Experimental Errors and Uncertainty: An Introduction

Prepared for students in AE 3051

by J. M. Seitzman
adapted from material by J. Craig

Outline

- Errors and types of error
- Statistic/probability: confidence levels
- Uncertainty analysis

Experimental Error

- **Error**: all measurements have some uncertainty

$$\text{error} = \varepsilon = u_{meas} - u_{exact}$$

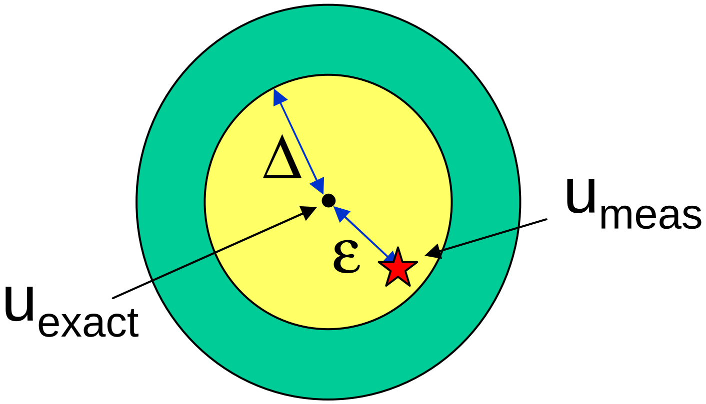
- **Objectives**

1. **Minimize** error so that

   $-\Delta \le \varepsilon \le +\Delta$ *within some uncertainty (statistical confidence)*

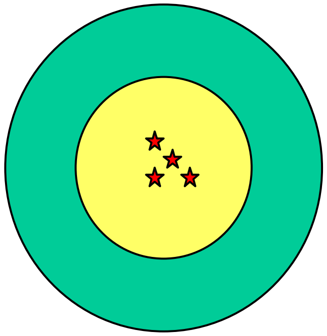
   or $u_{meas} - \Delta \le u_{exact} \le u_{meas} + \Delta$

2. **Estimate** error **(uncertainty)** to determine reliability, meaningfulness of data

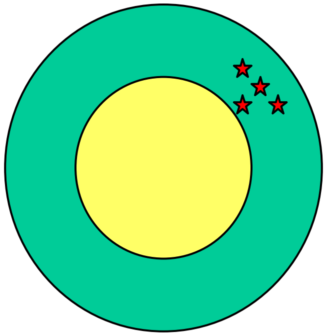


Accuracy and Precision

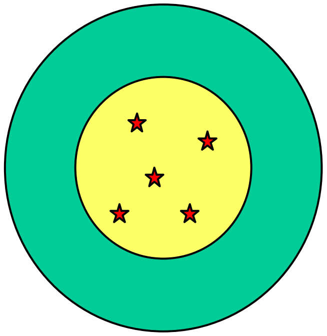
- **Accuracy:** also called ***systematic or bias error***
  - – denotes something repeatably “wrong” with the measurement or experiment
- **Precision:** also called ***random error or noise***
  - – denotes errors that change randomly each time you try to repeat experiment



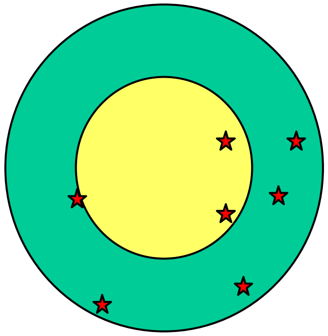
Good Accuracy
Good Precision



Good Precision
Poor Accuracy
(can calibrate)



Good Accuracy
Poor Precision
(can average)



Poor Accuracy
Poor Precision

Other Related Terms

- **Sensitivity**
  - Change in a **measurement device**’s output for a unit change in the measured (input) quantity, e.g., volts/Torr for the Baratron
- **Resolution**
  - Smallest increment of change in a system or property that a **measurement device** can reliably capture
- **Dynamic Range**
  - Maximum output of a **measurement device** divided by its resolution (or minimum measureable signal)

