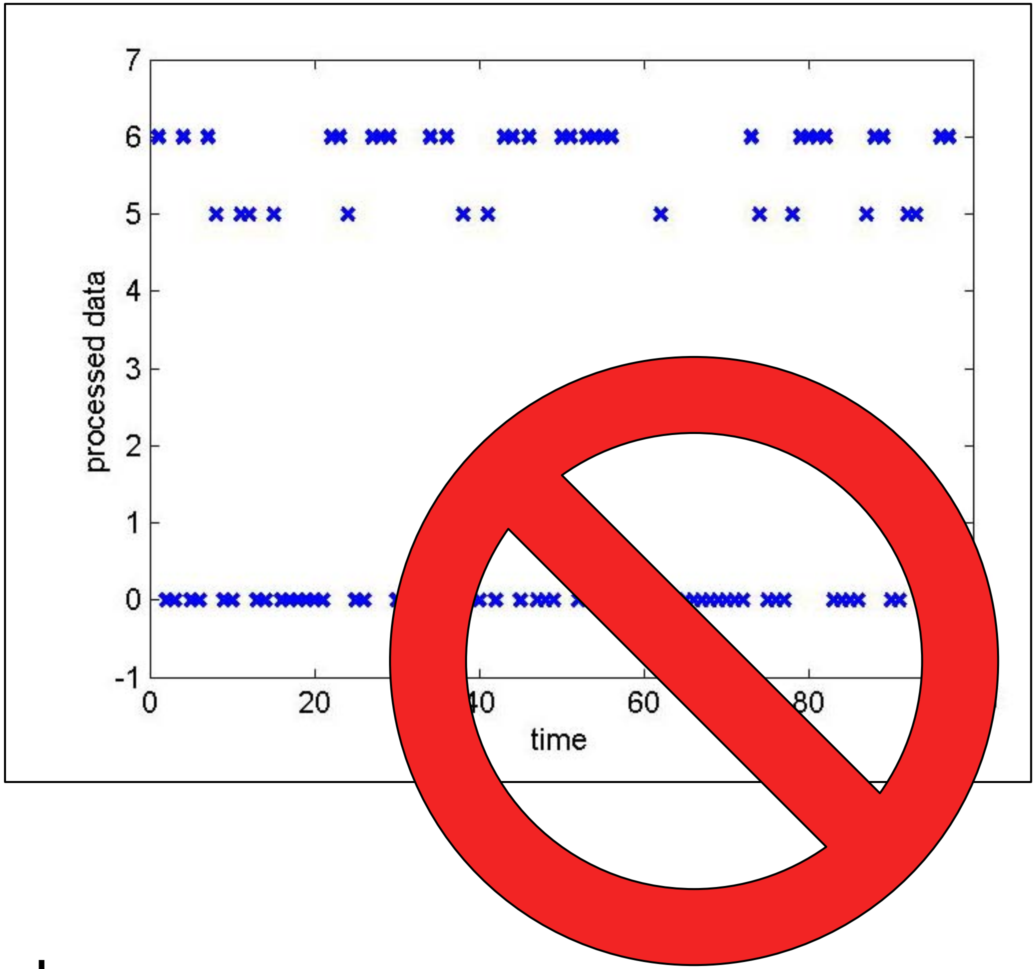
Data and Time Series Analysis Techniques

Ask Yourself:

- *Does the work stand up to scrutiny?*
  - Use of controls
  - Calibration
  - Data quality
  - Data processing
  - Documentation and record-keeping!



