457.309 Hydraulics and Laboratory
.05.01 Uncertainty Analysis (1)

Prepared by Jin Hwan Hwang
(2019.03.12)

Today's objectives

- Understanding uncertainty
- Learning how to determine bias and precision errors

Introduction

- No experimental measurement is perfect; on the contrary, the measurement of any variable contains inaccuracies to some degree.
- Thus, it is important for the experimentier to understand the potential sources and magnitudes of errors that could influence his/her measurements.
- Moreover, it is crucial that the experimenter reports the magnitudes of the potential erros, generally in the form of an *uncertainty* along with the measurements themselves.
- In this way, others using the data in engineering applications or for comparison to mathematical theories will know the “degree of goodness” of the data, i.e., the level of confidence to place in it.

Introduction

- For example, it would be completely pointelss to argue which of two theories, that differe by only a few percent agrees “best” with experimental measurements when the uncertainty in the data is ±20%!
- In addiation, uncertainty analysis can be used in the design phase of an experiment to ensure that the accuracy of the various measurements is sufficient to provide a “final answer” with an acceptable level of uncertinty.

Types of Experimental Errors

- The total error (δ):
  - In an experimental measurement, defined as the difference between the true values and the measureed value, is made up of two components: *bias error and precision* β *error.*
- Bias error ( ):
  - The fixed error, systematic error or simply bias
  - The constant component of the total error
  - The sources of bias generally composed of those due to instrument calibration, data acquisition, and data reduction

Types of Experimental Errors

- Bias error:
  - From calibration since no reference or standard is perfect.
  - Acquisition biases include environmental and installation effects, such as temperature, humidity, or electronic effects that causes calibration shifts, as well as biases inherent in the system that condition, acquire, and store the instrument output.
  - Data reduction category results, from example from the use of curve-fit equations or from computational resolution problems due to roundoff or truncation errors.

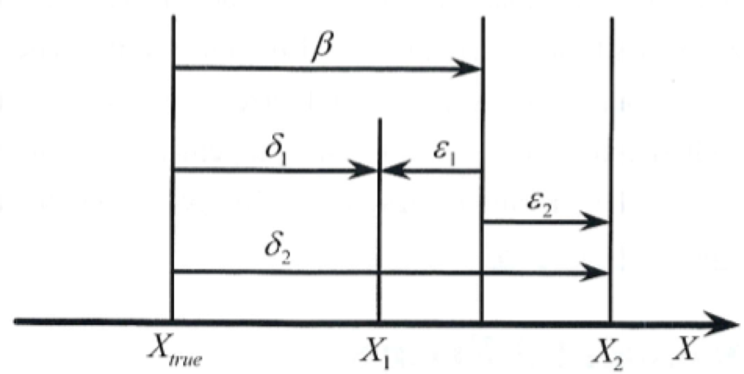
Types of Experimental Errors

- Bias error:
  - Clearly, bias errors can be reduced by careful instrument calibration, *but they can never be completely eliminated*, if for no other reason because of the bias error associated with the (imperfect) standard used in the calibration procedure itself

Types of Experimental Errors

- Precision error (ϵ):
  - The random error, repeatability error
  - Random electronic fluctuations, various influence of fr iction, or inability to maintain truly steady state conditi ons, i.e. any effect that causes lack of repeatability fro m trial to trial of a given experiment.
  - Generally follow a specific statistical distribution.
  - Can be estimated from a sample of N reading using st atistical techniques.

Types of Experimental Errors



– Suppose two measurements X_1 and X_2 of variable X are made.

– The bias error is fixed, it is the same for each measurement.

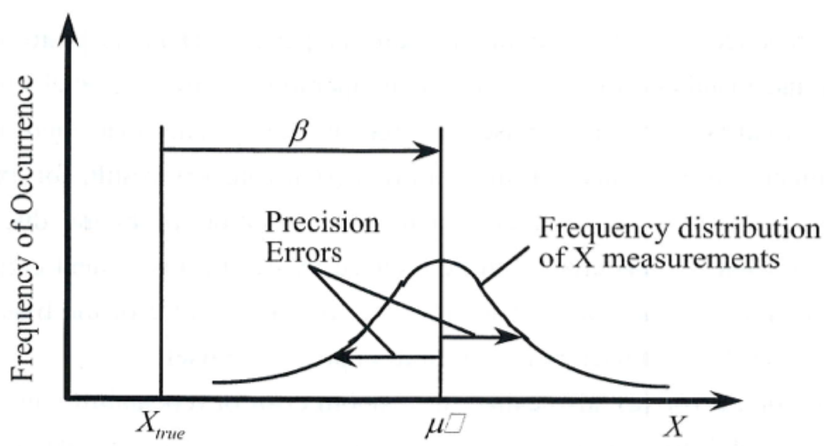
– The precision error is random, its magnitude is different fro each measurement.