Sensitivity
= 5000 psi/270°
= 18.5 psi/degree

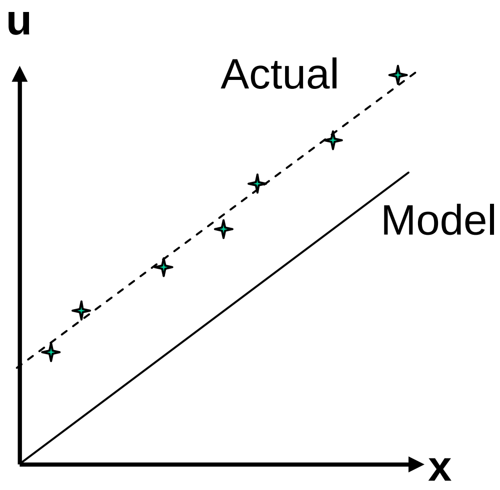
Resolution = 50 psi
Dyn. Range= 5000/50
= 100

Accuracy/Systematic Errors

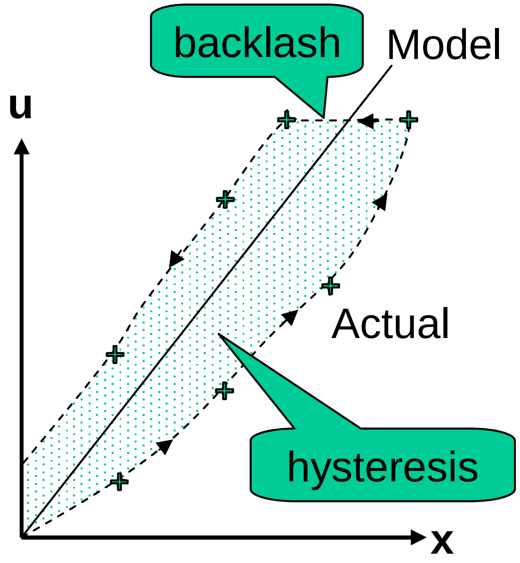
- **Sources**
  - **Measuring system** errors
    - difference between *model* of measuring system and *reality*
    - could be corrected, e.g., with better model of measurement
  - **Measured system** “errors”
    - influence of uncontrolled or unaccounted for variables in the experiment
    - the measured data may be “correct”, but may lead to an incorrect model of the object/process being studied
  - **Blunders**
    - human errors - misunderstandings

Some Systematic Measurement Errors

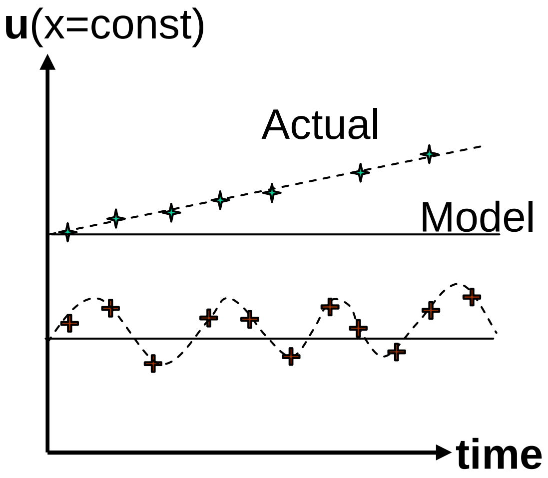
- u=u(x)



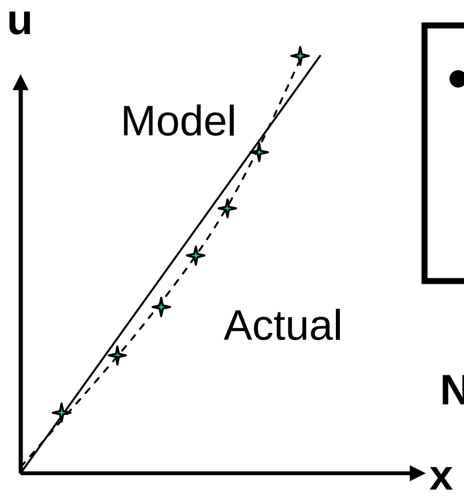
Nonzero offset - Background



Backlash & Hysteresis



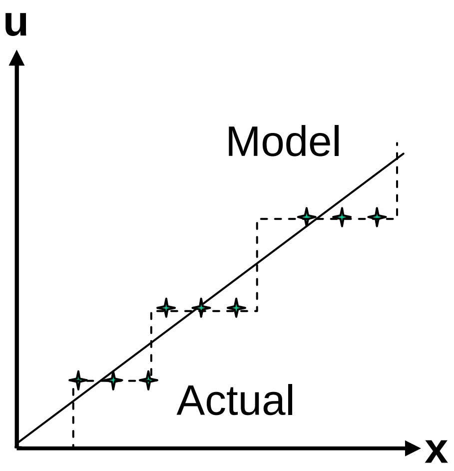
Drift (e.g., offset changing in systematic way with time)