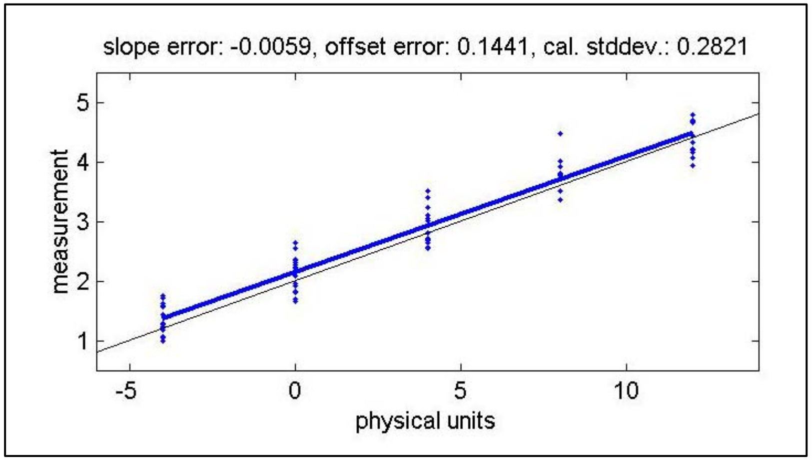
Controls

Did you really measure what you thought?

- *Rat Maze: Is the maze acoustically navigable? (R. Feynman)*
- *Mass Spectroscopy: When you put in a sample of known composition, are the other bins clean?*
- *When measuring electrical resistance, touch the probes together. Check a precision resistor too.*
- *Resonance in load measurement rigs?*
- *When measuring hull resistance, does zero speed give zero force?*

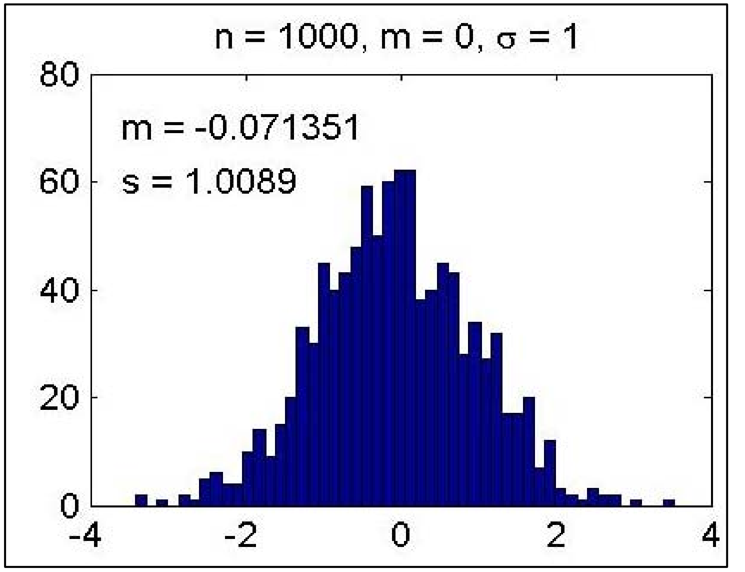
Take every opportunity to eliminate doubt!

Calibration



- More time can be spent on calibration than the rest of the experiment!
- Sensors should be calibrated and re-checked using independent references, such as:
  - *Manufacturer's specifications*
  - *Another sensor with very well-known calibration* ←→
  - *A tape measure, protractor, calipers, weights & balance, stopwatch, etc..*
- Calibration range should include the expected range in the experiment.
- Some statistics of the calibration:
  - *Precision of fit (r-value or σ)*
  - *Linearity (if applicable)*
- Understand special properties of the sensor, e.g., inherent nonlinearity, drift, PWM output

Sample Statistics



- *Sample* mean m:
- *Sample* standard dev. σ:

$$\sigma = \text{sqrt}\ [\ (\ (x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_n - m)^2\ )\ /\ (n-1)\ ]$$

- Error budgets for multiplication and addition (σA is standard deviation of A):