– The total error is each measurement, which is the difference between the measurement and the true value, will be different, according to the relation $\delta_i = \beta + \epsilon_i$

Types of Experimental Errors

- In an actural physcial experiment, the true value of the measurement variable X is, unfortunately, unknown.
- Therefore, it is impossible to specify the exact values of the bias and precision errors in a given measurement of X.
- The best alternative is to state that the true values of X lies within the interval

$$X_{best} \pm U_x$$



- The best estimate X_{best} is generally assumed to be the mean values of the N measurements take at a given condition and U_x is an estimate of the uncertainty (at C% confidence level in eng. 95%).

The Gaussian or Normal Distribution

- Parent population:
  - Infinite number of readings could be, in an actual experiment only a limited number of measurements can be obtained due to time and financial constraints:
- Sample distribution:
  - A reduced distribution consisting of a *finite* number of readings taken from the population is called the *sample distribution.*

The Gaussian or Normal Distribution

- Gaussian or normal distibution
  - For situations in which the variation in measurments results from a combination of many small errors of approximately equal magnitude and with each of these errors being equally likely to be positive or negarive, the frequency distribution of an infinite number of measurements is given by the *Gaussian, or normal distibution*

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

  - Where $f(X)dX$ is the probability that a single measurement of variable X will fall between X and $X+dX$.

The Gaussian or Normal Distribution

- Gaussian or normal distibution
  - The mean and standard deviation, respectively as

$$\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$\sigma = \lim_{N\to\infty} \left[\frac{1}{N} \sum_{i=1}^{N} \left(X_i - \mu \right)^2 \right]^{1/2}$$

Sample Distribution from a Gaussian Distribution

- In the preceding section it was assumed the distribution considered was a Gaussian parent population. This would indeed be the case if we took an infinite number of measurements for an experiment subject to random precision errors. (Impossible)
- Therfore, it is necessay to consider the statisitcal properties of *sample populations* consisting of a finite number of measurements drawn from a Gaussian parent population.

Sample Distribution from a Gaussian Distribution

- The mean of the sample population is computed from the usual relation

$$\bar{X}=\frac{1}{N}\sum_{i=1}^{N}X_i$$

  - N is the number of individual measurements.
  - The *precision index* (similar to the standard deviation) of the sample population is defined by

$$S_X=\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(X_i-\bar{X}_i\right)^2\right]^{\frac{1}{2}}$$

  - The factor N-1 occurs here instead of N because $\bar{X}$ in the definition has been calculated from this same sample, causing the loss of one degree of freedom.

Sample Distribution from a Gaussian Distribution

- Another important statistical parameter is the *precision index of the sample mean.* Consider the situation in which multiple sets of N readings each are obtained from a Gaussian population with mean μ and standard deviation σ.
- The means of each of these sets will not be identical and, in fact, they are normally distributed with mean μ and standard deviation

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}.$$

Parent standard deviation

- This relation demonstrates the intuitively epxected restuls that the standard deviation of the *means* of the individual sets of N measurements will be smaller than the standard deviation of the population of the individual measurements themselves.

Sample Distribution from a Gaussian Distribution

- However, the true standard deviation of the parent population is generally unknown, so the *precision index of the mean* is used, which is defined as

Precision index of the mean

$$S_{\bar{X}} = \frac{S_X}{\sqrt{N}}.$$

Sample precision index

- Let's move back to the parent information.
- By rearraining an expression from the preceding section for a Gaussian population with mean and standard deviation,

$$\mathrm{Prob}(X_I - 1.96\sigma \le \mu \le X_i + 1.96\sigma) = 0.95$$

Sample Distribution from a Gaussian Distribution

- In other words, the mean of the parent populatoin is within $\pm 1.96\sigma$ of any individual reading at a 95% confidence level. As just discussed, the mean $\bar{X}$ of a smaple population of size *N* drawn from this Gaussian population is itself normally distributed with standard deviation $\sigma / \sqrt{N}$. Therefore,

$$\text{Prob}\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{N}} \right) = 0.95$$

- Or with 95% confidence the mean of the pipulation is within $\pm 1.96\sigma / \sqrt{N}$ of the sample mean $\bar{X}$ computed from N measurements.

