MEDIAN-MEAN INEQUALITY

LARRY, ZHIRONG LI

Prove that the median is no more than one standard deviation away from the mean, i.e.,

$$|median\,(X) - mean\,(X)| \le std\,(X)$$

The definition of the median of a random variable is given by [1]

For any probability distribution on the real line with cumulative distribution function F, regardless of whether it is any kind of continuous probability distribution, in particular an absolutely continuous distribution (and therefore has a probability density function), or a discrete probability distribution, a median m satisfies the inequalities

$$\mathbb{P}(X \le m) \ge \frac{1}{2} \wedge \mathbb{P}(X \ge m) \ge \frac{1}{2}$$

or

$$\int_{-\infty}^{m} \mathrm{d}F(x) \ge \frac{1}{2} \;\wedge \int_{m}^{\infty} \mathrm{d}F(x) \ge \frac{1}{2}$$

For an absolutely continuous probability distribution with probability density function f, we have

$$\mathbb{P}(X \le m) = \mathbb{P}(X \ge m) = \int_{-\infty}^{m} f(x)\mathrm{d}x = \frac{1}{2}.$$

The above claim can be proved using Chebyshev inequality.

Proof. WLOG, suppose $median\,(X) > mean\,(X)$, if $median\,(X) - mean\,(X) > std\,(X)$, i.e., $median\,(X) > mean\,(X) + std\,(X)$, then based on one tailed Chebyshev inequality,

$$\frac{1}{2} \le \mathbb{P}\left(X \ge median\,(X)\right) < \mathbb{P}\left(X \ge mean\,(X) + std\,(X)\right) \le \frac{var\,(X)}{std\,(X)^2 + var\,(X)} = \frac{1}{2}$$

A contradiction is already arrived. It is similar for $median\,(X) < mean\,(X)$ case using the other version of one tailed Chebyshev inequality. ☐

The one-tailed Chebyshev inequality

$$\mathbb{P}\left(X - mean\,(X) \ge t\right) \le \frac{var\,(X)}{t^2 + var\,(X)}$$

where $t > 0$ can be proved by the following arguments.

Proof.

$$\begin{aligned} var\,(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}^2\left[X\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[X^2\,|A\right]\right] - \left(\mathbb{E}\left[\mathbb{E}\left[X\,|A\right]\right]\right)^2 \\ &= \mathbb{E}\left[\mathbb{E}\left[X^2\,|A\right] - \mathbb{E}^2\left[X\,|A\right]\right] + \mathbb{E}\left[\mathbb{E}^2\left[X\,|A\right]\right] - \left(\mathbb{E}\left[\mathbb{E}\left[X\,|A\right]\right]\right)^2 \\ &= \mathbb{E}\left[var\left(X\,|A\right)\right] + var\left(\mathbb{E}\left[X\,|A\right]\right) \end{aligned}$$

Date: Dec.18, 2011.

Let p, q, r denote the probability for $X > t, X = t, X < t$, respectively and let A be the indicator random variable accordingly with the following definition

$$A = \begin{cases} 1 & X > t \\ 0 & X = t \\ -1 & X < t \end{cases}$$

Hence WLOG,

$$\mathbb{E}[X] = p\mathbb{E}[X|A=1] + q\mathbb{E}[X|A=0] + r\mathbb{E}[X|A=-1] = 0$$

since $var(X+s) = var(X)$ where s is a constant so the one tailed Chebyshev inequality is equivalent to prove the X with $\mathbb{E}[X] = 0 = \mathbb{E}[\mathbb{E}[X|A]]$.

Notice that

$$\begin{aligned} var(X) &= \mathbb{E}[var(X|A)] + var(\mathbb{E}[X|A]) \\ &\geq var(\mathbb{E}[X|A]) \\ &= \mathbb{E}\left[(\mathbb{E}[X|A])^2\right] \\ &\geq pt^2 + qt^2 + r(\mathbb{E}[X|A=-1])^2 \end{aligned}$$

Keep in mind that

$$\begin{aligned} 0 &= p\mathbb{E}[X|A=1] + q\mathbb{E}[X|A=0] + r\mathbb{E}[X|A=-1] \\ &\geq pt + qt + r\mathbb{E}[X|A=-1] \end{aligned}$$

Hence

$$(\mathbb{E}[X|A=-1])^2 \geq \frac{(1-r)^2 t^2}{r^2}$$

so

$$var(X) \geq (1-r)t^2 + \frac{(1-r)^2 t^2}{r} = t^2(1-r)\left(\frac{r+1-r}{r}\right) = \frac{t^2(1-r)}{r} = \frac{t^2 s}{1-s}$$

where

$$s = 1 - r = p + q$$

Hence

$$var(X) \geq s\left(t^2 + var(X)\right)$$

Therefore

$$\frac{var(X)}{t^2 + var(X)} \geq s = \mathbb{P}(X \geq t)$$

□

It is worth checking out the following two articles about the proof for Chebyshev inequality and connection with the mode, median and mean of a random variable. [2, 3]

References

[1] Median @ http://en.wikipedia.org/wiki/Median
[2] Chebyshev's Inequalities @ http://www.mcdowella.demon.co.uk/Chebyshev.html
[3] Chebyshev's inequality and a one-tailed version @ http://www.btinternet.com/~se16/hgb/cheb.htm