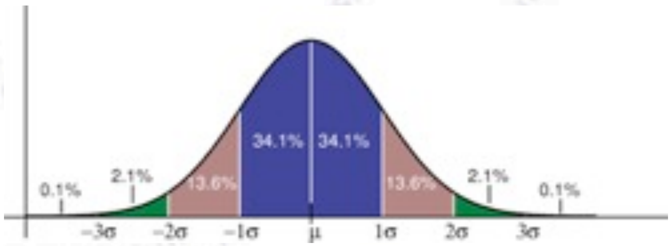
Applied Statistics

Mean and Width






Troels C. Petersen (NBI)



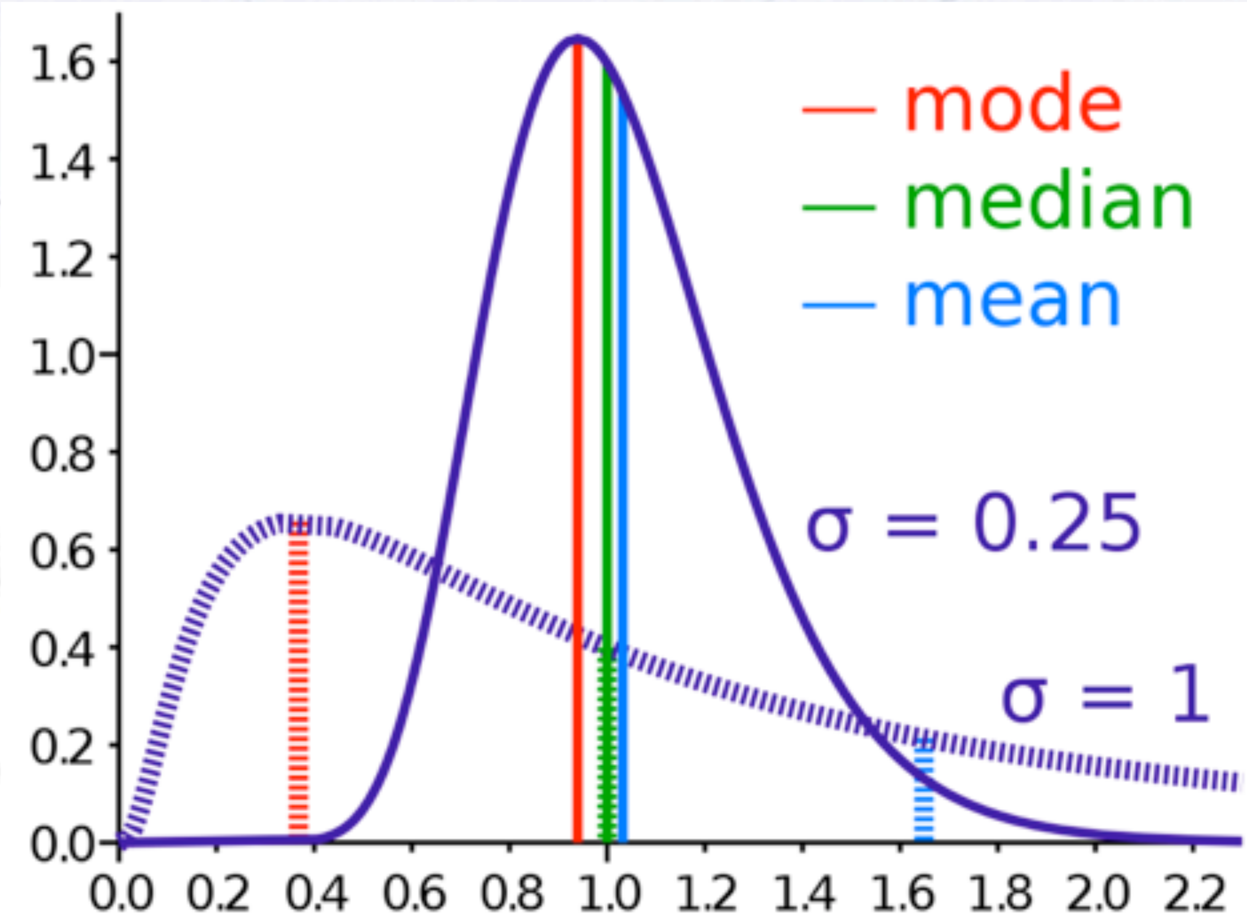
"Statistics is merely a quantization of common sense"

Defining the mean

There are several ways of defining "a typical" value from a dataset:

a) Arithmetic mean b) Mode (most probably) c) Median (half below, half above)

d) Geometric mean e) Harmonic mean f) Truncated mean (robustness)



Mean and Width

It turns out, that the best estimator for the **mean** is (as you all know):

$$\hat{\mu} = \frac{1}{N}\sum_i x_i = \bar{x}$$

For the **width** of the distribution (a.k.a. **standard deviation** or **RMS**) it is:

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_i (x_i - \mu)^2}$$

Note the “hat”, which means “estimator”. It is sometimes dropped...