$$(A + \sigma A)(B + \sigma B) \sim AB + A\sigma B + B\sigma A$$

Example: $(1.0 + \sigma 0.2)(3.0 + \sigma 0.3) \sim 3.0 + \sigma 0.9$

$$(A + \sigma A) + (B + \sigma B) = A + B + \sigma(A+B)$$

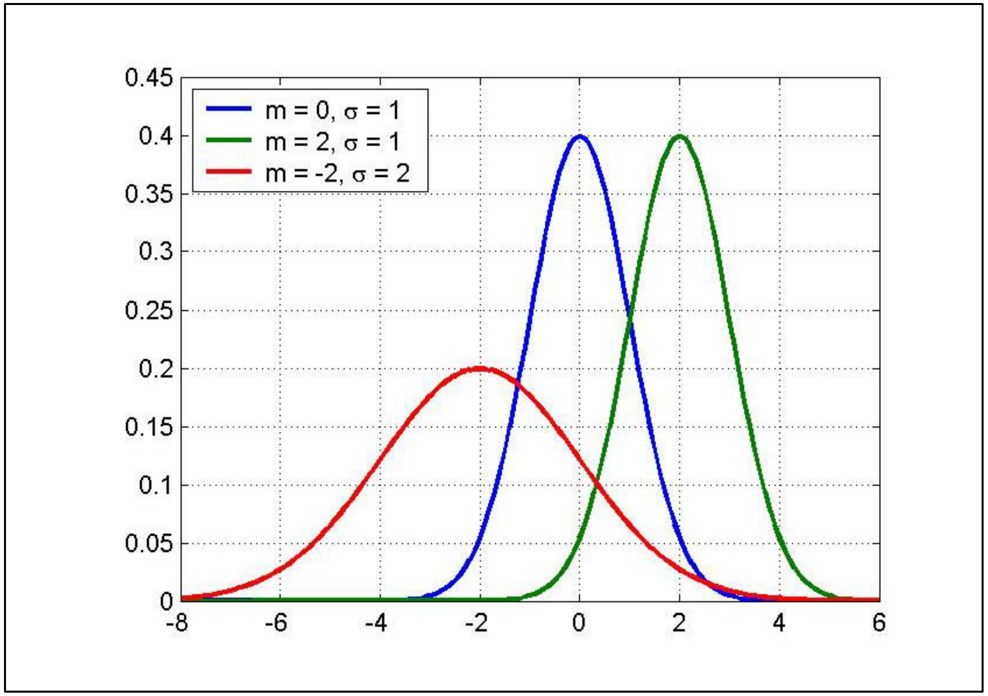
Example: $(1.0 + s0.2) + (3.0 + s0.3) = 4.0 + \sigma 0.5$

Gaussian (Normal) Distribution

Probability Density Function f(x) ~ *Histogram*

$$f(x) = \exp\left[- (x-m)^2 / 2\sigma^2 \right] / \mathrm{sqrt}(2\pi) / \sigma$$

This is the most common distribution encountered in sensors and systems.



+/- 1σ covers 68.3%

+/- 2σ covers 95.4%

+/- 3σ covers 99.7%

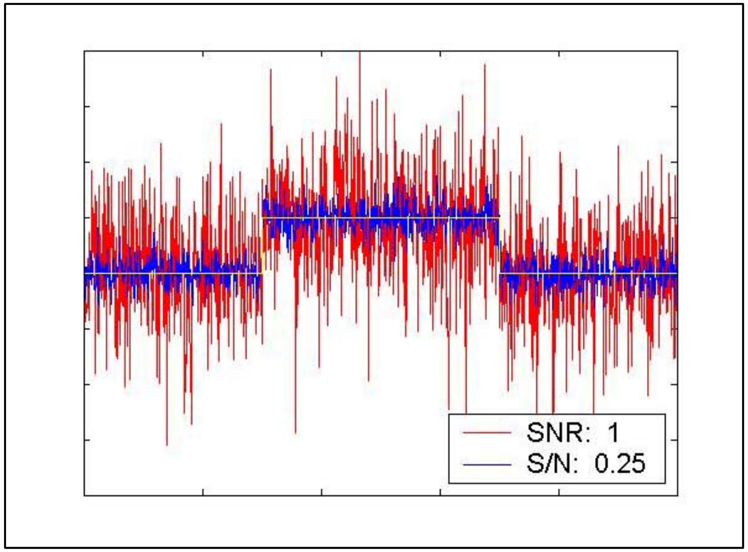
Area under f(x) is 1!