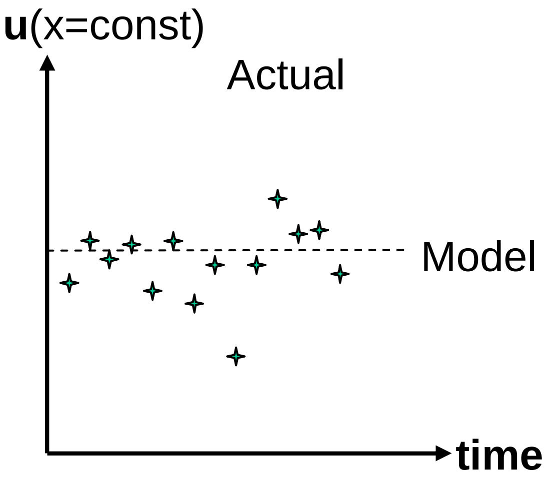
Nonlinearity

- Systematic errors can be eliminated/removed if they are known

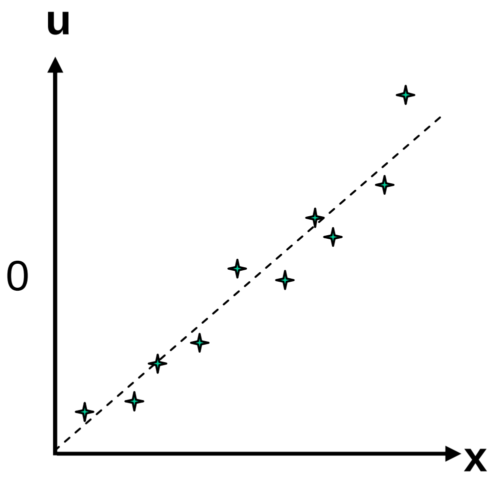
Quantization Error (digitized data – impacts resolution)



Some Random Measurement Errors



Background "Noise" (offset changing randomly with time)