Mean and Width

It turns out, that the best estimator for the **mean** is (as you all know):

$$\hat{\mu} = \frac{1}{N} \sum_i x_i = \bar{x}$$

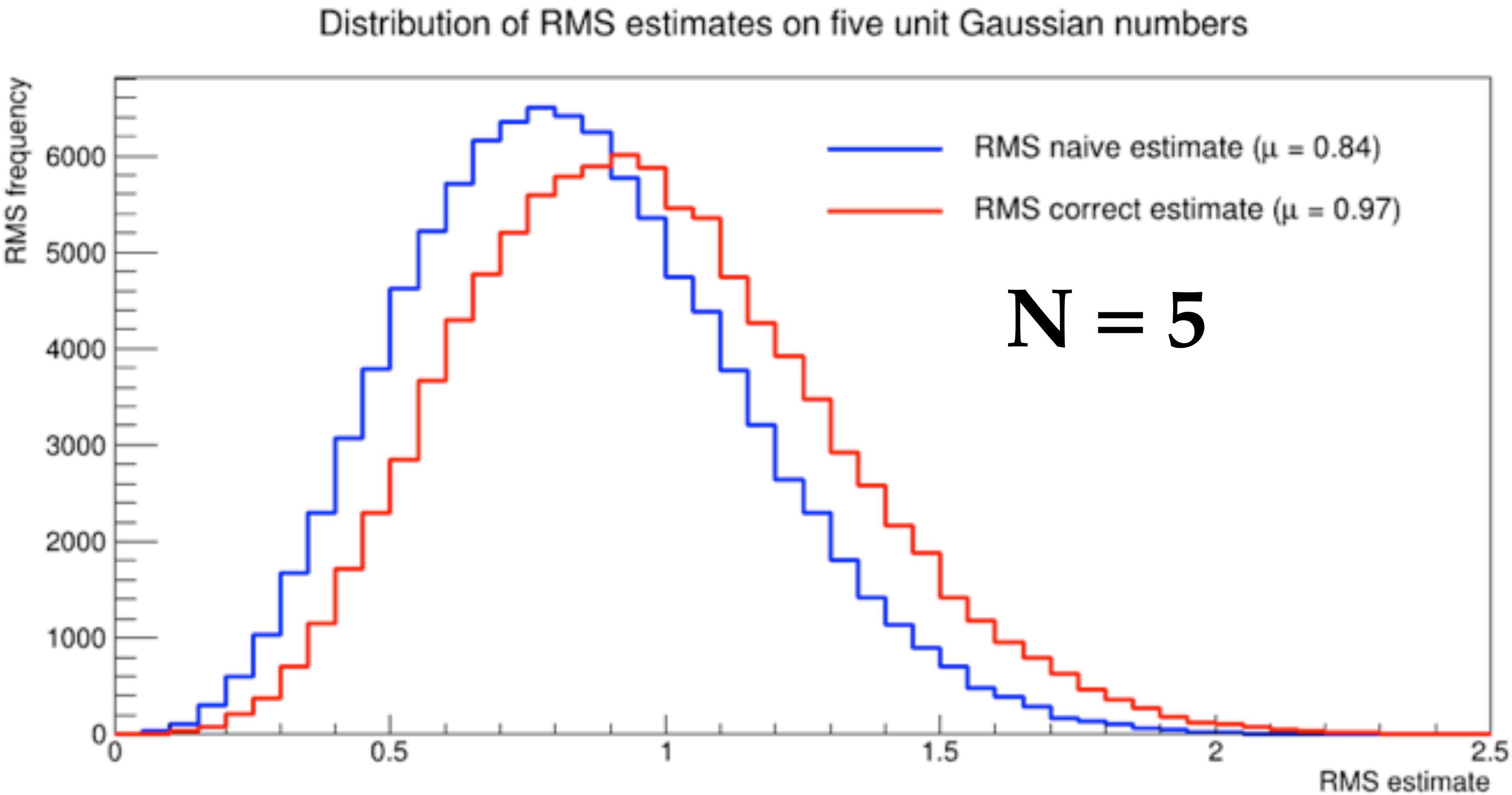
For the **width** of the distribution (a.k.a. **standard deviation** or **RMS**) it is:

$$\hat{s} = \sqrt{\frac{1}{N-1} \sum_i (x_i - \bar{x})^2}$$

Note the "hat", which means "estimator". It is sometimes dropped...

