## *Average And Median As Minimums*

In the following exposition, we prove two claims that were claimed in the class without proof and therefore enrich our view about those terms. Given a sequence $\{a_i\}_{i=1}^n$ we call

$$s(x) := \frac{1}{n}\sum_{i=1}^{n}(a_i - x)^2$$

the *mean square difference* of $x$ from $\{a_i\}_{i=1}^n$. Moreover, $\mu := \frac{1}{n}\sum_{i=1}^n a_i$ is called the *mean* or the *average* of the sequence.

**CLAIM 1.1.** *The average minimizes the mean square function.*

PROOF OF CLAIM 1.1. Differentiating $s$ yields

$$s'(x) = \frac{1}{n}\sum_{i=1}^{n} 2(a_i - x)\cdot(-1) = -\frac{2}{n}\sum_{i=1}^{n}(a_i - x) = 2x - \frac{2}{n}\sum_{i=1}^{n} a_i.$$

Comparing to zero then implies that $x = \frac{1}{n}\sum_{i=1}^n a_i = \mu$ which is indeed the mean of the sequence $\{a_i\}_{i=1}^n$. It is easy to verify that this is indeed the minimum as for every $\varepsilon > 0$ it holds that $s'(\mu + \varepsilon) > 0$ and $s'(\mu - \varepsilon) < 0$. ■

For our next purpose, we define the *mean of absolute deviations* as follows:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n}|a_i - x|.$$

**CLAIM 1.2.** *The median minimizes the mean of absolute deviation.*

PROOF OF CLAIM 1.2. We assume without loss of generality[1] that $\nu$ is bigger than more $m$ values than it's smaller than, then making $\nu$ smaller increases its distance from $m$ values and decreases its distance $n-m$ values, all by the same amount, say $d$, so the sum of the distances changes by $dm - d(n-m) = d(2m-n)$. Then $\nu$ will be minimized when $2m$ will be as closed as possible to $n$. That is, if $n$ is even then $m = n/2$. Otherwise, either $m = \lfloor n/2 \rfloor$ or $m = \lceil n/2 \rceil$. ■

After we proved what was required and state in the class we turn into generalize the first argument. To that end, let $\mu := \mathbb{E}[X]$ for a given random variable $X$. We claim that $\mu$ minimizes $f(m) = \mathbb{E}\left[(X-m)^2\right]$. Differentiating[2] yields

$$\frac{d}{dm}\mathbb{E}\left[(X-m)^2\right] = \mathbb{E}\left[\frac{d}{dm}(X-m)^2\right] = \mathbb{E}\left[2(X-m)\cdot(-1)\right] = 2m - 2\mathbb{E}[X].$$

Comparing to zero implies that $m = \mathbb{E}[X] = \mu$. This is indeed a minimum as $f''(m) = 2 > 0$. This argument gives one more justification for defining the variance to be $\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right]$.

[1] The second case is symmetric.

[2] We have assumed that the interchange of the differentiation and expectation operators is legitimate. This assumption can almost always be justified.