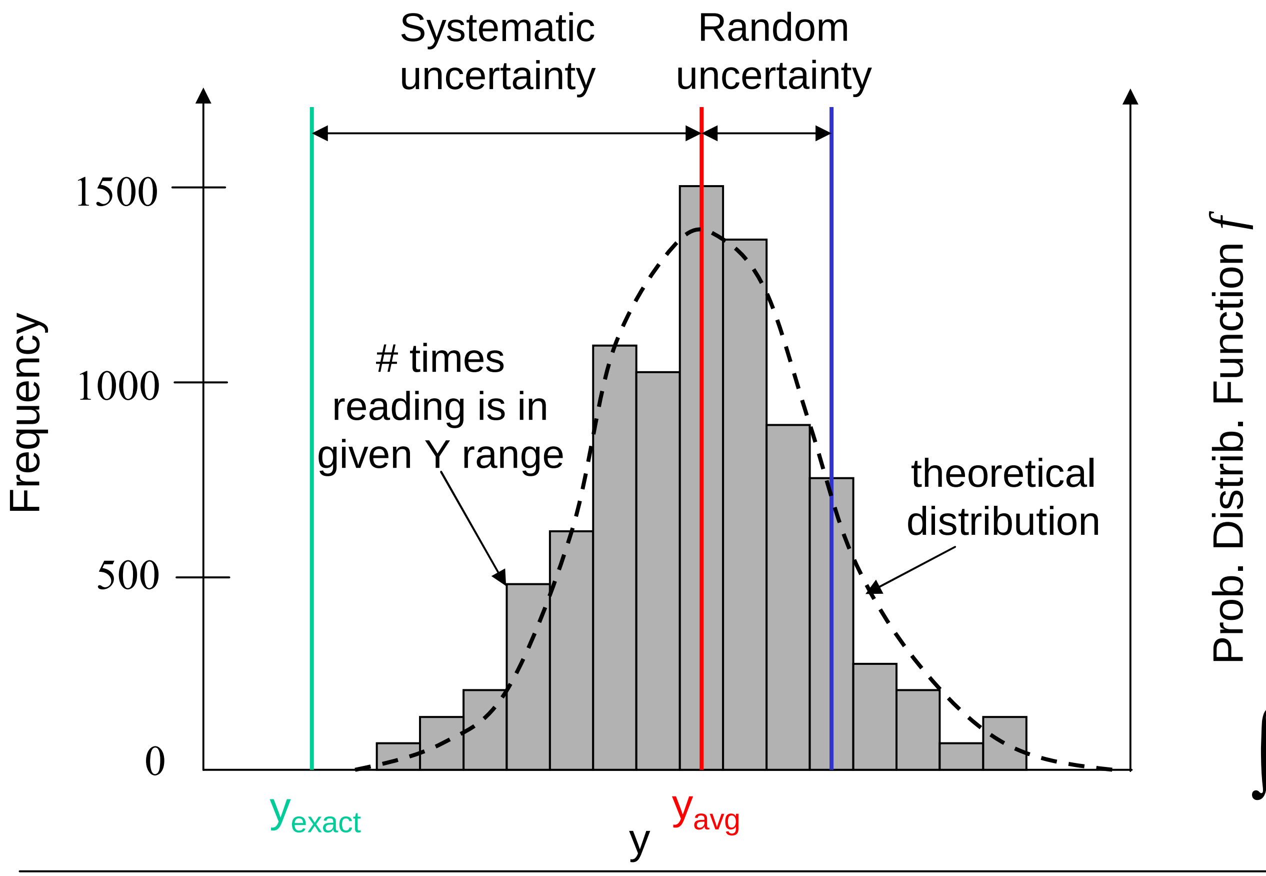


Detector "Noise" (random change in sensitivity of device)

- After data acquired, nearly impossible to separate random error (noise) sources
- Examine random error with statistical methods

Probability Distributions

- When we make measurements (i.e., take samples) of a system a number of times, we will get a distribution of results



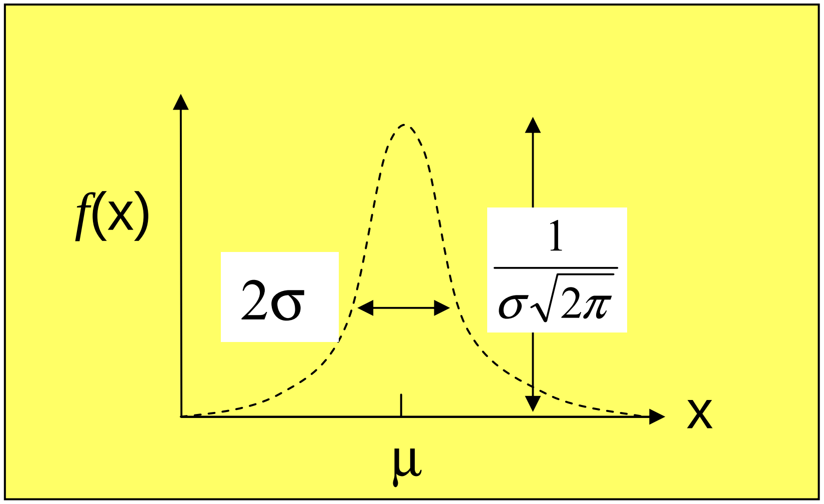
- We might even make just one measurement (sample) of a system that has a distribution of possible states

$$\int_{-\infty}^{\infty} f(y)dy = 1$$

Statistics and Probability

- Since we can not make an infinite number of measurements to determine the true probability distribution
  - we use statistics to make estimates based on assumed distribution function
- Some useful distribution functions
  - normal (Gaussian)
  - student’s t
  - log normal
  - exponential