How incorrect is the naive RMS?

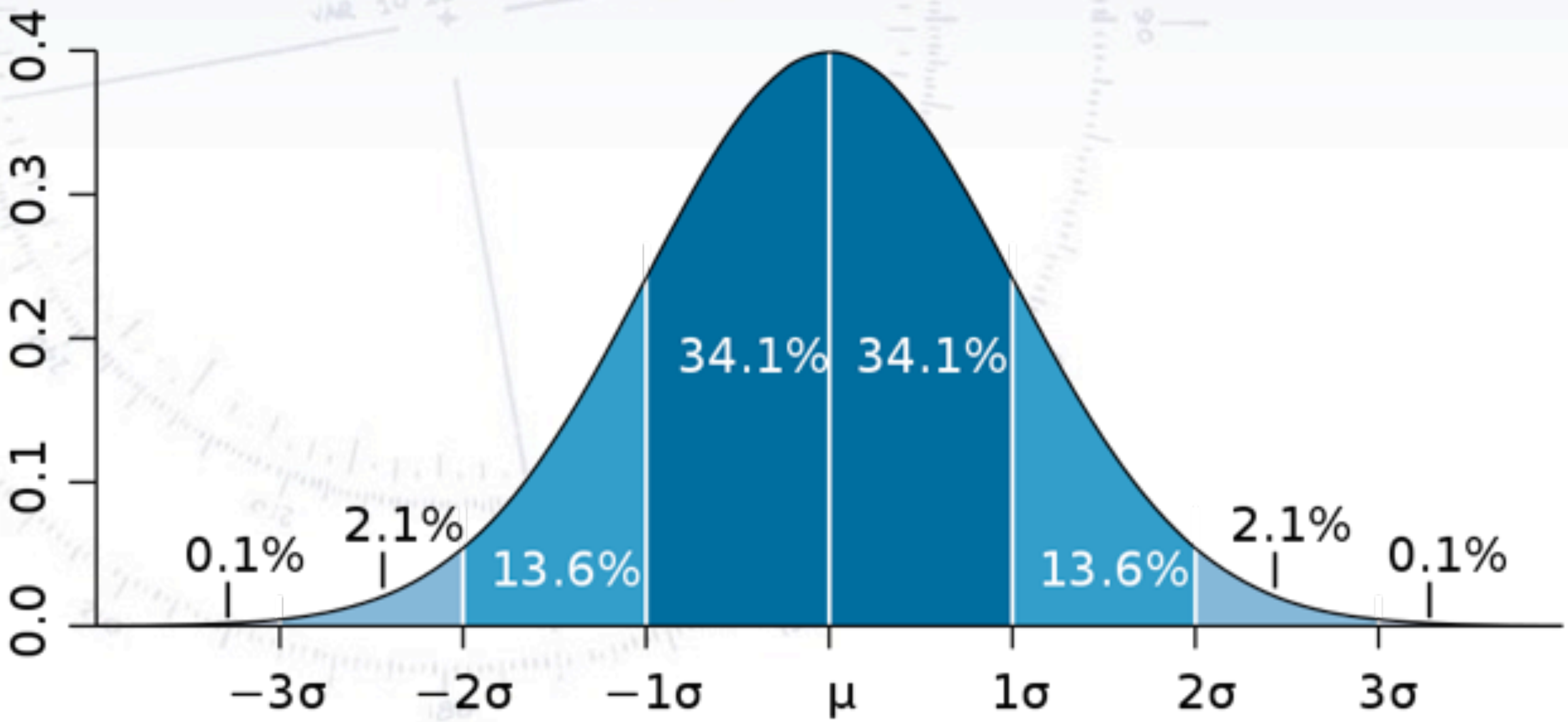
Such questions can most easily be answered by a small simulation:



So, the "naive" RMS underestimates the uncertainty a bit...

Relation between RMS and Gaussian width...

