}{2}\big(\min_{i=1,\dots,N} x_i + \max_{i=1,\dots,N} x_i\big)$ *the centre* of the sequence. Then, one can also think of *the radius* of the sequence as *the maximum deviation* about the centre, which is of course equal to half of **the range**:

$$\text{radius} := \max_{i=1,\dots,N} |x_i - \xi_\infty| = \frac{1}{2}\big(\max_{i=1,\dots,N} x_i - \min_{i=1,\dots,N} x_i\big).$$

While there is no obvious relation between the measures of central location ξ_p's at which F_p's are minimized, the following theorem tells us that there is a nice relation between the values of $\|X - \xi_p\|_p$'s.

Theorem 4. *For any sequence $X = (x_i)_{i=1}^N$, we have*

$$\Big(\frac{1}{N}\sum_{i=1}^N |x_i - \xi_p|^p\Big)^{1/p} \le \Big(\frac{1}{N}\sum_{i=1}^N |x_i - \xi_q|^q\Big)^{1/q},$$

whenever $1 \le p < q \le \infty$.

Proof. By the fact that ξ_p minimizes $\|X - \xi_p\|_p$ and Hölder's inequality, we have

$$\begin{aligned}\sum_{i=1}^N |x_i - \xi_p|^p &\le \sum_{i=1}^N |x_i - \xi_q|^p \\ &\le \Big(\sum_{i=1}^N |x_i - \xi_q|^q\Big)^{p/q}\Big(\sum_{i=1}^N 1\Big)^{1-p/q} \\ &= N^{1-p/q}\Big(\sum_{i=1}^N |x_i - \xi_q|^q\Big)^{p/q},\end{aligned}$$

whence

$$\frac{1}{N}\sum_{i=1}^N |x_i - \xi_p|^p \le \Big(\frac{1}{N}\sum_{i=1}^N |x_i - \xi_q|^q\Big)^{p/q}.$$

Taking the p-th root of both sides, we get the desired inequality. □

`Remark`. As a consequence of the above theorem, we have in particular the inequalities

$$\frac{1}{N}\sum_{i=1}^N |x_i - \xi_1| \le \Big(\frac{1}{N}\sum_{i=1}^N (x_i - \xi_2)^2\Big)^{1/2} \le \max_{i=1,\dots,N} |x_i - \xi_\infty|,$$

which relates **the mean absolute deviation** (about the median), **the standard deviation** (about the mean), and the maximum deviation (about the centre) of any sequence $(x_i)_{i=1}^N$.

2. More about the Mean and the Centre

Given a monotone sequence $(x_i)_{i=1}^N$ of real numbers, it is rather laborious to compute the mean, especially when N is large. In contrast, it is very easy for us to compute the centre, which is just half of the sum of the first and the last terms. Now, if one is curious to know how far the mean is from the centre, one may express the difference between the two values as

$$\frac{x_1+x_N}{2}-\frac{1}{N}\sum_{i=1}^{N}x_i=\frac{1}{N}\sum_{i=1}^{N-1}\Big(i-\frac{N}{2}\Big)\Delta_i$$

where $\Delta_i=x_{i+1}-x_i,\ i=1,\dots,N-1$. (Here we assume that $N\geq 2$.) Next, since $\sum_{i=1}^{N-1}\big(i-\frac{N}{2}\big)=0$, we have

$$\frac{x_1+x_N}{2}-\frac{1}{N}\sum_{i=1}^{N}x_i=\frac{1}{N}\sum_{i=1}^{N-1}\Big(i-\frac{N}{2}\Big)(\Delta_i-\xi)$$

for any $\xi\in\mathbf{R}$. Taking the absolute values, we get

$$\Big|\frac{x_1+x_N}{2}-\frac{1}{N}\sum_{i=1}^{N}x_i\Big|\leq\frac{1}{N}\sum_{i=1}^{N-1}\Big|i-\frac{N}{2}\Big|\,|\Delta_i-\xi|$$

for any $\xi\in\mathbf{R}$. By Hölder's inequality, we have in particular the following two estimates

$$\Big|\frac{x_1+x_N}{2}-\frac{1}{N}\sum_{i=1}^{N}x_i\Big|\leq\frac{N-2}{2N}\sum_{i=1}^{N-1}|\Delta_i-\xi| \tag{1}$$

and

$$\Big|\frac{x_1+x_N}{2}-\frac{1}{N}\sum_{i=1}^{N}x_i\Big|\leq S_N\max_{i=1,\dots,N-1}|\Delta_i-\xi| \tag{2}$$

for any $\xi\in\mathbf{R}$, where $S_N=\frac{N-2}{4}$ (if N is even) or $\frac{(N-1)^2}{4N}$ (if N is odd). The problem then is to choose $\xi=\xi_1$ in (1) at which $\sum_{i=1}^{N-1}|\Delta_i-\xi|$ is minimum, and to choose $\xi=\xi_\infty$ in (2) at which $\max\limits_{i=1,\dots,N-1}|\Delta_i-\xi|$ is minimum.

From Theorem 2, we know that in order to minimize $\sum_{i=1}^{N-1}|\Delta_i-\xi|$, we must first rearrange $(\Delta_i)_{i=1}^{N-1}$, to get a monotone sequence $(\Delta_i^*)_{i=1}^{N-1}$, and then choose $\xi=\Delta^*_{\frac{N}{2}}$ (if N is even) or $\frac{1}{2}(\Delta^*_{\frac{N-1}{2}}+\Delta^*_{\frac{N+1}{2}})$ (if N is odd). Meanwhile, from Theorem 3, we know that in order to minimize $\max_{i=1,\ldots,N-1} |\Delta_i - \xi|$ we must choose $\xi = \frac{1}{2}\big(\min_{i=1,\ldots,N-1} \Delta_i + \max_{i=1,\ldots,N-1} \Delta_i\big)$. If, for instance, N is even, then the estimate (1) will become

$$\Big|\frac{x_1 + x_N}{2} - \frac{1}{N}\sum_{i=1}^{N} x_i\Big| \le \frac{N-2}{2N}\sum_{i=1}^{N-1} |\Delta_i^* - \Delta_{\frac{N}{2}}^*|,$$

while the estimate (2) becomes

$$\Big|\frac{x_1 + x_N}{2} - \frac{1}{N}\sum_{i=1}^{N} x_i\Big| \le \frac{N-2}{8}\big(\max_{i=1,\ldots,N-1} \Delta_i - \min_{i=1,\ldots,N-1} \Delta_i\big).$$

Both estimates tell us that the mean will not be very far from the centre if the Δ_i's are approximately constant, as we expect. Note, however, that the second estimate is easier to compute than the first one.

References

[1] S.S. Dragomir and A. Mcandrew, "On trapezoid inequality via a Grüss type result and applications", *Tamkang J. Math.* **31** (2000), 193–201.

[2] H. Gunawan, "A note on Dragomir-Mcandrew's trapezoid inequalities", *Tamkang J. math.*, to appear.

E-mail address: `hgunawan@dns.math.itb.ac.id`