Normal/Gaussian Probability Distribution



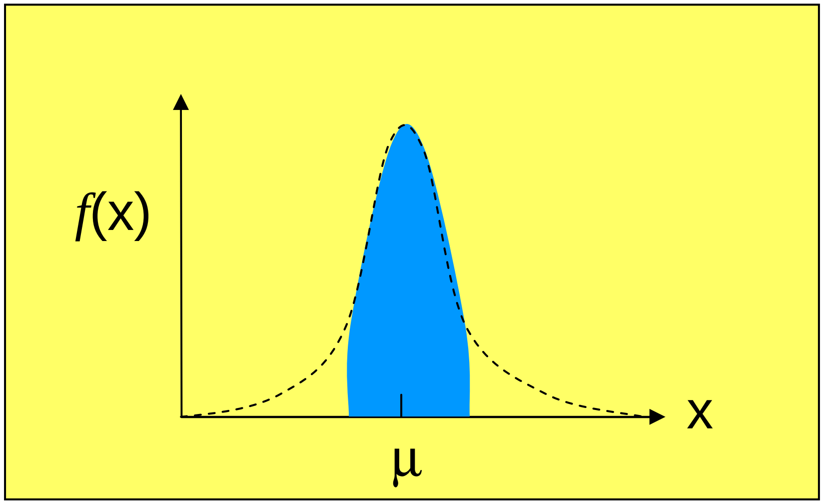
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

μ = mean
σ^2 = variance
σ = standard deviation

- Commonly used when measurements/measurement system:
  - made up from **many independent systems,** each with any kind of distribution
  - **# samples taken is very large** (e.g., sample means)
  - more…

Normal Distributions – Probability Range

- What fraction of values (combined probability) lie within given range from mean for a normal distribution?



One Sigma: $$\Pr\text{ob}(\mu-\sigma\le\mu\le\mu+\sigma)=\int_{\mu-\sigma}^{\mu+\sigma}f(x)dx=0.683$$

Two Sigma: $$\Pr\text{ob}(\mu-2\sigma\le\mu\le\mu+2\sigma)=\int_{\mu-2\sigma}^{\mu+2\sigma}f(x)dx=0.954$$

Three Sigma: $$\Pr\text{ob}(\mu-3\sigma\le\mu\le\mu+3\sigma)=0.997$$

Sample Statistics

- What if μ and σ are unknown (as is often the case)?
  - use estimates from measurements, $\bar{x}$ and s_x

$$\mu \cong \bar{x};\ \sigma \cong s_x$$

$$\mathit{Sample\ mean} = \bar{x} = \sum_{i=1}^{N} x_i \Big/ N$$

Mean Square

Square of mean

$$\mathit{Sample\ variance} = s_x^{\ 2} = \sum_{i=1}^{N} (x_i - \bar{x})^2 \Big/ N-1 = \left(\frac{\sum_{i=1}^{N} x_i^{\ 2}}{N} - \bar{x}^2 \right) \frac{N}{N-1}$$

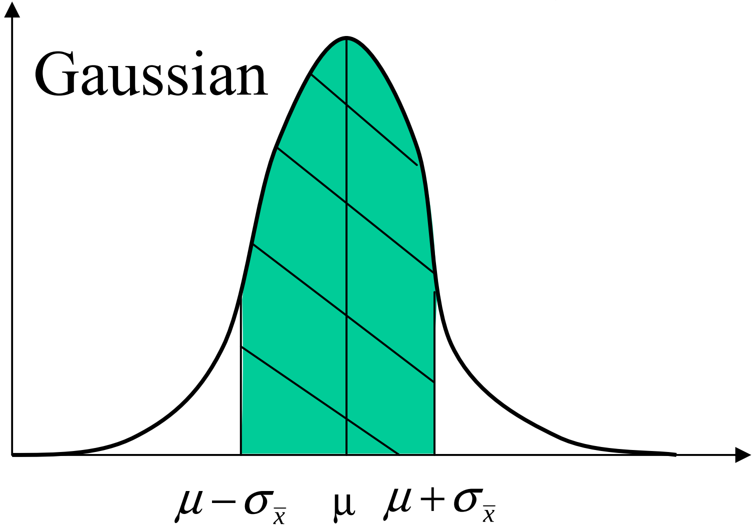
  - use (N-1) for s_x^2 because we have N independent x_i but if also know x_{mean} then only need to know (N-1) x_i to compute last remaining x_i
    - ⇒only (N-1) "degrees of freedom" for this calculation

Uncertainty Estimates

- Question: If one takes N (large) readings and computes $\bar{x}$, how confident can you be that the average is really close to the true mean (μ)?
- Confidence intervals are way to describe this

Prob = c% that μ *lies in shaded area defined by* $\bar{x} = \mu \pm a_c \sigma_{\bar{x}}$

variance of $\bar{x}$ distrib.