When a distribution is Gaussian, the RMS corresponds to the Gaussian width σ:



Mean and Width

What is the **uncertainty on the mean?** And how quickly does it improve with more data?

$$\hat{\sigma}_\mu = \hat{\sigma}/\sqrt{N}$$

<u>Example:</u>

Cavendish Experiment

(measurement of Earth's density)

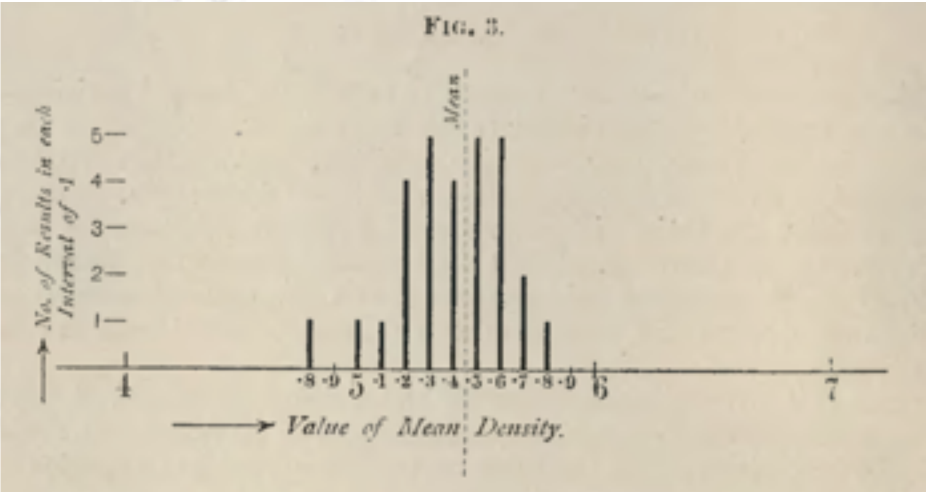
N = 29

mu = 5.42

sigma = 0.333

sigma(mu) = 0.06

Earth density = 5.42 ± 0.06



Mean and Width

What is the **uncertainty on the mean?** And how quickly does it improve with more data?

Please, commit to memory now!

$$\hat{\sigma}_\mu = \hat{\sigma}/\sqrt{N}$$

<u>Example:</u>

Cavendish Experiment

(measurement of Earth's density)

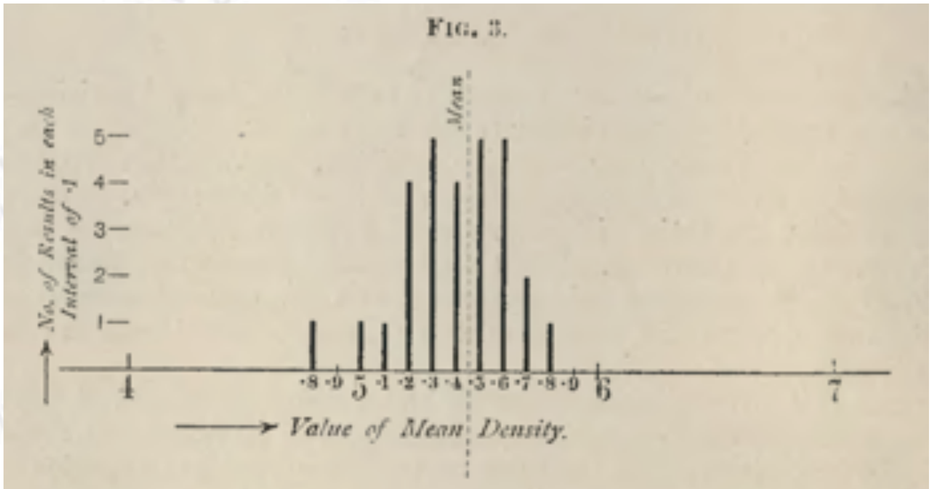
N = 29

mu = 5.42

sigma = 0.333

sigma(mu) = 0.06

Earth density = 5.42 ± 0.06



Mean and Width

The calculation of the mean and RMS is often simplified (especially in programs) by the following classic calculation/reduction:

$$V(x) = \sigma_x^2 = \frac{1}{N}\sum_i (x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$$