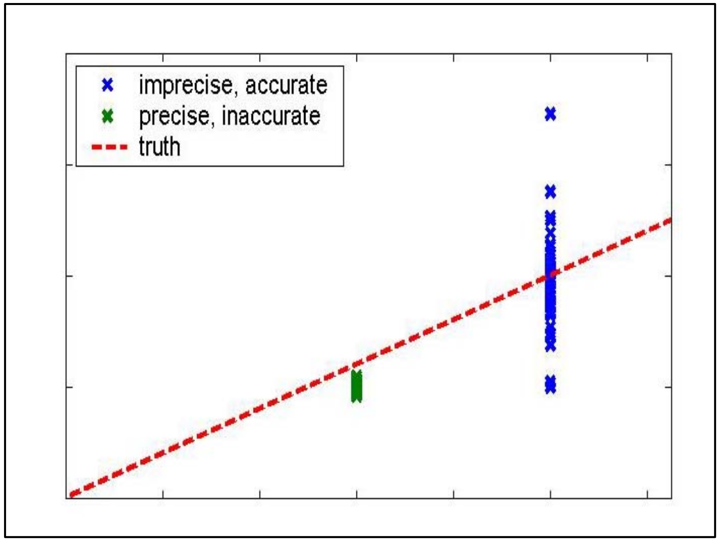
Data and Sensor Quality

- Signal-to-Noise Ratio (SNR): compares σ to the signal you want

- Repeatability/Precision: If we run the same test again, how close is the answer?

- Accuracy: Take the average of a large number of tests – is it the right value?