$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \approx \frac{s_x}{\sqrt{N}}$$

$a_c = ?$

Confidence Levels

- For normal distribution

Error Level Name	Error Level	Prob. that Error is Smaller	Prob. that Error is Larger
Probable Error	± 0.67 σ	50%	1:2
One Sigma	± σ	68%	~ 1:3
90% error	± 1.65 σ	90%	1:10
“Two” Sigma	± 1.96 σ	95%	1:20
Three Sigma	± 3 σ	99.7%	1:370
Maximum Error	± 3.29 σ	99.9	1:1000
Four Sigma	± 4 σ	99.994%	1:16000
Six Sigma	± 6 σ	99.9999999%	1:1.01e9

Six Sigma is used for many electronic manufacturing processes

back

$Example:$

$$c = 95\% \Rightarrow a_c = 1.96, so\ \bar{x} = \mu \pm 1.96\sigma_{\bar{x}}$$

$$With\ 95\%\ confidence, \mu\ will\ \ fall\ within\ \bar{x} \pm 1.96\sigma_{\bar{x}}$$

Confidence Intervals: Examples

- Example 1: Find 95% confidence limits for $\bar{x}$ = 75 psi when s_x=8.3 psi for N = 50 samples

$$Answer:\ \bar{x} \pm 1.96\, S_x \big/ \sqrt{N} = 75 \pm 1.96 \frac{8.3}{\sqrt{50}} = 75 \pm 2.3\ psi$$

- Example 2: Find 99.9% confidence limits for previous case (use previous table):

$$Answer:\ \bar{x} \pm 3.29\, S_x \big/ \sqrt{N} = 75 \pm 3.29 \frac{8.3}{\sqrt{50}} = 75 \pm 3.9\ psi$$

- Example 3: How many samples (N) are required to assume that mean is within ± 5% with 95% confidence?

$$Answer:\ 75 \pm 1.96 \frac{8.3}{\sqrt{N}} = 75 \pm \left(75 \times 0.05\right) psi$$

$$or:\ N = 18.8 \approx 19\ samples$$

Remarks on Sample Size

- Above strictly works only when
  - N=large
  - or when σ is known (don't have to estimate from s_x)
- When N=small, then we should not approximate σ by s_x or use Gaussian statistics,
  - leads to replacing the normal distribution with the student's t-distribution, a subject beyond this introduction
  - **for the purposes of this class**, you may assume $N \geq 10$ is large

Analysis of Combined Uncertainties

- Many times experimental results are the result of several independent measurements combined using a theoretical formula (e.g., $\dot{m} = \rho u A = \frac{p}{RT} u \pi D^2 / 4$ for mass flowrate through a pipe).
- How do uncertainties in each variable contribute to whole?

If $y=y(x_1, x_2, \ldots x_N)$ is a linear function, a statistical theorem states that:

$$\sigma_y = \left[\left(\frac{\partial y}{\partial x_1} \sigma_{x_1} \right)^2 + \left(\frac{\partial y}{\partial x_2} \sigma_{x_2} \right)^2 + \cdots \left(\frac{\partial y}{\partial x_N} \sigma_{x_N} \right)^2 \right]^{1/2}$$

For uncertainties, u_i, that are small compared to x_i we can use a Taylor Series expansion in u_i:

$$y(x_1 + u_1, x_2 + u_2, \ldots x_N + u_N) = y(x_1, x_2, \ldots x_N) + \frac{\partial y}{\partial x_1} u_1 + \frac{\partial y}{\partial x_2} u_2 + \ldots \frac{\partial y}{\partial x_N} u_N$$

so that y is now a linear function of the uncertainties. Applying this in the first equation yields:

$$u_y = \left[\left(\frac{\partial y}{\partial x_1} u_1 \right)^2 + \left(\frac{\partial y}{\partial x_2} u_2 \right)^2 + \cdots \left(\frac{\partial y}{\partial x_N} u_N \right)^2 \right]^{1/2}$$

