Ref: Wikipedia "Bessel's Correction"

If have a set of measurements $x_1, x_2, x_3, \ldots, x_N$,

then can construct a sample dist'n with mean

$\bar{x} = \frac{1}{N}\sum_i x_i$ & std. dev. $s^2 = \frac{1}{N-1}\sum(x_i - \bar{x})^2$.

The parent dist'n has mean $\mu = \lim_{N\to\infty}\left(\frac{1}{N}\sum_i x_i\right)$

& std. dev. $\sigma^2 = \lim_{N\to\infty}\left[\frac{1}{N}\sum(x_i-\mu)^2\right]$.

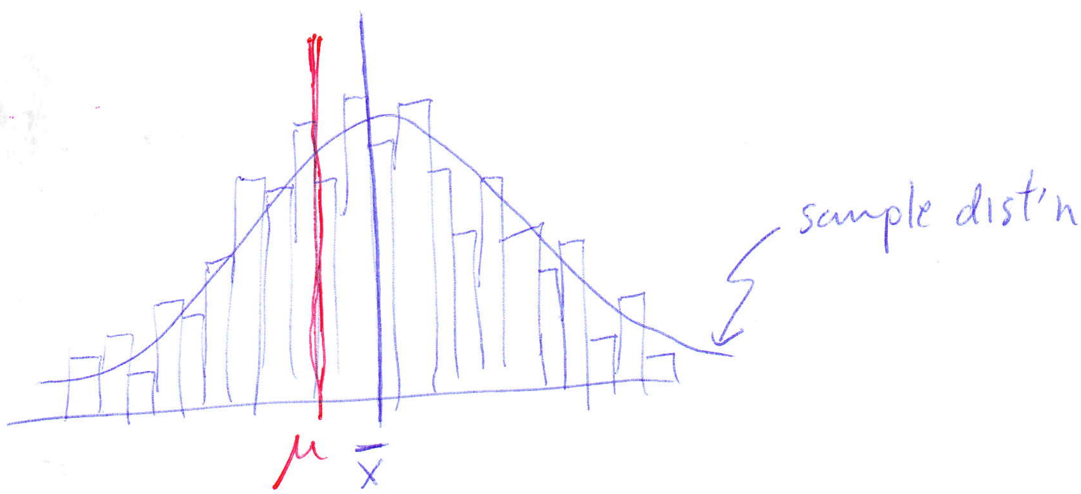
Goal is to try to understand why s^2 has factor $\frac{1}{N-1}$

instead of $\frac{1}{N}$ as in σ^2.

<u>Note</u>: when N is large, the difference between

$\frac{1}{\sqrt{N}}$ & $\frac{1}{\sqrt{N-1}}$ is negligible.

First, one can intuitively understand that, if one was to estimate σ^2 using $\frac{1}{N}\sum(x_i-\bar{x})^2$, the result would be too small.



the x_i measurements, by construction, are closer to $\bar{x}$ than they are to μ (mean of parent dist'n).

$\therefore$ the average $(x_i-\bar{x})^2$ will be a little too small.

Dividing by $N-1$ instead of N compensates for this effect. To see why the correction factor is $\frac{N}{N-1}$ requires a little more work.

Before beginning, establish some notation.

$E(\ldots)$ is the "expectation value" of ...

For example $E(x_i) = \mu$

We will make use of $E((x_i - \mu)^2) = Var(x_i)$

where $Var(x_i)$ is the variance of x_i.

$$Var(x_i) = \sigma^2$$

In addition, $E((\bar{x} - \mu)^2) = Var(\bar{x})$

$Var(\bar{x})$ is variance of $\bar{x}$

$$Var(\bar{x}) = \frac{\sigma^2}{N}$$

this result (discussed on next page) is consistent with the expectation that our determination of $\bar{x}$ should improve as N increases.

Recall from PHYS 231, we demonstrated that if we take 1000 meas of a quantity & find $\bar{x}$ & s & then we take 10,000 meas of the same quantity we find same $\bar{x}$ & s. i.e. $\bar{x}$ & s of a distribution determined by experimental setup & not by the no. of measurements.

However, expect our estimate of $\bar{x}$ to improve as N increases. We demonstrated that if we repeat N meas. of x many many times that the dist. of the determined $\bar{x}$'s has width $\sqrt{\frac{\sigma^2}{N}} \Rightarrow \mathrm{Var}(\bar{x}) = \frac{\sigma^2}{N}$

Now, let's determine $E\left(\sum [x_i - \bar{x}]^2\right)$

$$= E\left(\sum \left[(x_i-\mu)-(\bar{x}-\mu)\right]^2\right) \quad \text{(add \& subtract } \mu\text{)}$$

$$= E\left(\sum \left[(x_i-\mu)^2 - 2(x_i-\mu)(\bar{x}-\mu) + (\bar{x}-\mu)^2\right]\right)$$

$$= E\left(\sum (x_i-\mu)^2 - 2(\bar{x}-\mu)\sum(x_i-\mu) + (\bar{x}-\mu)^2 \underbrace{\sum 1}_{N}\right)$$

$$\therefore E\left(\sum [x_i-\bar{x}]^2\right)$$

$$= E\left(\sum (x_i-\mu)^2 - 2(\bar{x}-\mu)\left[\underbrace{\sum x_i}_{N\bar{x}} - \mu \underbrace{\sum 1}_{N}\right] + N(\bar{x}-\mu)^2\right)$$

$$= E\left(\sum (x_i-\mu)^2 \underbrace{- 2(\bar{x}-\mu)N(\bar{x}-\mu) + N(\bar{x}-\mu)^2}_{-N(\bar{x}-\mu)^2}\right)$$

$$= E\left(\sum (x_i-\mu)^2 - N(\bar{x}-\mu)^2\right)$$

$$= E\left(\sum (x_i-\mu)^2\right) - N\,E\left((\bar{x}-\mu)^2\right)$$

$$= \sum \underbrace{E\left((x_i-\mu)^2\right)}_{\mathrm{Var}(x_i) = \sigma^2} - N \underbrace{E\left((\bar{x}-\mu)^2\right)}_{\mathrm{Var}(\bar{x}) = \frac{\sigma^2}{N}}$$

$$\therefore E\left(\sum (x_i-\bar{x})^2\right) = \sum \sigma^2 - \sigma^2 = N\sigma^2 - \sigma^2$$

$$= \sigma^2(N-1)$$

$$\therefore E\left(\underbrace{\frac{1}{N}\sum(x_i-\bar{x})^2}\right) = \frac{N-1}{N}\sigma^2$$

this expression <u>underestimates</u> σ^2 by a factor of $\frac{N-1}{N}$

However,

$$E\left(\underbrace{\frac{1}{N-1}\sum(x_i-\bar{x})^2}_{s^2 \text{ is a good estimator of } \sigma^2}\right) = \frac{N-1}{N-1}\sigma^2 = \sigma^2$$

$\therefore$ standard deviation of sample distribution given by

$$\boxed{s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

& is good estimator of the standard deviation of parent dist'n σ^2.