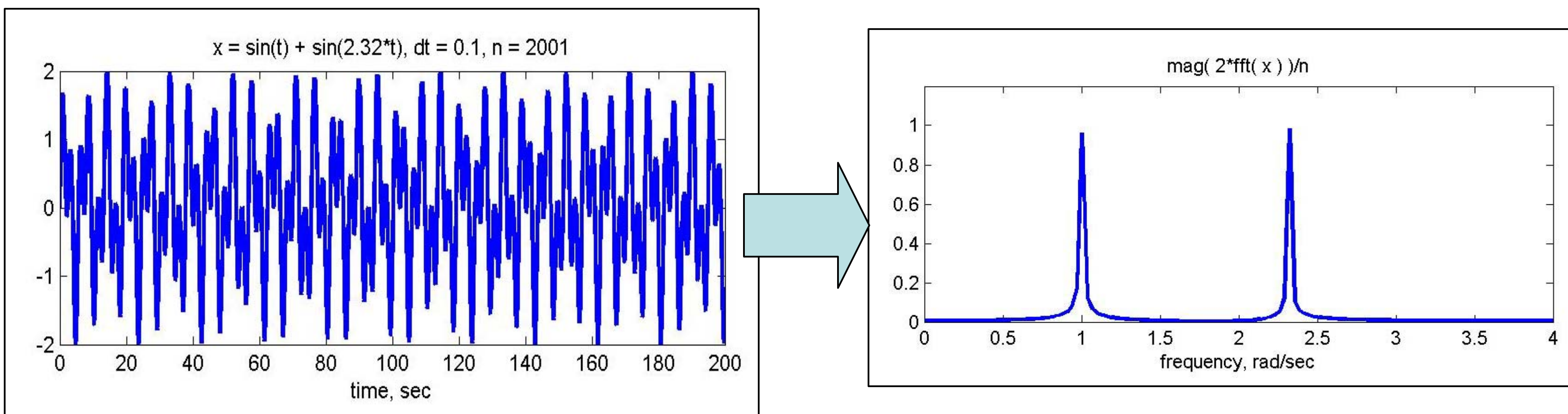


Time and Frequency Domain

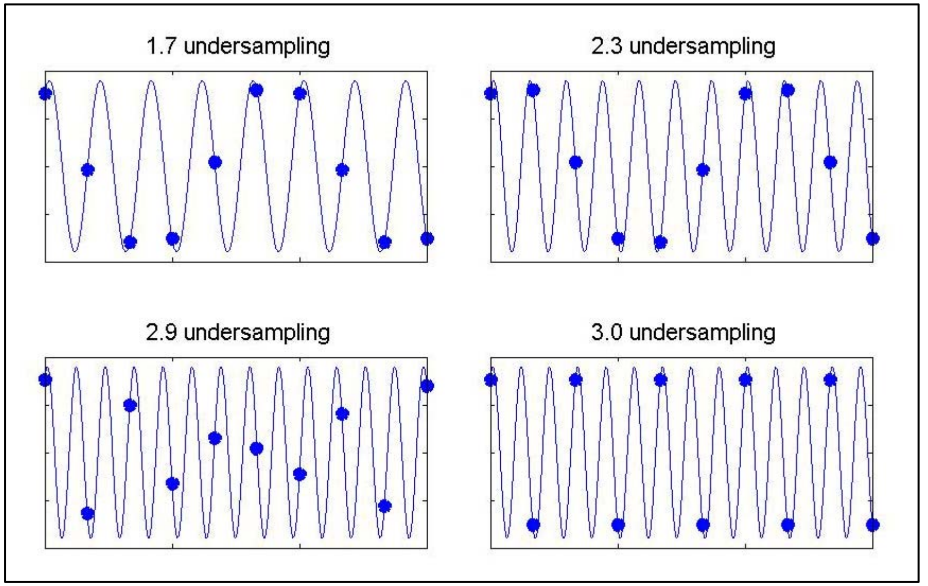
- Fourier series/transforms establish an **exact** correspondence between these domains, e.g.,

$$\begin{cases} X_m = \int_0^T \cos(2\pi\, m\, t / T)\, z(t)\, dt * 2/T, \quad m = 0,1,2,\ldots \\ Y_m = \int_0^T \sin(2\pi\, m\, t / T)\, z(t)\, dt * 2/T \end{cases}$$

$$z(t) = X_0/2 + \sum X_m \cos(2\pi\, m\, t / T) + \sum Y_m \sin(2\pi\, m\, t / T)$$

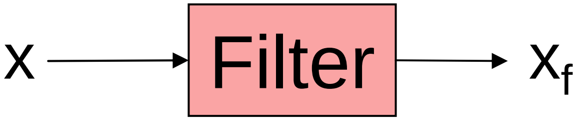


Time Resolution in Sampled Systems



- The Sampling Theorom shows that the highest frequency that can be detected by sampling at frequency $\omega_s = 2\pi/\Delta t$ is the Nyquist rate: $\omega_N = \omega_s / 2$.
- Higher frequencies than this are "aliased" to the range below the Nyquist rate, through "frequency folding." *This includes sensor noise!* ← anti-aliasing filters
- The required rate for "visual" analysis of the signal, and phase and magnitude calculation is much higher, say ten samples per cycle.

Filtering of Signals



Use good judgement!

filtering brings out trends, reduces noise

filtering obscures dynamic response

<u>Causal filtering</u>: $x_f(t)$ depends only on past measurements – appropriate for real-time implementation

Example: $x_f(t) = (1-\varepsilon)\, x_f(t-1) + \varepsilon\, x(t-1)$

<u>Acausal filtering</u>: $x_f(t)$ depends on measurements at future time – appropriate for post-processing

Example: $x_f(t) = [\, x(t+1) + x(t) + x(t-1)\,] / 3$

A first-order filter transfer function in the freq. domain (where $j\omega$ is the derivative operator):

$$x_f(j\omega) / x(j\omega) = \lambda / (j\omega + \lambda)$$

At low ω, this is approximately 1 (that is, λ/λ)