Combining Bias and Precision Uncertainties

- We noted earlier that errors in each measured variable (y_i) will include both bias (systematic) and precision (random) components
- These can usually be treated as independent and therefore the uncertainties for each can be combined into a total:

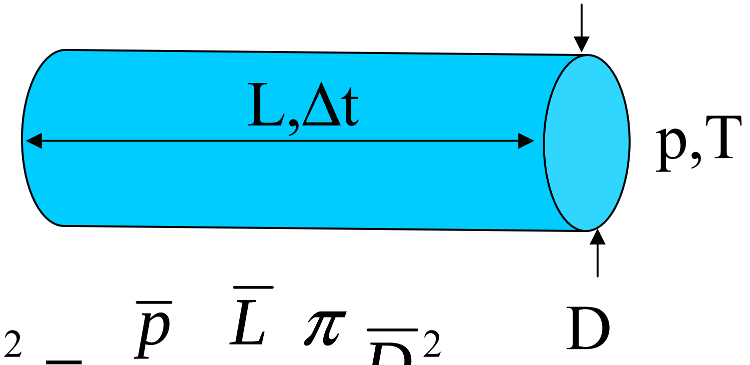
$$u_{y_i-total} = \left[u_{y_i-precision}{}^2 + u_{y_i-bias}{}^2\right]^{1/2}$$

  - note: if they are not independent, combining in other ways may be necessary
- Including confidence intervals

$$\boxed{u_{y_i-total} = \left[\left(a_c \sigma_{\bar{y}_i}\right)^2 + u_{y_i-bias}{}^2\right]^{1/2}}$$

Example: Uncertainty Calculation

- Consider measurement of mass flowrate through a round pipe



$$\overline{\dot{m}} = \rho v A = \frac{p}{RT} v \frac{\pi}{4} D^2 = \frac{\overline{p}}{R\overline{T}} \frac{\overline{L}}{\overline{\Delta t}} \frac{\pi}{4} \overline{D}^2$$

- The fractional uncertainty in $\dot{m}$ can then be shown to be:*

$$\frac{u_{\dot{m}}}{\dot{m}} = \left[\left(\frac{u_p}{p} \right)^2 + \left(\frac{u_T}{T} \right)^2 + \left(\frac{u_L}{L} \right)^2 + \left(\frac{u_{\Delta t}}{\Delta t} \right)^2 + \left(2 \frac{u_D}{D} \right)^2 \right]^{1/2}$$

* e.g., $\left(\frac{\partial \dot{m}}{\partial p} \right) \frac{u_p}{\dot{m}} = \left(\frac{1}{RT} \frac{L}{\Delta t} \frac{\pi}{4} D^2 \right) \frac{u_p}{\dot{m}} = \left(\frac{\dot{m}}{p} \right) \frac{u_p}{\dot{m}} = \frac{u_p}{p}$

Example: Uncertainty Calculation (cont'd)

Assume the following uncertainties ($u_{x\text{-precision}} = s_x/N^{1/2}$)