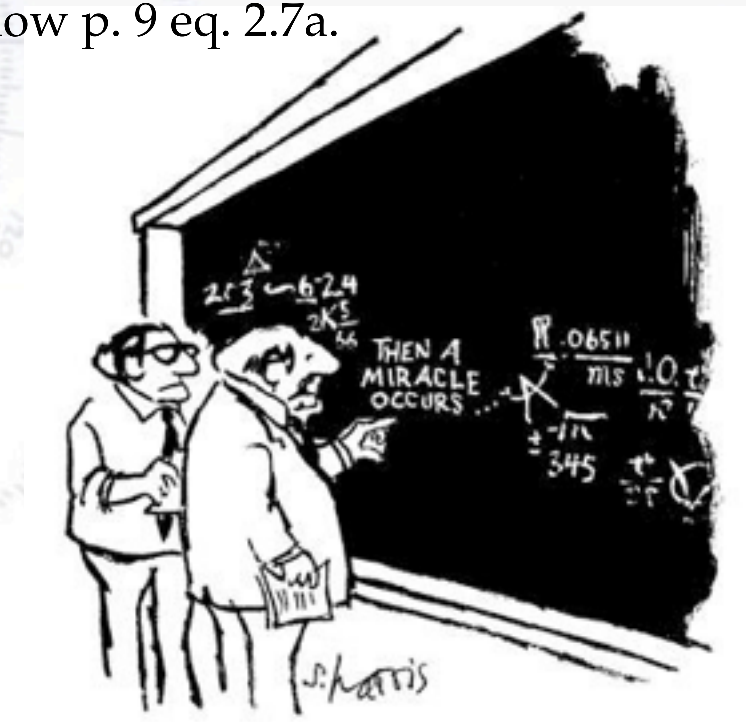
If you want to see how this is deduced, see Barlow p. 9 eq. 2.7a.

Thus, in a program, it is useful to define:

```
Sum0 = Σ 1 = N
Sum1 = Σ x
Sum2 = Σ x²
```

and then obtain:

```
Mean = Sum1 / Sum0
RMS = sqrt(Sum2/Sum0 - mean²)
σ(Mean) = RMS / sqrt(Sum0)
```



"I think you should be more explicit here in step two."

Weighted Mean

What if we are given data, which has different uncertainties?
How to average these, and what is the uncertainty on the average?

$$\hat{\mu} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}$$

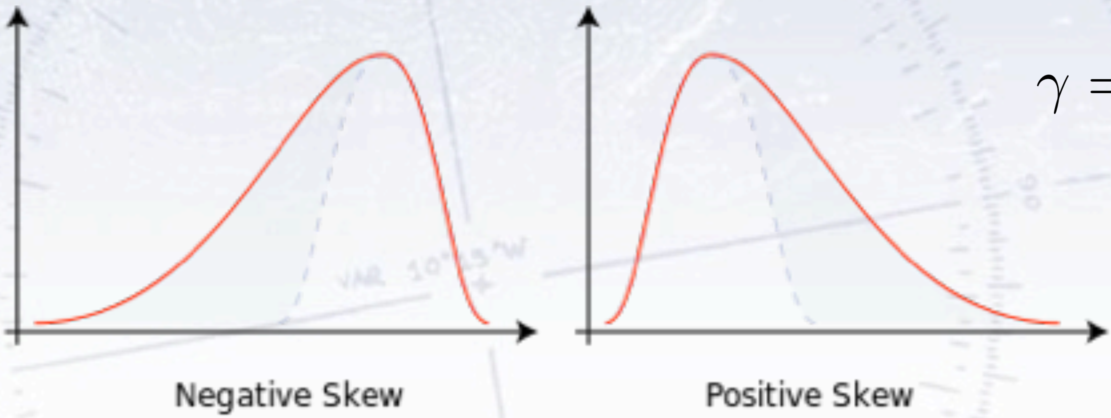
For measurements with varying uncertainty, there is no meaningful RMS!
The uncertainty on the mean is:

$$\hat{\sigma}_\mu = \sqrt{\frac{1}{\sum 1/\sigma_i^2}}$$

Can be understood intuitively, if two persons combine 1 vs. 4 measurements

Skewness and Kurtosis

Higher moments reveal something about a distributions asymmetry and tails:



$$\gamma = \frac{\frac{1}{N}\sum_i (x_i - \bar{x})^3}{(\frac{1}{N}\sum_i (x_i - \bar{x})^2)^3/2}$$

$$\kappa = \frac{\frac{1}{N}\sum_i (x_i - \bar{x})^4}{(\frac{1}{N}\sum_i (x_i - \bar{x})^2)^2} - 3$$