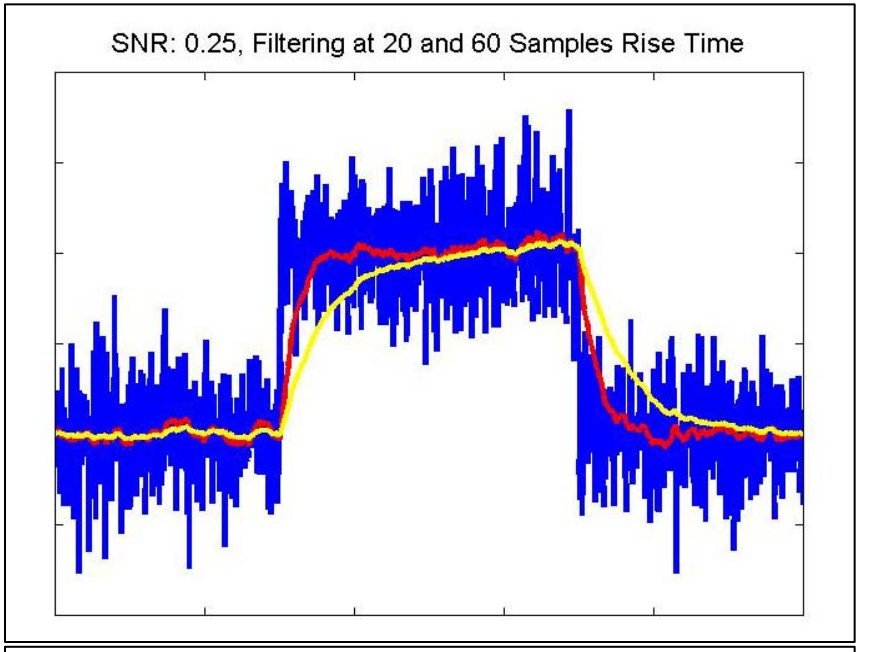
At high ω, this goes to 0 magnitude, with 90 degrees phase lag ($\lambda/j\omega = -j\lambda/\omega$)

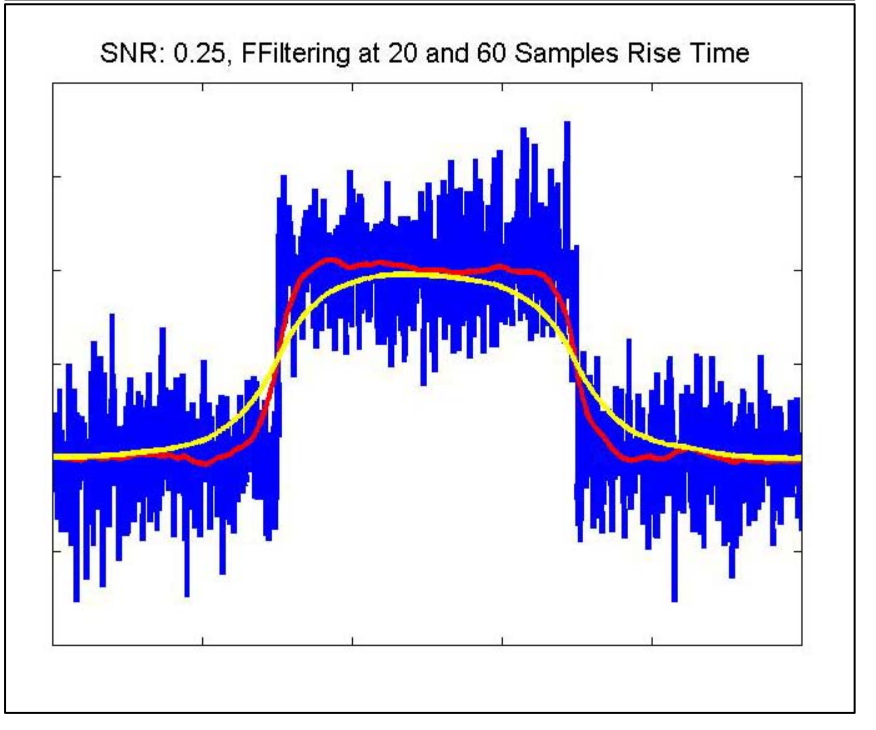
Time domain equivalent:

$$dx_f/dt = \lambda (x - x_f)$$

In discrete time, try

$$x_f(k) = (1-\lambda\Delta t)\, x_f(k-1) + \lambda\Delta t\; x(k-1)$$





- BUT linear filters will not handle outliers very well!
- First defense against outliers: find out their origin and eliminate them at the beginning!
- <u>Detection</u>: Exceeding a known, fixed bound, or an impossible deviation from previous values. *Example: vehicle speed >> the possible value given thrust level and prior tests.*
- Second defense: set data to NaN (or equivalent), so it won't be used in calculations.
- Third defense: try to fill in.

  Example:

  if abs(x(k) – x(k-1)) > MX,

  x(k) = x(k-1) ;

  end;

  → Limited usefulness!