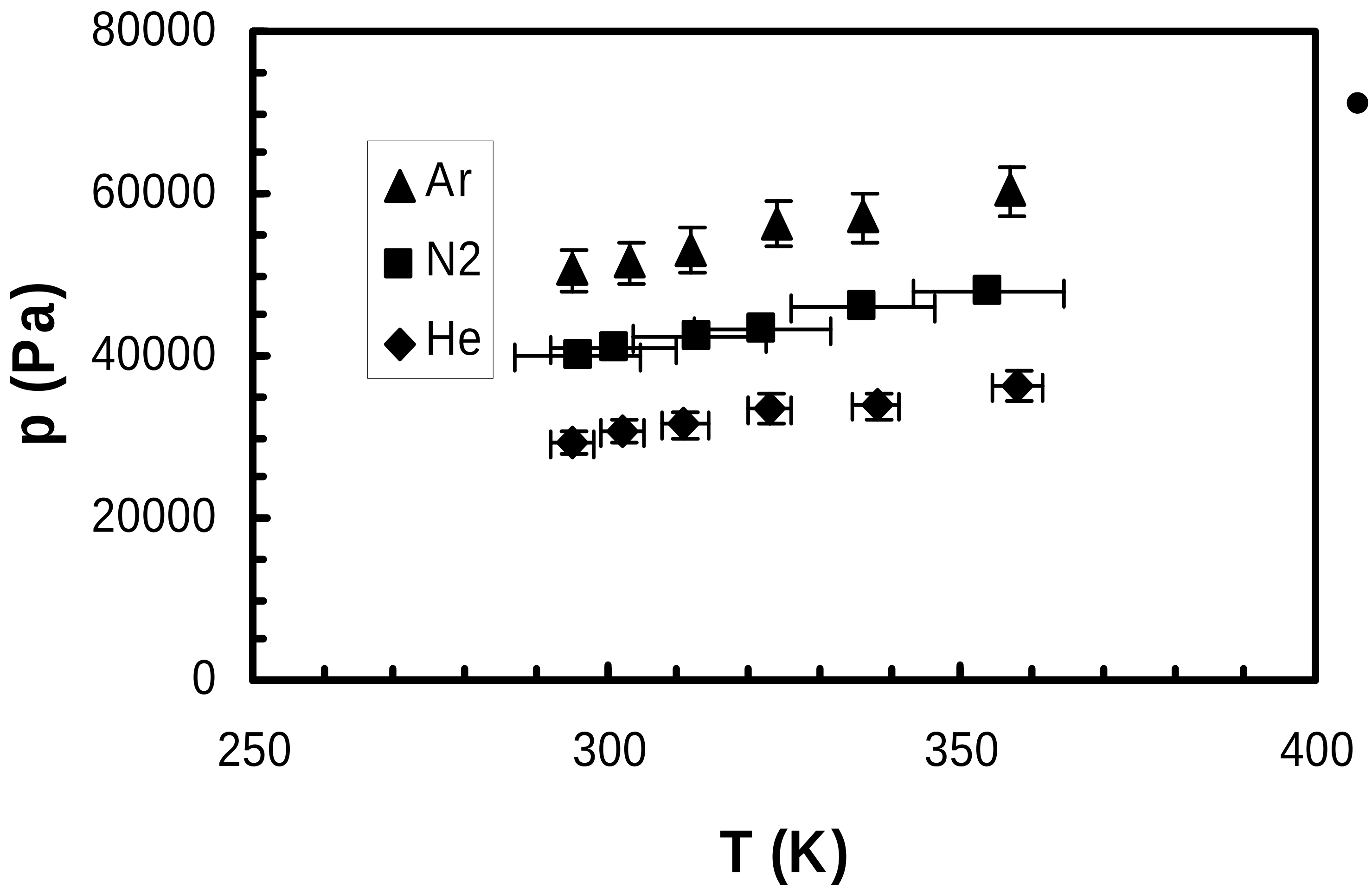
includes factor of 2 for D term

Variable	Accuracy (u_x/x)	Precision (u_x/x)	Notes
p	0.4%	0.1%	Pressure transducer with 8-bit digitizer
T	2%	0.4%	Temp. transducer with 1% full-scale linearity error used at half-scale
Δt	0.01%	2%	Accurate clock, but starting/stopping uncertainty of 0.01 sec for 0.5 sec measurement
L	0.1%	—	Only measured once with ruler having maximum 0.5 mm reading error over 0.5 m pipe length
D	1%	—	Only measured once with with ruler having maximum 0.5 mm reading error over 0.05 m diameter
Summed $(\Sigma u^2)^{1/2}$	2.5%	2.0%	Δt meas. dominates precision error D meas. dominates accuracy error
95% Confidence	2.5%	4.0%	With 95% confidence level precision error dominant

Total uncertainty in a **overall** measurement of $\dot{m}$ is then:

$$\frac{u_{\dot{m}}}{\dot{m}} = \left[\underset{\substack{\uparrow \\ \text{systematic}}}{u_{accuracy}}{}^2 + \underset{\substack{\uparrow \\ \text{random}}}{u_{precision}}{}^2\right]^{1/2} = 4.7\%$$

Plotting Uncertainty - Error Bars



- Can have error bars in vertical and/or horizontal coordinates