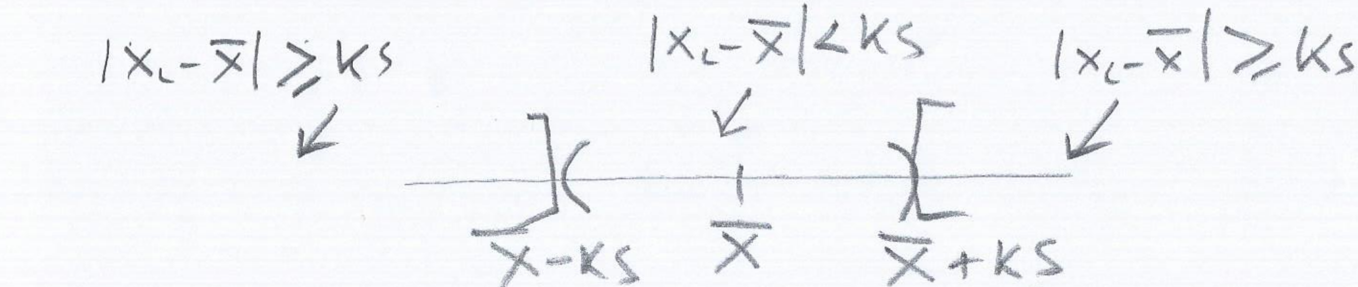

imagine a sample of size n ie $\{x_1, x_2, \ldots, x_n\}$

Then $\bar{x}$ = mean & s = sample standard deviation.

Let $S_k = \{i,\ 1 \leq i \leq n : |x_i - \bar{x}| < ks\}$ $\quad k > 1$

Then S_k is the set of all indices i of x_i such that x_i is in the interval above $(\bar{x}-ks, \bar{x}+ks)$.

Now let $N(S_k)$ = nu. of elements in S_k = number of x's in the interval

Then $(n-1)s^2 = \sum_i (x_i - \bar{x})^2$ — def of s^2

$= \sum_{i \in S_k} (x_i - \bar{x})^2 + \sum_{i \notin S_k} (x_i - \bar{x})^2$ — since i corresponds to x_i in or out of the interval

$\therefore (n-1)s^2 \geq \sum_{i \notin S_k} (x_i - \bar{x})^2$ — since $\sum_{i \in S_k} (x_i - \bar{x})^2 \geq 0$

now when x_i is out of the interval we must have $|x_i - \bar{x}| \geq ks$

$\therefore (x_i - \bar{x})^2 \geq k^2 s^2$

$\therefore (n-1)s^2 \geq (n - N(S_k))k^2 s^2$ — because there are n total x_i values $N(S_k)$ in the interval and $n - N(S_k)$ out

Divide by nk^2s^2 $\Longleftrightarrow$ $\frac{(n-1)\cancel{s^2}}{nk^2\cancel{s^2}} \geq \frac{(n-N(S_k))\cancel{k^2s^2}}{n\cancel{k^2s^2}} \Longleftrightarrow \frac{(n-1)}{nk^2} \geq 1 - \frac{N(S_k)}{n}$

So we get

$$\frac{N(S_k)}{n} \geq 1 - \frac{(n-1)}{nk^2}$$ — this is one form of Chebyshev

$$\Longleftrightarrow \frac{N(S_k)}{n} \geq 1 - \frac{\cancel{n}}{\cancel{n}k^2} + \frac{1}{nk^2}$$

$$\Longleftrightarrow \frac{N(S_k)}{n} \geq 1 - \frac{1}{k^2} + \frac{1}{nk^2}$$ — this last term is positive so

$$\Longleftrightarrow \boxed{\frac{N(S_k)}{n} > 1 - \frac{1}{k^2}}$$ — this is Chebyshev proper

This says the proportion of x values in the interval is $> 1 - \frac{1}{k^2}$ OR better the proportion of data within k standard deviations of the mean is $>$

$$1 - \frac{1}{k^2}$$