Sample Distribution from a Gaussian Distribution

- However, the two preceding expressions are not particularly useful in a realistic expriment because the population standard deviation is unknown. In general, the prevision index of the sample S_x is what is known, although S_x approaches as the number of sample readings N approaches infinity.
- It is convinient to define the 95% confidence limit for a sample of N measurements from a Gaussian population as the *precision limit* P_x, where

$$P_X = tS_x$$

Sample Distribution from a Gaussian Distribution

- And also the *precision limit of the mean*,

$$P_{\bar{x}} = tS_{\bar{X}} = \frac{tS_X}{\sqrt{N}}$$

- The ANSI/ASME standard on *measurement uncertaiinty* suggests that ***the value of t be given by the t distribution with N-1 degrees of freedom***, ν, ***for N<31 and by t=2.0 for N > 30.***
- The t-distribution is used because the variables $(X-\mu)/S_X$ and $(\bar{X}-\mu)/\left(S_{\bar{X}}/\sqrt{N}\right)$ are found to follow this distribution rahter than the Gaussian distribution.
- v and t values should be checked in the stasistics textbook

The *t* distribution

ν	C = 0.900	0.950	0.990	0.995	0.999
1	6.314	12.706	63.657	127.321	636.619
2	2.920	4.303	9.925	14.089	31.598
3	2.353	3.182	5.841	7.453	12.924
4	2.132	2.776	4.604	5.598	8.610
5	2.015	2.571	4.032	4.773	6.869
6	1.943	2.447	3.707	4.317	5.959
7	1.895	2.365	3.499	4.029	5.408
8	1.860	2.306	3.355	3.833	5.041
9	1.833	2.262	3.250	3.690	4.781
10	1.812	2.228	3.169	3.581	4.587
11	1.796	2.201	3.106	3.497	4.437
12	1.782	2.179	3.055	3.428	4.318
13	1.771	2.160	3.012	3.372	4.221
14	1.761	2.145	2.977	3.326	4.140
15	1.753	2.131	2.947	3.286	4.073
16	1.746	2.120	2.921	3.252	4.015
17	1.740	2.110	2.898	3.223	3.965
18	1.734	2.101	2.878	3.197	3.922
19	1.729	2.093	2.861	3.174	3.883
20	1.725	2.086	2.845	3.153	3.850
21	1.721	2.080	2.831	3.135	3.819
22	1.717	2.074	2.819	3.119	3.792
23	1.714	2.069	2.807	3.104	3.768
24	1.711	2.064	2.797	3.090	3.745
25	1.708	2.060	2.787	3.078	3.725
26	1.706	2.056	2.779	3.067	3.707
27	1.703	2.052	2.771	3.057	3.690
28	1.701	2.048	2.763	3.047	3.674
29	1.699	2.045	2.756	3.038	3.659
30	1.697	2.042	2.750	3.030	3.646
40	1.684	2.021	2.704	2.971	3.551
60	1.671	2.000	2.660	2.915	3.460
120	1.658	1.980	2.617	2.860	3.373
∞	1.645	1.960	2.576	2.807	3.291

*Given are the values of *t* for a confidence level C and number of degrees of freedom $\nu = N - 1$.

Example

- A mercury-in-glass thermometer is used to make 1 consecutive tem perature measurements of a heated pool of water: 97.25, 97.5, 96.7 5, 97.0, 98.5, 97.75, 97.5, 97.25, 97.25, and 97.0 °F

Computing the mean and precision index of the sample:

X_i	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$
97.25	−0.125	0.0156
97.5	0.125	0.0156
96.75	−0.625	0.3906
97.0	−0.375	0.1406
98.5	1.125	1.2656
97.75	0.375	0.1406
97.5	0.125	0.0156
97.25	−0.125	0.0156
97.25	−0.125	0.0156
97.0	−0.375	0.1406
$\sum X_i = 973.75$	$\sum (X_i - \overline{X}) = 0$	$\sum (X_i - \overline{X})^2 = 2.1563$

$$\overline{X} = \frac{\sum X_i}{N} = \frac{973.75}{10} = 97.375 \approx 97.38\ ^\circ F$$

$$S_X = \left[\frac{1}{N-1} \sum (X_i - \overline{X})^2\right]^{\frac{1}{2}} = \left[\frac{2.1563}{9}\right]^{\frac{1}{2}} \approx 0.49\ ^\circ F$$

Example

Using the t-distribution table with ν = N – 1 = 9 degrees of freedom at the 95% confidence level, the value **t = 2.262** is found.

$$\therefore P_X = tS_X = 2.262(0.49°F) = 1.11°F$$

and

$$P_{\bar{X}} = \frac{tS_X}{\sqrt{N}} = \frac{2.262(0.49°F)}{\sqrt{10}} = 0.35°F$$

(a) Based on the calculations above, an additional measurement is expected to lie in the following interval at 95% confidence:

$\bar{X} \pm P_X$ = **97.38 ± 1.11°F,** or in the approximate range: **(96.3, 98.5)°F**

Another way to interpret the calculation for P_X is that the population mean μ is expected to lie within $\pm P_X = \pm 1.11$°F of any individual measurement at 95% confidence.

(b) From the discussion above, at 95% confidence, the population mean μ will lie within

$\bar{X} \pm P_{\bar{X}}$ = **97.38 ± 0.35°F,** or in the approximate range: **(97.0, 97.7)°F**